

# SCIENTIFIC REPORTS

OPEN

## Construction of a High-Density Genetic Map and Quantitative Trait Locus Mapping in the Sea Cucumber *Apostichopus japonicus*

Meilin Tian\*, Yangping Li\*, Jing Jing\*, Chuang Mu, Huixia Du, Jinzhuang Dou, Junxia Mao, Xue Li, Wenqian Jiao, Yangfan Wang, Xiaoli Hu, Shi Wang, Ruijia Wang & Zhenmin Bao

Received: 02 March 2015  
Accepted: 28 August 2015  
Published: 06 October 2015

Genetic linkage maps are critical and indispensable tools in a wide range of genetic and genomic research. With the advancement of genotyping-by-sequencing (GBS) methods, the construction of a high-density and high-resolution linkage maps has become achievable in marine organisms lacking sufficient genomic resources, such as echinoderms. In this study, high-density, high-resolution genetic map was constructed for a sea cucumber species, *Apostichopus japonicus*, utilizing the 2b-restriction site-associated DNA (2b-RAD) method. A total of 7839 markers were anchored to the linkage map with the map coverage of 99.57%, to our knowledge, this is the highest marker density among echinoderm species. QTL mapping and association analysis consistently captured one growth-related QTL located in a 5 cM region of linkage group (LG) 5. An annotated candidate gene, retinoblastoma-binding protein 5 (*RbBP5*), which has been reported to be an important regulator of cell proliferation, was recognized in the QTL region. This linkage map represents a powerful tool for research involving both fine-scale QTL mapping and marker assisted selection (MAS), and will facilitate chromosome assignment and improve the whole-genome assembly of sea cucumber in the future.

Genetic linkage maps have become a critical and indispensable tool in an extensive range of genetic and genomic research<sup>1</sup>. A genetic map with sufficient density and resolution is one of the essential prerequisites for marker-assisted selection (MAS), comparative analysis of genomic synteny, fine mapping of quantitative trait loci (QTL), positioning of candidate genes and facilitating the chromosome assignment of whole-genome scaffolds. The resolution and density of a genetic linkage map primarily depends on the number of mapping individuals and mapping families, as well as that of candidate genetic markers and the corresponding accuracy of genotyping in the mapping families<sup>2</sup>. Therefore, it is understandable that the improvement of linkage map resolution is always accompanied by the technical innovation of marker development. Most of the first generation linkage maps for important aquatic species were constructed using both dominant and codominant markers, such as AFLP, RAPD and SSR markers, in tilapia<sup>3</sup>, channel catfish<sup>4</sup>, Pacific oyster<sup>5</sup>, and white shrimp<sup>6</sup>. However, these maps did not demonstrated a sufficient capability in studies addressing the fine mapping of QTLs or MAS due to a low density and the technical limitations of genetic markers<sup>7</sup>. With advances in molecular genetics, co-dominant markers including microsatellites (SSR) and single nucleotide polymorphisms (SNPs), began to be applied in the construction of linkage maps due to their high abundance and mapping potential<sup>8</sup>. The low efficiency and high cost of traditional strategies for marker development become the major obstacles that restricting the

Ministry of Education Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, 5 Yushan Road, Qingdao 266003, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to S.W. (email: swang@ouc.edu.cn) or R.W. (email: rjwang@ouc.edu.cn)

	Raw-reads	High-quality reads	Sequencing depth
Male	29,077,897	24,588,767(85%)	81×
Female	37,471,929	36,811,645(98%)	104×
Progeny	3,210,181	2,838,723(91%)	17×

**Table 1.** Summary of data filtering of *A. japonicus*.

usage of these techniques in the generation of high-density genetic linkage maps<sup>9</sup>. Recent advances in next-generation sequencing technologies have enabled the identification of sufficient genetic molecular markers at a reasonable price, thereby promoting the development of several high-throughput SNP genotyping methods, particularly the restriction site-associated DNA (RAD) technique<sup>10</sup> and its derivative methods. Among such techniques, 2b-RAD<sup>11</sup>, a representative optimized method that integrates the advantages of simpler library preparation, even and variable genome coverage, and a flexible genotyping platform, has been successfully applied to aquaculture species with large genomes<sup>12–15</sup>.

Echinoderms are the second largest group of deuterostomes, and they are the largest phylum without any freshwater or terrestrial (land-based) species<sup>16</sup>. Echinoderms are important both biologically and evolutionarily<sup>15</sup>. Biologically, they constitute one of the groups that is most abundantly distributed in the biotic desert of the deep sea, in addition to shallower ocean regions<sup>17</sup>. The two remarkable traits of echinoderms, tissue regeneration and aestivation, have become research hotspots attracting increasing attention from biologists<sup>18</sup>. Evolutionarily, it is believed that the radiation of echinoderms was responsible for the Mesozoic Marine Revolution<sup>19</sup>, and ossified skeletons of echinoderms are considered to be the major contributors to many limestone formations, providing valuable resources for tracking the ancient environment<sup>20</sup>. In addition, echinoderms comprise several species that are commercially important food resources around the world, with production of over 177 thousand tons being reported in 2012 and a total value of up to 622 million dollars<sup>21</sup>. Hence, genetic inheritances underlying economically important traits, such as growth and reproduction, have been central tasks in echinoderms genetic and breeding studies on echinoderms. Great progress has recently been achieved using RNA-Seq technology for the large-scale identification of candidate genes and SNPs in the echinoderms transcriptome<sup>22</sup>. However, the development of linkage maps has stalled in the whole echinoderm phylum. The only linkage maps constructed thus far are for *Strongylocentrotus nudus*, *Strongylocentrotus intermedius* and *Apostichopus japonicus*, using AFLP, SSR and SNP markers<sup>23,24</sup>. The density and resolution of these maps is generally low (mostly 7.0–17.1 cM), thus limiting their application in QTL related studies.

The sea cucumber *Apostichopus japonicus*, a representative echinoderm species, has become one of the most dominantly and economically important aquaculture species in the Western Pacific Ocean along the coasts of China, Japan, Korea and Russia, with considerable edible and medicinal value<sup>25</sup>. Sea cucumber breeding was first studied in China in 1954, and great progress was subsequently made in the breeding and larval rearing of sea cucumbers in the late 1990s<sup>26</sup>. In 2012, the farming area of *A. japonicus* in China had reached 153 thousand square hectometers, and production was more than 109 thousand tons<sup>27</sup>. The great expansion of large-scale sea cucumber farming has led to an increasing demand to promote aquaculture traits using MAS. Significant progress has been made in sea cucumber genomic research, particularly in the development of genetic and genomic resource including various types of polymorphic markers<sup>22</sup>, two low density linkage maps<sup>23,28</sup> and great numbers of transcriptome contigs<sup>29</sup>. In this study, we describe the construction of a genetic linkage map with over seven thousand markers for *A. japonicus*, which is, to our knowledge, the first high-density, high-resolution genetic map to be reported for echinoderms. One major growth-related QTL interval was identified utilizing this map. This linkage map represents a powerful tool for research involving both fine-scale QTL mapping and MAS and will facilitate chromosome assignment for sea cucumber whole-genome assembly in the future.

