A Male-Specific Genetic Map of the Microcrustacean *Daphnia pulex* Based on Single-Sperm Whole-Genome Sequencing

Sen Xu,*,[1] Matthew S. Ackerman,* Hongan Long,* Lydia Bright,* Ken Spitze,* Jordan S. Ramsdell,[†]
W. Kelley Thomas,[†] and Michael Lynch*

*Department of Biology, Indiana University, Bloomington, Indiana 47405, and [†]Hubbard Center for Genome Studies and Department of Molecular Cellular and Biomedical Sciences, University of New Hampshire, Durham, New Hampshire 03824

ABSTRACT Genetic linkage maps are critical for assembling draft genomes to a meaningful chromosome level and for deciphering the genomic underpinnings of biological traits. The estimates of recombination rates derived from genetic maps also play an important role in understanding multiple aspects of genomic evolution such as nucleotide substitution patterns and accumulation of deleterious mutations. In this study, we developed a high-throughput experimental approach that combines fluorescence-activated cell sorting, whole-genome amplification, and short-read sequencing to construct a genetic map using single-sperm cells. Furthermore, a computational algorithm was developed to analyze single-sperm whole-genome sequencing data for map construction. These methods allowed us to rapidly build a male-specific genetic map for the freshwater microcrustacean *Daphnia pulex*, which shows significant improvements compared to a previous map. With a total of mapped 1672 haplotype blocks and an average intermarker distance of 0.87 cM, this map spans a total genetic distance of 1451 Kosambi cM and comprises 90% of the resolved regions in the current *Daphnia* reference assembly. The map also reveals the mistaken mapping of seven scaffolds in the reference assembly onto chromosome II by a previous microsatellite map based on F_2 crosses. Our approach can be easily applied to many other organisms and holds great promise for unveiling the intragenomic and intraspecific variation in the recombination rates.

KEYWORDS meiosis; fluorescence-activated cell sorting; single cell; whole-genome amplification

CONSTRUCTING a genetic linkage map that encompasses as much genomic sequence as possible is critical for current endeavors of *de novo* genomic assembly (*e.g.*, Kawakami *et al.* 2014; International Cassava Genetic Map Consortium (ICGMC) 2015) and for deciphering the genomic underpinnings of the biological traits in the species of interest (Lynch and Walsh 1998). Genetic maps can be utilized in several ways to help achieve these goals. At the very least, a linkage map with an adequate number of genetic markers can serve as the backbone for orienting and assembling segments of DNA (*i.e.*, scaffolds) into chromosomes. The developed genetic markers can also be used in QTL association mapping and genomic-scanning efforts to investigate genetic loci underlying ecological tolerance, adaptation, and disease. Furthermore, a linkage map provides genome-wide estimates of the meiotic recombination rate, which plays a significant role in the distribution of genetic diversity (Nachman 2001), rate of adaptation (Bachtrog and Charlesworth 2002), accumulation of deleterious mutations (Hussin *et al.* 2015), and nucleotide substitution (Duret and Arndt 2008).

Currently, the most common approach for constructing a genetic linkage map is based on genotyping a large number of molecular markers (*e.g.*, SNPs, microsatellites) from a large number of offspring (usually on the order of hundreds) derived from various kinds of crossing schemes (*e.g.*, backcrosses, F_2s, recombinant inbred lines) or from family trios to estimate the frequencies of recombination between markers across the genome. Although most estimates of recombination rates in a diverse range of taxa come from studies based on this idea, crossing experiments are laborious, often inefficient (*e.g.*, low hatching/survival rate of progeny), and unattainable in cases

where manipulative crossing is impossible. Furthermore, this approach may introduce biases in estimating offspring genotype frequencies if certain classes of recombinant genotypes cause lethality and/or low viability.