## Results

**2b-RAD sequencing and selection of markers.** Single-end sequencing of 2b-RAD libraries was conducted using the Illumina HiSeq2000 platform on 100 progenies and their parents. A total of over 387 million reads were generated, including 29 million reads from the male parent, 37 million reads from the female parent, and 321 million reads from the progenies (3.2 million reads per progeny). The statistical sequencing depth corresponded to 81× in the male, 104× in the female and 17× in the 100 progenies. After quality trimming, 85% of the reads from the male parent, 98% of the reads from the female parent, and 91% of the reads from the progenies were carried forward for analysis (Table 1 and Additional File A1). The raw read data were archived at the NCBI Sequence Read Archive (SRA) under Accession Number PRJNA274721.

The genotyping procedure was performed using RADtyping v1.0<sup>30</sup>. The result showed that clustering of parental reads generated 186,692 representative reference tags, including 125,616 codominant tags and 61,076 dominant tags, 7769 low quality tags were removed according to their sequencing depth. In this study, we defined “dominant tags” and “dominant markers” as 2b-RAD tags that appear in only one parent. And “codominant tags” was defined as 2b-RAD tags that appear in both parents with allelic

	Number
Markers detected in parents	104,771
Markers polymorphic and with high genotype percentage	32,012
Markers used for linkage mapping after Mendelian ratio trimming	11,306
Markers used for female linkage mapping	6311
Markers used for male linkage mapping	5517

**Table 2. Markers selected for linkage mapping.**

variations, while “codominant markers” represented those SNPs that detected in “codominant tags” (Supplementary Figure F1). Utilizing the constructed reference, 47,049 codominant markers and 57,722 dominant markers were detected. After initial genotyping, a total of 32,012 markers that showed polymorphism in at least one parent and 3 progenies were retained, including 20,169 codominant markers and 11,843 dominant markers (Table 2). Then the mendelian fit test and genotyping percentage were performed for further trimming. In total 11,306 markers that fit mendelian ratios ( $P \geq 0.05$ ) and high genotyping rate (got available genotype over 80 progenies) were used to construct sex specific map. Ultimately, 6311 and 5517 markers were used for female-specific and male-specific linkage map construction, respectively (Table 2 and Additional File A2).

**Linkage mapping and analysis of recombination rates.** Linkage maps were first constructed separately for the male parent and female parent using Joinmap 4.0<sup>31</sup> (LOD = 5.0). In total 6311 and 5517 markers were used to construct female and male map separately, after linkage mapping, 2302 and 1525 markers were eliminated. For both sex-specific maps, 22 linkage groups (LGs) were obtained (Supplementary Figure F2 and F3), which were consistent with the haploid chromosome number of *A. japonicus*<sup>32</sup>. The female genetic map consisted of 4009 markers mapped on 2908 distinct positions spanned 3728.9 cM, with an average marker interval of 0.93 cM, ranging from 0.72 in LG5 to 1.31 in LG14, while the male genetic map contained 3992 markers on 2138 distinct positions, with a much shorter genetic length (1824.3 cM) and shorter average marker interval (0.46 cM) (Table 3). The marker interval of the male genetic map ranged from 0.26 in LG8 to 0.99 in LG7 (Table 3). Overall, the length of the female genetic map was 1904.6 cM longer than the male genetic map, with an average female-to-male length ratio of 2.11:1. This ratio varied by linkage groups, ranging from 1.21:1 in LG1 to 3.45:1 in LG12 (Table 3).

Considering the relationship between interval distance and recombination, this result demonstrated significantly higher recombination rates in a majority of the linkage groups of the female compared with the male (t-test,  $p < 0.0001$ ). Significantly different recombination rates (female-to-male recombination rate ratio  $> 2.5$ ) between the female and male maps were observed in LG2, LG3, LG4, LG5, LG8, LG17, and LG19 (Table 3). Across all of these linkage groups, the female:male recombination ratio ranges from 0.66:1 in LG 20 to 4.86:1 in LG 19 (Table 3). In total, 162 markers were shared between the two sexes, and the average recombination rates of these markers in each linkage group were not the same as for the sex-specific markers (Fig. 1). The female:male ratios for the recombination rates of shared markers for each linkage group ranged from 0.67:1 in LG5 to 4.32:1 in LG12. The highest recombination rate ratios (greater than 2.5) for markers shared between the sexes were observed in LG8 4.32:1, 2.61:1 in LG12 and 3.64:1 in LG17 (Supplementary Table T1), whereas the corresponding average recombination ratios for the sex-specific maps were 3.28:1, 1.44:1 and 2.58:1 (Table 3). While the recombination rates of shared markers in LG2 and LG14 were similar between the sexes (0.92:1, 1.06:1, respectively), various recombination rates were observed for the sex-specific maps (Supplementary Table T1).

The consensus linkage map for *A. japonicus* was constructed after integration of the two sex-specific maps using 162 heterozygous markers from both parents. This consensus map comprised 7839 markers representing 4884 distinct positions spanning 3706.6 cM, with an average resolution of 0.47 cM (Fig. 2). The size of each linkage group ranged from 100.6 cM to 268.8 cM, and the number of loci varied from 174 to 618 per linkage group, with an average of 168.5 cM and 356 loci per group (Table 4). According to the constructed linkage maps, the patterns of marker distribution across chromosomes can be examined (Supplementary Figure 2, 3 and Fig. 2). Clustered markers were commonly found in regions around the center of linkage groups and less frequently observed around terminal regions. This result was consistent with observations in various genetic maps of aquatic species previously<sup>33,34</sup>.

**Map length, map coverage and map validation.** According to the two different algorithms, the expected map length was estimated to be 3716.9 cM ( $G_{e1}$ ) and 3728.5 cM ( $G_{e2}$ ), with an average expected map length of 3722.7 cM ( $G_e$ ) being obtained. The map coverage of this consensus map was 99.57% based on the expected map lengths according to the average length calculated from the following formulas: 1)  $G_{e1} = G_{of} + 2s$ , in which  $G_{of}$  represents the observed length of each linkage group, and  $s$  represents the

LG	Female-specific map					Male-specific map					F:M ratio	
	Mapped markers	Distinct positions	Genetic length (cM)	Marker interval (cM)	Average recombination rate	Mapped markers	Distinct positions	Genetic length (cM)	Marker interval (cM)	Average recombination rate	Map length ratio	Average recombination rate ratio
1	233	178	171.8	0.74	0.005	288	176	142.4	0.49	0.007	1.21	1.49
2	154	89	115.4	0.75	0.039	263	120	89.6	0.34	0.136	1.29	3.54
3	130	109	136.6	1.05	0.072	188	115	82.8	0.44	0.214	1.65	2.98
4	169	125	179.9	1.06	0.065	221	95	84.6	0.38	0.175	2.13	2.70
5	205	135	148.1	0.72	0.071	213	95	78.6	0.37	0.151	1.88	2.13
6	95	67	99.8	1.05	0.126	127	63	78.4	0.62	0.339	1.27	2.70
7	184	140	168.5	0.92	0.246	68	40	67.7	0.99	0.183	2.49	0.74
8	193	110	148.3	0.77	0.056	312	106	82.5	0.26	0.182	1.80	3.28
9	223	159	204.4	0.92	0.083	219	120	79.7	0.36	0.167	2.56	2.01
10	355	251	271.1	0.76	0.009	276	129	114.9	0.42	0.012	2.36	1.33
11	158	117	170.8	1.08	0.031	107	73	75.9	0.71	0.039	2.25	1.24
12	280	212	218.7	0.78	0.021	193	95	63.3	0.33	0.030	3.45	1.44
13	186	146	211.5	1.14	0.027	175	102	78.1	0.45	0.056	2.71	2.06
14	86	62	112.4	1.31	0.059	93	63	66.1	0.71	0.135	1.70	2.31
15	203	143	159.3	0.78	0.034	184	107	78.4	0.43	0.065	2.03	1.92
16	142	95	162.6	1.15	0.057	120	67	60.8	0.51	0.104	2.67	1.83
17	179	129	156.6	0.87	0.035	216	101	81.8	0.38	0.091	1.91	2.58
18	186	157	210.7	1.13	0.047	181	116	80.0	0.44	0.099	2.63	2.13
19	89	76	138.1	1.55	0.046	201	147	83.7	0.42	0.224	1.65	4.86
20	260	185	201.0	0.77	0.134	81	54	61.7	0.76	0.088	3.26	0.66
21	149	114	163.0	1.09	0.088	134	76	95.0	0.71	0.165	1.72	1.89
22	150	109	180.3	1.20	0.096	132	78	98.3	0.74	0.176	1.83	1.83
Total	4009	2908	3728.9	0.93	0.066	3992	2138	1824.3	0.46	0.129	2.11	2.17

**Table 3.** Summary of sex-specific linkage maps of *A. japonicas*.

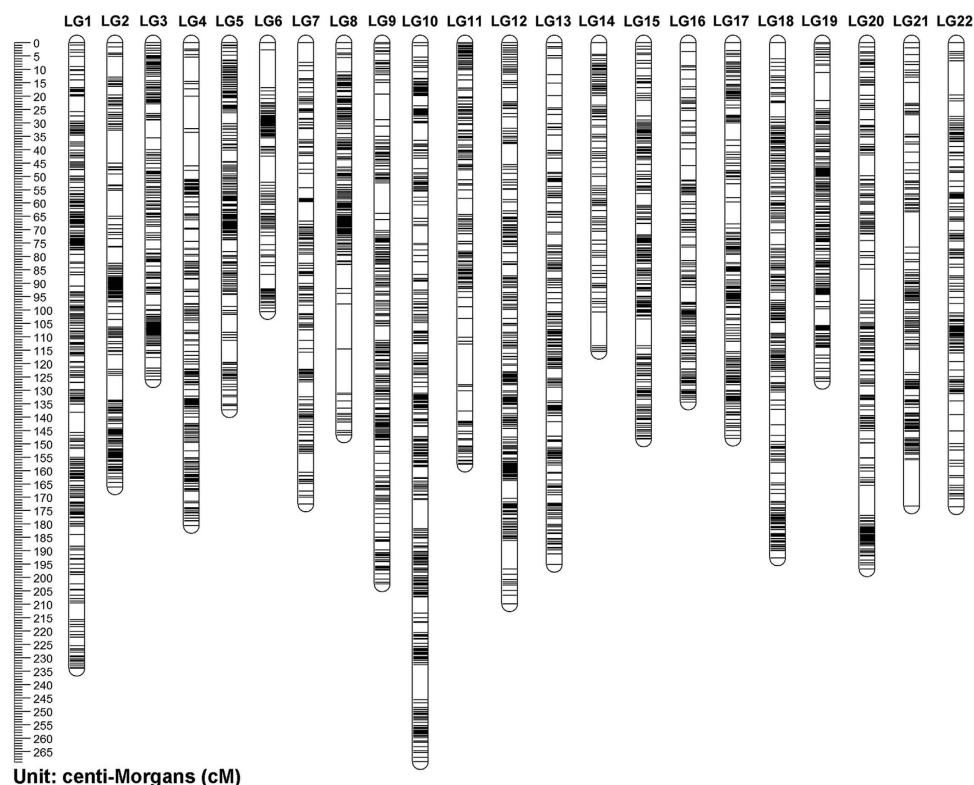
average interval of the whole map; 2)  $G_{e2} = L \times (m + 1)/(m - 1)$ , where  $L$  represents the observed length of the linkage groups, and  $m$  is the number of markers for the corresponding group<sup>35</sup>.

To confirm the reliability of the genotyping information used in linkage map construction, we performed re-sequencing of both parents and randomly selected 10 progenies. The results showed that the genotyping consistency was as high as 96.9%, 95.0% and 94.06% in the male parent, female parent and offspring, respectively (Supplementary Table T2). The validation results provided solid evidence of an acceptable quality of the high-density linkage map.