With whole-genome sequencing becoming increasingly economical, interpretation of population-genomic data in a coalescence framework can generate estimates of historical recombination (*i.e.*, crossover) rates needed for creating genetic maps (Stumpf and McVean 2003; McVean *et al.* 2004). This approach is gaining popularity and has been successfully applied to many organisms, including human (McVean *et al.* 2004; Myers *et al.* 2005), chimpanzee (Auton *et al.* 2012), dogs (Auton *et al.* 2013), and *Arabidopsis* (Choi *et al.* 2013), uncovering genetic factors that regulate crossover events such as *PRDM9* in humans (Myers *et al.* 2010) and H2A.Z nucleosomes at promoters in *Arabidopsis* (Choi *et al.* 2013). Nonetheless, one prerequisite for applying this method is the availability of a genome assembly with reasonable quality for the species of interest or for a closely related species. This is because crossover events are inferred on the basis of patterns of linkage disequilibrium in the population of interest, which can be severely distorted if the mapping of short-sequence reads is performed on an assembly where physical ordering of genomic sites is erroneous. It should also be noted that the linkage-disequilibrium method can yield only estimates for historical population-level and sex-averaged crossover rates. Thus, this method is not suitable for examining the variation of recombination rate between individuals and sexes (*e.g.*, Coop and Przeworski 2007; Kong *et al.* 2010; Brandvain and Coop 2012; Comeron *et al.* 2012; Bauer *et al.* 2013). Furthermore, it is well known that crossover, the reciprocal exchange of DNA between homologous chromosomes, only represents a small proportion (often <20%) of the total recombination events, with the rest generating nonreciprocal, gene-conversion events (Langley *et al.* 2000; Malkova *et al.* 2004; Morrell *et al.* 2006; Mancera *et al.* 2008; Yang *et al.* 2012; Lynch *et al.* 2014).

The most recent development in linkage map construction methodologies is the deployment of single-cell whole-genome sequencing on gametes such as sperm (Lu *et al.* 2012; Wang *et al.* 2012). Gametes are products of meiosis, bearing the genetic signature of meiotic recombination. Directly examining the haplotypes of gametes can offer an unbiased view of the recombination process and potentially yields insight into genetic factors regulating meiotic recombination events. The idea of utilizing single sperm to examine the recombination rate between a few markers on a small genomic scale coincided with the time when PCR (polymerase chain reaction) technology started revolutionizing the field of molecular biology in the 1980s (Li *et al.* 1988; Cui *et al.* 1989). However, this approach has not gained popularity until recently, mainly because of two technical hurdles, *i.e.*, the rapid isolation of a large number of single gametes and the low amount of DNA present in each gamete. Problems inherent in sorting a large number of in-dividual gametes made it implausible to gain enough power to estimate recombinant frequencies and disentangle individual gamete recombination patterns. Additionally, the extremely low amount of DNA present in each gamete (e.g., ∼3 pg of DNA in a single human sperm) prevented genotyping a sufficient number of genetic markers. Nonetheless, technological advances in microfluidics and flow cytometry have made it feasible to rapidly isolate single cells (Wang *et al.* 2012), and whole-genome amplification techniques now can propagate a single-copy genome to an amount of DNA that allows subsequent whole-genome sequencing/genotyping (Pan *et al.* 2008; Zong *et al.* 2012). Given the abundance of sperm cells, the combination of these techniques with whole-genome sequencing has led to the successful construction of a marker-dense male-specific genetic map for humans (Lu *et al.* 2012; Wang *et al.* 2012). Nonetheless, to our knowledge this single-sperm whole-genome sequencing approach has not been applied to other organisms.

In this study, we developed a high-throughput experimental workflow that isolates, whole-genome amplifies, and sequences single sperm from the North American microcrustacean *Daphnia pulex* (Crustacea, Anomopoda) to build a genetic map. *Daphnia* are keystone zooplankton species in global freshwater ecosystems (*e.g.*, lakes and woodland ponds) and are model organisms for toxicological, evolutionary, and biomedical research (Colbourne *et al.* 2011). *Daphnia* typically reproduce by cyclical parthenogenesis. Under good environmental conditions, females produce directly developing and genetically identical daughters. However, with unfavorable conditions (*e.g.*, food shortage), males are produced by environmental sex determination and engage in sexual reproduction to produce diapausing embryos. A major motivation for us to develop single-sperm sequencing for building a genetic map in *D. pulex* is the difficulty in efficiently hatching the diapausing embryos needed to form an offspring panel from crossing experiments (*e.g.*, Cristescu *et al.* 2006).