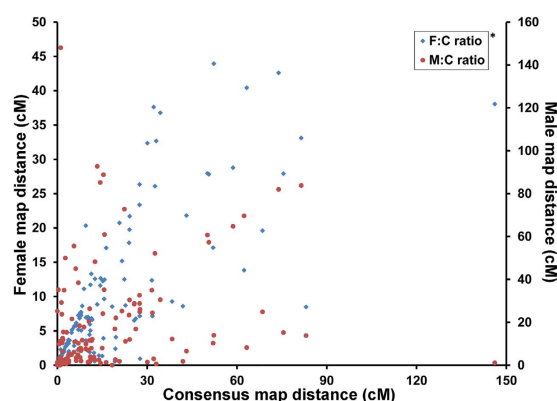
**QTL mapping and association analyses of growth trait.** Multiple QTL model (MQM mapping) in MapQTL<sup>36</sup> package was used for analyzing the body weight information of 100 progenies. After QTL mapping, a linkage region on LG5 was detected to pass the corresponding LOD threshold ( $\text{LOD} \geq 3.7$ ) of the corresponding permutation test and was considered as the major QTL in this study. This QTL (SNP37318) was situated at 79.95–84.95 cM in LG5, explaining 11.8% of the phenotypic variation in body weight observed in *A. japonicus*. In total 7839 markers on the linkage map were performed for association analysis using TASSEL 5.0<sup>37</sup> and the P-value was adjusted to WFDR-P by WFDR (Weighted False Discovery Rate)<sup>38</sup> with cutoff 0.1 (Additional File A3). Association analysis showed similar distribution patterns to the QTL mapping analysis across all linkage groups (Fig. 3a), supporting the results of the QTL analyses overall. The nine closest loci (Fig. 3b) located near the growth-related QTL genomic window ( $\pm 2.5$  cM) were defined as the candidate locus relevant to the grow traits of *A. japonicus*. Eight of these nine loci could be mapped onto eight scaffolds of the *A. japonicus* draft genome (unpublished data), among which one 47.2 kb scaffold (Contig98127) had available gene annotation information (Table 5 and Fig. 3b). One annotated gene, retinoblastoma-binding protein 5 (*RbBP5*), which has been reported to be an important regulator in cell proliferation<sup>39</sup>, was 190 bp downstream of the candidate markers (SNP50634).

## Discussion

In the present work, we constructed the first high-density genetic map reported for echinoderms, in the sea cucumber aquaculture species *A. japonicas*. This map possesses the highest marker density among



**Figure 1. Recombination rate of shared markers between sex-specific maps and consensus map.** This diagram was constructed using the recombination rate of 162 shared markers shared by sex-specific maps and consensus map. The left Y-axis represents shared marker interval in female map and right Y-axis represents the shared marker interval in male map, the X-axis stands for shared marker interval in consensus map. The blue dots stand for shared marker interval ratio between female map and consensus map (F:C ratio), while the red dots represent shared marker interval ratio between male map and consensus map (M:C ratio). \*F:C ratio represents the recombination rate correlation of shared markers between female map and consensus map; M:C ratio stands for the recombination rate ratio between male map and consensus map.



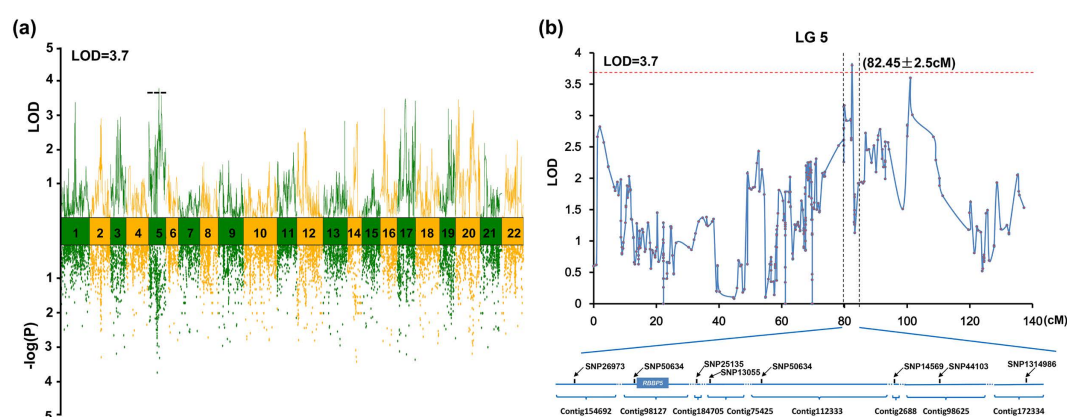
**Figure 2. The high-density consensus linkage map of *A. japonicus*.** The consensus map which contained 7839 markers in 22 linkage groups was constructed through combining the male and female linkage maps (Supplementary Figure F2 and F3).

all of the genetic linkage maps constructed for any echinoderm species. Taking advantage of 2b-RAD technology, we were able to cost-effectively genotype 32,012 polymorphic markers in 102 sea cucumbers (two parents and 100 offspring) from one full-sibling families (Table 2). Such a high-density linkage



Linkage group	Consensus map						
	Mapped markers	Female-specific makers	Male-specific markers	Shared makers	Distinct positions	Genetic length (cM)	Marker interval (cM)
1	506	202	264	40	339	234.0	0.46
2	416	151	257	8	208	166.1	0.40
3	309	110	171	28	215	126.1	0.41
4	383	155	212	16	213	180.5	0.47
5	415	193	208	14	227	137.3	0.33
6	219	86	120	13	127	100.6	0.46
7	245	174	59	12	173	172.5	0.70
8	495	181	299	15	206	146.7	0.29
9	437	212	211	14	274	202.4	0.46
10	618	332	257	29	367	268.8	0.43
11	257	139	97	21	182	157.7	0.61
12	462	263	180	19	296	209.8	0.45
13	348	162	160	26	235	195.1	0.56
14	174	78	83	13	120	115.5	0.66
15	374	183	166	25	237	148.2	0.40
16	257	135	114	8	157	134.4	0.52
17	391	173	201	17	226	148.0	0.38
18	361	175	174	12	267	192.7	0.53
19	285	79	186	20	218	126.7	0.44
20	334	244	73	17	232	196.8	0.59
21	280	142	129	9	187	173.2	0.62
22	273	134	122	17	178	173.6	0.63
Total	7839	3703	3743	393	4884	3706.6	0.47

**Table 4.** Summary of the consensus linkage map of *A. japonicus*.



**Figure 3.** Growth-related QTL mapping and association analysis in *A. japonicus* among all linkage groups (a) and the candidate region of the major QTL in LG5 (b). The dashed line indicates the genome-wide significance threshold where  $\text{LOD} = 3.7$ . The vertical dashed line stands for the growth trait related QTL window ( $82.45 \pm 2.5 \text{ cM}$ ).

map is expected to be a valuable resource for genomic analyses and fine-scale QTL mapping in the sea cucumber *A. japonicus*.

“High density” refers to a compact marker interval ( $< 2 \text{ cM}$ ), high map coverage and even marker distribution in a genetic linkage map<sup>40</sup>. This term usually indicates that tens of thousands candidate polymorphic markers are obtained for species with large ( $> 500 \text{ MB}$ ) and complex genomes<sup>41</sup>. Therefore, a cost-effective sequencing and genotyping plan which can balance the marker quantity, sequencing

Maker	Genetic distance (cM)	Scaffold ID	Scaffold length (bp)	Annotation
SNP2336306	80.01	–	–	–
SNP26973	80.61	Contig154692	44,607	–
SNP50634	82.04	Contig98127	47,215	<i>RBBP5</i>
SNP25135	82.19	Contig184705	2493	–
SNP37318	82.45	–	–	–
SNP12502	83.11	–	–	–
SNP13055	83.26	Contig75425	26,055	–
SNP14569	83.26	Contig112333	98,411	–
SNP3266	83.26	Contig2688	3525	–
SNP44103	83.26	Contig98625	40,991	–
SNP1314986	83.33	Contig172334	26,407	–
SNP35302	84.42	–	–	–

**Table 5. Information of growth-related QTL window in LG5.**

coverage, sequencing cost and genotyping accuracy, is the foremost guaranteed for the construction of large-scale linkage maps in the non-model invertebrate species without the whole genome reference, such as sea cucumber, *A. japonicus*<sup>30</sup>. An optimized 2b-RAD sequencing plan<sup>11</sup> was applied in this study, which can easily provide the sequencing coverage required for a given sequencing depth, considering the genome size (880 MB) reported for sea cucumber<sup>42</sup>. As other next-generation sequencing-based technology, 2b-RAD possess the following advantages: 1) lower sequencing cost due to the sequencing of a subset of BsaXI sites derived from RR libraries, rather than a whole genome; 2) an adjustable marker density due to utilizing selective adaptors; and 3) a relatively even marker distribution, benefiting from the ubiquitous restriction sites in the genome<sup>11</sup>.

High levels of polymorphism and heterozygosity have come to be considered universal hallmarks of most invertebrate genomes, including those of several mollusks, chordates and echinoderms<sup>43–45</sup>. In sea urchin, the existence of over 2% genomic polymorphism markedly decreased the performance of genome assembly compared with other vertebrates<sup>44</sup>. Not coincidentally, a high degree of genome polymorphism and heterozygosity have significantly increased the difficulty of the assembly process in the ongoing starfish genome project (EchinoBase, <http://www.echinobase.org/Echinobase/PmBase>). Conversely, the level of genomic polymorphism can be inferred from the development of polymorphic markers, such as AFLPs, SSRs and SNPs<sup>46</sup>. In this study, 32,012 polymorphic markers were identified in the evaluated full-sib mapping family after 2b-RAD genotyping (Table 2). This number was 4.29 times higher than that obtained in a previous 2b-RAD study in *Chlamys farreri*<sup>12</sup>. One technical difference between these two projects is the utilization of two types of selective adaptors (5'-NNT-3' and 5'-NNA-3') in this study, while one type of selective adaptor (5'-NNT-3') were used in *Chlamys farreri*<sup>12</sup>. The selective adaptor used in this study means one type of adaptor with a single selective base at 3' end. Theoretically, compared with using only one type of selective adaptor, the extra type of selective adaptor will quadruple the restriction sites in the genome leading to the proportional expansion of the polymorphic markers (Supplementary Figure F4). However, after the consideration on the smaller genome size of *A. japonicus* (~880 MB)<sup>42</sup> than *C. farreri* (~1.24 GB)<sup>47</sup>, the total amount of polymorphic markers identified from sea cucumber was still relatively high. The abundance of polymorphic markers implied the high polymorphisms, complexity, and heterozygosity features of *A. japonicus* genome, which required special attentions and considerations in the future whole genome assembly.

The construction of genetic linkage maps for echinoderms is still in its infancy. To date, only a few genetic linkage maps have been constructed for echinoderm species<sup>23,24,28</sup>. However, these maps were generally constructed at a low density, using several hundred molecular markers at most. Even for the sea cucumber *A. japonicus*, for which the greatest number of linkage maps are available, the mapping density was still above 7.0 cM<sup>23,28</sup>. An insufficient map density has been one of the barriers limiting the development of both genomics and genetics. In this study, we adopted the double pseudo-testcross strategy to construct the first high-density SNP genetic linkage map for *A. japonicus* by taking advantage of parental heterozygosity<sup>42,48</sup>. The final consensus linkage map contained 7839 markers with a resolution of 0.47 cM (Table 4), surpassing the resolution of not only the previous genetic maps for *A. japonicus* but also that of linkage maps reported for other echinoderms (7.0–17.1 cM)<sup>23,24,28</sup>, and is even competitive with other genetic linkage maps constructed through 2b-RAD (0.39–0.41 cM)<sup>12,15</sup>.

Another common issue of RAD-based method is allele/locus dropout problems in genotyping for dominant tag. Here we utilized a new analytical tool, RADtyping, which provides a statistical algorithm to avoid sampling error for accurate *de novo* dominant genotyping in mapping population<sup>30</sup>. Specifically,

we first removed the low-quality sites that are not supported by parental reads in sufficient depth (i.e. the requirements  $d_{p1} < l_{p1}$  and  $d_{p2} < l_{p2}$ ), where the thresholds are determined by:

$$l_{pj} = \max \left\{ \sum_{0 \leq m \leq d} \text{poisson}(m|C_j) \leq 0.05 \right\} \quad (1)$$

$C_j$  is the mean sequence depth of the  $j$ th parent ( $j = 1, 2$ ). Then, the low-coverage sites in the progenies with  $d_{pro}$  less than  $l_{pro}$  were removed, where  $d_{pro}$  is calculated for each site by summarizing all progenies having reads derived from that site, and the threshold  $l_{pro}$  is determined for each site using this formula again with parameter  $C_j$  being the average depth of all progenies. At the stage of the dominant genotyping, supposing that the cluster depth of the  $i$ th site for the  $j$  progeny is  $d_{ij}$ , this site is called as “presence” if  $d_{ij} > l_d$ , “absence” if  $d_{ij} = 0$ , and unknown if  $d_{ij}$  ranges from 0 to  $l_d$ , where the threshold  $l_d$  is determined using formula with  $C$  being the mean sequencing depth of the  $i$ th site. The performance of above statistical algorithm has been thoroughly evaluated using both simulated and real sequencing datasets previously with genotyping accuracy higher than 97.6% for all cases<sup>30</sup>.