Although there exists a microsatellite-based genetic map for the current *Daphnia* genome assembly, this map anchors only 73 scaffolds, which accounts for only 73.9 of the 200 Mbp *Daphnia* genome (Cristescu *et al.* 2006; Colbourne *et al.* 2011). Furthermore, there is a need for a genetic map specifically for *D. pulex* because the current *Daphnia* assembly and genetic map are both derived from *Daphnia arenata*, an endemic species to Oregon rather than *D. pulex* that occurs in most of the temperate regions in North America. Although *D. arenata* and *D. pulex* are nearly indistinguishable with respect to morphology, *D. arenata* shows significantly reduced heterozygosity relative to *D. pulex*, with a pairwise nucleotide diversity of 0.0013 *vs.* 0.0119 in *D. pulex* based on six protein-coding loci (Omilian and Lynch 2009). Although *D. arenata* is still paraphyletic with respect to *D. pulex* for the mitochondrial genome (Colbourne *et al.* 1998; Lynch *et al.* 2008), these two species display a nuclear genomic divergence of ∼2% (Tucker *et al.* 2013). Nonetheless,

it remains largely an open question whether the nuclear genetic divergence between these two species involve any genetic map changes, *i.e.*, chromosomal rearrangements such as inversions. Comparative analysis of the genetic maps of these two species will offer insight into the process of their genetic divergence and speciation.

The experimental procedure presented here employs fluorescence-activated cell sorting (FACS), the whole-genome amplification technique multiple annealing and looping-based amplification cycles (MALBAC) (Zong *et al.* 2012), and the Illumina Hi-Sequation 2500 sequencing platform. Because this procedure relies on equipment such as flow cytometers and standard PCR thermal cyclers that are readily available to the research community, it can be easily applied and/or adapted to many other organisms with or without existing reference genomic assemblies to rapidly generate genetic maps. Moreover, we provide a computer program for extraction of haplotype blocks that are free of recombination events (*i.e.*, crossover and gene conversion) to build a male-specific genetic map in combination with the genetic map software MSTMap (Wu *et al.* 2008). This method constitutes a valuable approach for future studies that examine sex-specific genetic maps, fine-scale recombination patterns, and individual recombination variation.

Materials and Methods

Daphnia culture

The *Daphnia* isolate (PA42) used in this study was sampled from Portland Arch (latitude: 40°13, longitude: −87°20), Indiana, in May 2013. This isolate, which reproduces by cyclical parthenogenesis, was maintained under benign laboratory conditions at 20° and fed *ad libitum* with a suspension of *Scenedesmus obliquus* to enable essentially indefinite parthenogenetic reproduction.

Single-sperm isolation

We collected 15 mature, parthenogenetically produced males from the mass culture of the PA42 isolate. These males are genetically identical except for *de novo* mutations, which can be safely ignored given their low frequency (on the order of 10^{-9} events/base/generation). Sperm was collected from each male by squeezing the abdominal part of the individual in a drop of ultrapure distilled H_2O under a cover slip. The presence of sperm was confirmed by examination under microscope. The pooled collection of sperm was transferred to 50 μl PBS buffer and stained using Hoechst 33528 (100 μg/μl, Sigma-Aldrich), a dye that binds to double-strand DNA. Then, we used a FACS Aria II SORP Flow Cytometer (BD Biosciences) to isolate single sperm. Lasers used were a 488 nm 100 mW for light scatter detection and a 355 nm 20 mW for Hoechst detection. A FSC–PMT was used for optimal small particle discrimination. A 70-μm nozzle was used at 45 psi. Sperm cells were dispensed into regular 96-well PCR plates. Each sperm was deposited into 5 μl cell lysis buffer (30 mM Tris, 2 mM EDTA, 20 mM KCl, 0.2% Triton-X100, 50 mM DTT, and

500 μm/ml protease) and lysed for 3 hr at 50°, 20 min at 75°, and 5 min at 80°.

Whole-genome amplification and sequencing

We whole-genome amplified 104 single-sperm cells following the MALBAC protocol (Zong *et al.* 2012). In brief, MALBAC consists of a preamplification stage and a second-stage PCR amplification. In the preamplification cycles, a pair of quasi-degenerate primers is used to initiate overlapped amplicons throughout the whole genome. The quasi-degenerate primers consist of a 27-bp fragment of 5′-GTGAGTGATGGTTGAGG TAGTGTGGAG-3′ and eight variable nucleotides attached to the 3′ end. For each reaction, 3 μl ThermPol buffer (New England Biolabs), 1 μl dNTP (10 mM), 21 μl H_2O, and 0.15 μl primers (50 μM) were added. Each sample is then denatured at 94° for 3 min and quenched on ice immediately. With samples staying on ice, 0.6 μl Bst large fragment (New England Biolabs) is added to each reaction. The following thermal regime is used to generate random amplicons across the genome: 10° for 45 sec, 15° for 45 sec, 20° for 45 sec, 30° for 45 sec, 40° for 45 sec, 50° for 45 sec, 65° for 2 min, 95° for 20 sec, followed by quenching on ice and adding 0.6 μl *Bst* large fragment to each sample. Subsequently, the samples are subject to five rounds of amplification, each of which consists of 10° for 45 sec, 15° for 45 sec, 20° for 45 sec, 30° for 45 sec, 40° for 45 sec, 50° for 45 sec, 65° for 2 min, 95° for 20 sec, 58° for at least 20 sec, followed by quenching on ice and adding 0.6 μl *Bst* large fragment to each sample. After the preamplification stage, a standard PCR amplification is performed with each sample to generate 1–2 μg DNA to be used for downstream applications. Each reaction consists of 3 μl ThermoPol Buffer (New England Biolabs), 1 μl dNTP (10 mM), 26 μl H_2O, 0.15 μl primer 5′-GTGAGTGATGGTTGAGGTAGTGTGGAG-3′ (100 mM), and 1 μl DeepVentR exo- (New England Biolabs). The PCR thermal regime consists of 22 rounds of 94° for 20 sec, 59° for 20 sec, 65° for 1 min, 72° for 2 min, followed by 72° for 5 min.

To eliminate the possibility of samples containing multiple sperm cells, 12 microsatellite markers, each from 1 of the 12 chromosomes in the *Daphnia* genome, were genotyped using an ABI 3730 genetic analyzer (Life Technologies). The allele sizes of the genotyped microsatellite loci were analyzed using Genemapper software 4.0 (Life Technologies). None of the samples presented evidence of multiple cells, *i.e.*, more than one allele for the whole suite of loci.

Subsequently, the whole-genome-amplified DNA of each sperm was sheared to an average fragment size of 350 bp on a Covaris S2 shearing machine. Short-read sequencing library preparation followed the standard protocol of Illumina and was done by the Center for Genomics and Bioinformatics at Indiana University, Bloomington. Short-read sequencing was performed on an Illumina HiSeq2500 platform with 150-bp paired end reads.

Construction of genetic maps

Our approach can be applied to organisms with or without an existing reference assembly. This is because the

whole-genome sequences of all sperm samples can be used for creating a *de novo* assembly, although the completeness of the assembly depends on the whole-genome amplification coverage across the entire genome. To demonstrate this, we built a *de novo* sperm reference assembly using the assembler Platanus (Kajitani *et al.* 2014) based on the pooled whole-genome sequences of all sperm. The pooled sequences for *de novo* assembly were normalized using the software BBNorm (http://sourceforge.net/projects/bbmap/) to reduce the redundancy and remove errors of raw reads. We used the default settings in Platanus for building contigs and scaffolds. All scaffolds that were possibly from contaminant DNA (*e.g.*, bacteria, algae, and human) or shorter than 1000 bp were removed from the final sperm reference assembly.