The sea cucumber *Apostichopus japonicus* is one of the most important species of commercial echinoderms in aquaculture. Growth traits are of particular interest to *A. japonicus* researchers due to their high commercial significance in the sea cucumber culture industry. QTL mapping represents an efficient approach for identifying the genetic loci underlying these traits, to allow marker-assisted selection to be applied in genetic breeding. One major QTL related to total body weight was mapped to a 79.95–84.95 cM region in LG5 in this study (Fig. 3b). Thus, this region was considered to be a candidate genomic region involved in controlling the growth of *A. japonicus*. Although the gapped draft genome assembly (unpublished data) may limit QTL candidate gene mining, a growth-related gene, retinoblastoma-binding protein 5 (*RbBP5*), was recognized within the loose window close to this QTL (Fig. 3b). *RbBP5* encodes a ubiquitously expressed nuclear protein that belongs to a highly conserved subfamily of WD repeat proteins. This protein binds directly to retinoblastoma protein, which regulates cell proliferation<sup>49</sup>. *RbBP5* has also been reported to show the ability to regulate *HOXA9* and *HOXC8* to control cell growth and cell differentiation<sup>50</sup>. Knockdown of *RbBP5* led to a reduction of the expression of the HOX gene family members *HOXA9* and *HOXC8*, which play a crucial role in embryonic and fetal development<sup>51</sup>. Our QTL analysis provided a candidate gene that could be exploited in the future selective breeding of *A. japonicus* with the aim of improving sea cucumber aquaculture production. Further studies are necessary to explore the detailed regulatory mechanisms of *RbBP5* involved in cell proliferation and growth in sea cucumber with more families and populations.

## Materials and Methods

**Resource family and DNA extraction.** A total of nine full-sib families were established in April 2009 through mating between a pair of common male and nine different females. The parents were selected from wild population in Penglai, Shandong province. And the families were constructed in an aquatic breeding farm in the city of Weihai, Shandong province of China. Body weights were measured and recorded for all families at the age of one year. One of these families exhibiting high within-family variation in growth traits was chosen for linkage and QTL analysis. A total of 100 one-year-old offspring were randomly selected from this family, and mixed tissues from all specimens, including parents and offspring, were collected and stored in 70% ethanol for DNA extraction. Genomic DNA was prepared following previously described traditional phenol-trichloromethane DNA extraction procedures<sup>52</sup>. The body tissues of experimental individuals were firstly cut into pieces and digested in 2% CTAB by holding at 56°C for 1 hour. After digestion process, cycles of phenol-trichloromethane extraction were conducted to eliminate protein. Then DNA was precipitated in ethanol and washed by 70% ethanol for 2–3 times to eliminate micro molecule. RNA digestion was conducted by adding proper quantities of RNase and holding at 37°C for 30 minutes. After concentration and integrity examination, DNA samples could be used for further experiments. All of the procedures involved in the handling and treatment of sea cucumbers during this study were approved by the Ocean University of China Institutional Animal Care and Use Committee (OUC-IACUC) prior to the initiation of the study. And all experiments and relevant methods were carried out in accordance with approved guidelines and regulations of OUC-IACUC.

**2b-RAD sequencing.** Before the experiment was carried out, a digital-enzyme-cut analysis was conducted on the draft genome assembly to confirm the sufficient amount of restriction sites. According to the analysis of the restriction sites, the two types of selective adaptors (5'-NNT-3' and 5'-NNA-3') with a selective base on the 3' end were finally decided to generate reasonable quantities of tags used for further analysis. For the construction of 2b-RAD libraries, two parents and 100 offspring were used, following the standard protocol<sup>11</sup>. Briefly, DNA was digested with BsaXI at 37°C for 4 hours and then ligated to adaptors at 4°C for 16 hours to provide the complementary sequences for primers. For the parents, standard BsaXI libraries were constructed, while for progenies, reduced representation (RR) libraries were constructed using adaptors with 5'-NNT-3' and 5'-NNA-3' overhangs to target a subset of BsaXI fragments in *A. japonicus*. Pre-amplification reactions were performed using BsaXI primers to enrich the DNA fragments ligated to adaptors (14 cycles, annealing temperature 60°C). The amplification



products were purified via retrieval from 8% polyacrylamide gels. The 2<sup>nd</sup> amplification proceeded using a BsaXI-barcode primer with two selective bases and the pre-amplification products as a template to distinguish libraries for different individuals (7 cycles, annealing temperature 60°C). The 2b-RAD libraries were then pooled for single-end sequencing using the Illumina HiSeq2000 system.

**Preprocessing and genotyping of sequence data.** Raw reads were processed for initial trimming using a homemade Perl script. Adaptor sequences, unreliable reads (no restriction sites or 'N' at the end of reads), long homopolymeric regions (>10 bp), low-quality sequences (>5 positions with a quality <0) and mitochondrial origin were removed<sup>30</sup>. RADtyping v1.0 (<http://www2.ouc.edu.cn/mollusk/detailen.asp?Id=727>) was employed in the following genotyping procedure due to its acceptable performance for 2b-RAD data<sup>12</sup>. Sequences from parents with BsaXI tags were first clustered under default parameters ( $-m\ 3$ ,  $-M\ 2$ ). The clusters with low coverage (<8) or excessive coverage (>3000) were removed to avoid the interference of sequencing errors and repetitive elements. A total of 7769 clusters were removed and the remaining clusters were used to construct the reference sequences and identify the markers. The clusters were further classified into codominant tags and dominant tags for subsequent codominant markers and dominant markers genotyping. Sequencing reads of offspring were mapped on the constructed high-quality reference. For codominant markers with sufficient reads coverage (6–500), we used iML algorithm<sup>53</sup> to estimate their homozygous or heterozygous. On the other side, dominant markers were recorded as “presence” or “absence” by RADtyping to reflect its existence. Markers that showed polymorphism in at least one parents and 3 progenies were kept for further trimming. Then the markers that fit Mendelian ratios ( $P \geq 0.05$ ) and sufficient genotype rate (got available genotype over 80 progenies) were used to construct sex specific map.

**Linkage map construction and map validation.** The dominant markers that segregated at a 1:1 ratio in the map family were obtained and categorized as  $lm \times ll$  (markers from the male parent) or  $nn \times np$  (markers from the female parent). Markers present in both parents that segregated at a 1:2:1 ratio were also retrieved and were defined as  $hk \times hk$  (markers in both parents). Prior to markers grouping, the markers were tested for goodness-of-fit based on the expected Mendelian ratios through chi-square tests to eliminate markers that significantly deviated from the expected ratios ( $p\text{-value} \leq 0.05$ ). Male- and female-specific linkage maps were then grouped and constructed using the qualified markers with JoinMap 4.0<sup>31</sup>. The genetic positions of the markers were determined using an LOD score cut-off of 6.0, and the recombination frequencies were converted into map distances in centi-Morgans (cM) through the Kosambi mapping function. The consensus map was then established by integrating sex-specific maps through shared markers using MergeMap<sup>54</sup>, map weight was set to 1.0 for both sex-specific maps. The visualized linkage maps were subsequently drawn using MapChart 2.2<sup>55</sup>.

To confirm the reliability of the genotyping information used in linkage map construction, re-sequencing was employed in two parents and 10 progenies using the HiSeq2000 platform as full technical replicates to confirm the accuracy of genotyping (supplementary table T2).

**Estimation of expected map length and map coverage.** Once the linkage map was constructed, the expected map length ( $G_e$ ) of *A. japonicus* was estimated in two different ways. Expected map lengths was calculated by the following formulas: 1)  $G_{e1} = G_{of} + 2s$ , in which  $G_{of}$  represents the observed length of each linkage group, and  $s$  represents the average interval of the whole map; 2)  $G_{e2} = L \times (m + 1) / (m - 1)$ , where  $L$  represents the observed length of the linkage groups, and  $m$  is the number of markers for the corresponding group<sup>35</sup>. Briefly, two different methods were employed to estimate expected map length  $G_{e1}$  and  $G_{e2}$ <sup>35</sup>, and the average of these two indexes was used as the expected map length. The observed map coverage ( $C_{oa}$ ) for *A. japonicus* was determined by calculating the  $G_{oa}/G_e$  value, where  $G_{oa}$  was the observed total genetic map length.

**QTL mapping of growth traits.** QTL mapping analysis was performed for the body weight of *A. japonicus* using MapQTL<sup>36</sup>. The LOD scores were first analyzed using the multiple QTL model (MQM mapping), after which genome-wide and chromosome-wide LOD significance thresholds at the 95% level were determined with a 1000-permutation test for body weights, and QTLs with LOD scores greater than the LOD threshold at 95% were declared significant. Once a QTL was detected, the confidence interval was calculated followed the protocol of Li's method<sup>56</sup>.

Association analyses between genotypes and growth traits were conducted using TASSEL 5.0<sup>37</sup>, which performs a least squares fixed effects linear model of quantitative trait on the genotype. Integrating the result of QTL mapping and association analysis, when a QTL was captured for significant correlation with body weight, sequencing tags were extracted. Then both homolog based analysis (BLAST and gene-wise alignment) and de novo gene prediction (Augustus<sup>57</sup>) was performed on the corresponding genomic scaffolds.

## References

1. Knapik, E. W. *et al.* A microsatellite genetic linkage map for zebrafish (*Danio rerio*). *Nature genetics* **18**, 338–343 (1998).
2. Daw, E. W., Thompson, E. A. & Wijsman, E. M. Bias in multipoint linkage analysis arising from map misspecification. *Genetic epidemiology* **19**, 366–380 (2000).

3. Kocher, T. D., Lee, W.-J., Sobolewska, H., Penman, D. & McAndrew, B. A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* **148**, 1225–1232 (1998).
4. Liu, Z., Karsi, A., Li, P., Cao, D. & Dunham, R. An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* **165**, 687–694 (2003).
5. Li, L. & Guo, X. AFLP-based genetic linkage maps of the Pacific oyster *Crassostrea gigas* Thunberg. *Marine Biotechnology* **6**, 26–36 (2004).
6. Pérez, F., Erazo, C., Zhinaula, M., Volckaert, F. & Calderón, J. A sex-specific linkage map of the white shrimp *Penaeus (Litopenaeus) vannamei* based on AFLP markers. *Aquaculture* **242**, 105–118 (2004).
7. Sun, X., Habier, D., Fernando, R. L. *et al.* Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian methods. *BMC proceedings* **5**, (Suppl 3): S13, doi: 10.1186/1753-6561-5-S3-S13 (2011).
8. Wang, S. *et al.* Quality assessment parameters for EST-derived SNPs from catfish. *BMC genomics* **9**, 450 (2008).
9. Baxter, S. W. *et al.* Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS one* **6**, doi: 10.1371/journal.pone.0019315c (2011).
10. Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A. & Johnson, E. A. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome research* **17**, 240–248 (2007).
11. Wang, S., Meyer, E., McKay, J. K. & Matz, M. V. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* **9**, 808–810 (2012).
12. Jiao, W. *et al.* High-resolution linkage and quantitative trait locus mapping aided by genome survey sequencing: building up an integrative genomic framework for a bivalve mollusc. *DNA research* **21**, 85–101 (2014).
13. Lowry, D. B. *et al.* The genetics of divergence and reproductive isolation between ecotypes of *Panicum hallii*. *New Phytologist* **205**, 402–414 (2015).
14. Cui, Z. *et al.* High-density linkage mapping aided by transcriptomics documents ZW sex determination system in the Chinese mitten crab *Eriocheir sinensis*. *Heredity*, doi: 10.1038/hdy.2015.26 (2015).
15. Shi, Y. *et al.* High-density single nucleotide polymorphisms linkage and quantitative trait locus mapping of the pearl oyster, *Pinctada fucata martensii* Dunker. *Aquaculture* **434**, 376–384 (2014).
16. Sprinkle, J. *Origin and early evolution of the Metazoa* [Lipps, Jere H., Signor, Philip W. (eds.)] [375–398] (Springer, 1992).
17. Blunt, J. W., Copp, B. R., Munro, M. H., Northcote, P. T. & Prinsep, M. R. Review: marine natural products. *Natural Product Reports*, **28**(2), 196–268 (2011).
18. Carnevali, M. & Burighel, P. Regeneration in echinoderms and ascidians. *eLS*, doi: 10.1002/9780470015902.a0022102 (2010).
19. Signor, III, P. W. & Brett, C. E. The mid-Paleozoic precursor to the Mesozoic marine revolution. *Paleobiology*, 229–245 (1984).
20. Clarkson, E. *Invertebrate palaeontology and evolution* [Clarkson, E. (ed.)] [311–314] (John Wiley & Sons, 2009).
21. Árni, M. M. *Fishery Statistics Summary tables (B-1)*. *FAO Yearbook. Fishery and Aquaculture Statistics*. (2014). Available at: <ftp://ftp.fao.org/FI/STAT/summary/b-1.pdf>. (Accessed: 30th June 2014).
22. Zhou, Z. *et al.* Transcriptome sequencing of sea cucumber (*Apostichopus japonicus*) and the identification of gene-associated markers. *Molecular ecology resources* **14**, 127–138 (2014).
23. Yan, J. *et al.* A genetic linkage map of the sea cucumber (*Apostichopus japonicus*) based on microsatellites and SNPs. *Aquaculture* **404**, 1–7 (2013).
24. Zhou, Z. *et al.* AFLP linkage map of sea urchin constructed using an interspecific cross between *Strongylocentrotus nudus* (♀) and *S. intermedius* (♂). *Aquaculture* **259**, 56–65 (2006).
25. Sloan, N. Echinoderm fisheries of the world: a review. *AA Balkema, Rotterdam* 109–124 (1985).
26. Chen, J. Present status and prospects of sea cucumber industry in China. *Ile* **1**, 25–38 (2005).
27. He, C. & Huang, G. On *Apostichopus japonicus* culture in China and major culture provinces. *Modern Fisheries Information* **29**(1), doi: 10.3969/j.issn.1004-8340.2014.01.005 (2014).
28. Li, Q., Chen, L. & Kong, L. A genetic linkage map of the sea cucumber, *Apostichopus japonicus* (Selenka), based on AFLP and microsatellite markers. *Animal genetics* **40**, 678–685 (2009).
29. Du, H. *et al.* Transcriptome sequencing and characterization for the sea cucumber *Apostichopus japonicus* (Selenka, 1867). *PLoS one* **7**, doi: 10.1371/journal.pone.0033311 (2012).
30. Fu, X. *et al.* RADtyping: an integrated package for accurate de novo codominant and dominant RAD genotyping in mapping populations. *PLoS one* **8**, doi: 10.1371/journal.pone.0079960 (2013).
31. Van Ooijen, J. JoinMap<sup>®</sup> 4, Software for the calculation of genetic linkage maps in experimental populations. *Kyazma BV, Wageningen* **33**, doi: 10.1371/journal.pone.0098628 (2006).
32. Okumura, S.-I. *et al.* Chromosome number and telomere sequence mapping of the Japanese sea cucumber *Apostichopus japonicus*. *Fisheries Science* **75**, 249–251 (2009).
33. Mohideen, M.-A. P., Moore, J. L. & Cheng, K. C. Centromere-Linked Microsatellite Markers for Linkage Groups 3, 4, 6, 7, 13, and 20 of Zebrafish (*Danio rerio*). *Genomics* **67**, 102–106 (2000).
34. Sakamoto, T. *et al.* A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* **155**, 1331–1345 (2000).
35. Chakravarti, A., Lasher, L. K. & Reefer, J. E. A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* **128**, 175–182 (1991).
36. Van Ooijen, J. MapQTL<sup>®</sup> 5, Software for the mapping of quantitative trait loci in experimental populations. *Kyazma BV, Wageningen* **63** (2004).
37. Bradbury, P. J. *et al.* TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
38. Sun, L., Craiu, R. V., Paterson, A. D. & Bull, S. B. Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. *Genetic epidemiology* **30**, 519–530 (2006).
39. Park, J. A. *et al.* Retinoblastoma protein regulates cell proliferation, differentiation, and endoreduplication in plants. *The Plant Journal* **42**, 153–163 (2005).
40. Slate, J. *et al.* Gene mapping in the wild with SNPs: guidelines and future directions. *Genetica* **136**, 97–107 (2009).
41. Heller-Uszynska, K. *et al.* Diversity arrays technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Molecular breeding* **28**, 37–55 (2011).
42. Liu, J. *et al.* Genome size determination of sea cucumber (*Apostichopus japonicus*). *Journal of Fisheries of China* **5**, 008 (2012).
43. Hedgecock, D. *et al.* The case for sequencing the Pacific oyster genome. *Journal of Shellfish research* **24**, 429–441 (2005).
44. Sodergren, E. *et al.* The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**, 941–952 (2006).
45. Zhang, G. *et al.* The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**, 49–54 (2012).
46. Borevitz, J. O. *et al.* Large-scale identification of single-feature polymorphisms in complex genomes. *Genome research* **13**, 513–523 (2003).
47. Zhang, L. *et al.* Fosmid library construction and initial analysis of end sequences in Zhikong scallop (*Chlamys farreri*). *Marine biotechnology* **9**, 606–612 (2007).

48. Grattapaglia, D. & Sederoff, R. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* **137**, 1121–1137 (1994).
49. Saijo, M. *et al.* Molecular cloning of a human protein that binds to the retinoblastoma protein and chromosomal mapping. *Genomics* **27**, 511–519 (1995).
50. Lawrence, H. J. *et al.* Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. *Blood* **106**, 3988–3994 (2005).
51. Crawford, B. D. & Hess, J. L. MLL core components give the green light to histone methylation. *ACS chemical biology* **1**, 495–498 (2006).
52. Zhan, A. *et al.* Development and characterization of 45 novel microsatellite markers for sea cucumber (*Apostichopus japonicus*). *Molecular Ecology Notes* **7**, 1345–1348 (2007).
53. Dou, J. *et al.* Reference-free SNP calling: improved accuracy by preventing incorrect calls from repetitive genomic regions. *Biology direct* **7**, 17 (2012).
54. Wu, Y., Close, T. J. & Lonardi, S. On the accurate construction of consensus genetic maps. *Comput Syst Bioinformatics Conf.* **7**, 285–296 (2008).
55. Voorrips, R. MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**, 77–78 (2002).
56. Li, H. A quick method to calculate QTL confidence interval. *Journal of genetics* **90**, 355 (2011).
57. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic acids research* **32**, W309–W312 (2004).

## Acknowledgments

This project was supported by National High Technology Research and Development Program of China (863 Program, 2012AA10A412), and National Key Technology R&D Program of China (2011BAD13B06). The authors wish to thank Lingling Zhang, Yun Li, Tianqi Li, Jingjing Yan, and Liheng Zhang for their technical assistance in various experiments.

## Author Contributions

M.T., Y.L. and J.J. conducted the major part of the research including preparation of the 2bRAD-seq data, bioinformatics analysis and manuscript preparation. C.M., H.D., J.D., J.M., X.L., W.J., Y.W. and X.H. were involved in one or more processes of DNA extraction or bioinformatics analysis. R.W. and S.W. conceived, designed and guided the research and involved in manuscript preparation. All authors read and approved the final manuscript.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Tian, M. *et al.* Construction of a High-Density Genetic Map and Quantitative Trait Locus Mapping in the Sea Cucumber *Apostichopus japonicus*. *Sci. Rep.* **5**, 14852; doi: 10.1038/srep14852 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>