The 27-bp primer sequences for the whole-genome amplification reactions were computationally removed from the ends of raw reads when present using the software CLC Genomics Workbench (v. 7, CLC Bio). The processed raw reads for each sample were mapped to the *Daphnia* reference assembly (Colbourne *et al.* 2011) and the sperm reference assembly using the short-read mapping function implemented in CLC Genomics Workbench with default settings. However, reads mapped to multiple locations were removed from further analysis. The haplotype for each position of a single sperm was determined using a consensus approach, where a base call is made with the support of >80% of the reads. To avoid sequencing errors, PCR artifacts, and potential mapping errors, we also require at least two forward and two reverse reads to validate the consensus call. Because only heterozygous loci are informative for analyzing recombination events, only sites where two nucleotides were found across the entire set of sperm samples were kept.

We developed an algorithm implemented in Python (Supporting Information, File S5, phasingHaplotype.py) to detect haplotype blocks that are free of recombination events to be used as markers for genetic map construction. Because this set of sperm is derived from the recombination of two parental haplotypes, we randomly assign either 0 or 1 (designating the two parental haplotypes) to the same two-locus haplotype for the first pair of sites on each scaffold (Figure 1B). Every pair of sites across the samples has a maximum of four haplotypes when a crossover or gene conversion event happens, three haplotypes for biased gene conversion events, and two haplotypes for no recombination. We then consecutively examine each pair of sites and extend the haplotype assignment. A switching of phase occurs only when three or four haplotypes evidently exists.

Once the haplotype-phase assignment was done, we selected haplotype blocks (a minimum of two SNP sites) free of evident recombination events in the set of sequenced sperm samples as genetic markers for genetic map construction. The selected haplotype blocks were coded as either 0 or 1. We used the software MSTMap (Wu *et al.* 2008) to construct the linkage map with its default settings. MSTMap implements an efficient algorithm to determine the correct order of a large number of genetic markers (10,000–100,000) by computing the minimum spanning tree. MSTMap is significantly better at recovering the correct order of markers from noisy data compared to most other software, which is helpful for dealing with the missing data in our data set.

MareyMap database for estimating recombination rate

Integrating the genetic map data into a physical map allows the estimation of recombination rate for genomic regions along a chromosome using different curve fitting methods (Rezvoy *et al.* 2007). This is the so-called Marey map method (Chakravarti 1991). We combined the sperm genetic map data with the current *Daphnia* reference assembly (Colbourne *et al.* 2011) and created a database (File S3) using the R package MareyMap (Rezvoy *et al.* 2007), which allows users to estimate recombination rates for most of the regions on major scaffolds in the current assembly.

Data availability

The binary alignment files for sperm samples were deposited at NCBI Sequence Read Archive under study no. SRP058678.

Results

Single-cell genome sequencing, mapping, and SNP density

We sequenced the whole genomes of 104 single sperm (Figure 1A) from the *D. pulex* isolate PA42 from Portland Arch, Indiana, with 150-bp paired end reads. After trimming the PCR and Illumina adapter sequences from the raw reads and removing reads that mapped to multiple locations in the *Daphnia* reference assembly (Colbourne *et al.* 2011), we found that the aligned reads covered on average 77.3 Mbp ± 14.7 (SD) of the 200-Mbp *Daphnia* genome, which is equivalent to 52.8% of the resolved regions in the reference assembly. The average coverage per site per sperm sample is 12.5 ± 4.5 (SD).

We identified a total of 1,537,288 heterozygous sites from the parental diploid genotype PA42 following a set of stringent criteria (see *Materials and Methods*). With a 2% estimated heterozygosity based on whole-genome sequencing of the PA42 genotype, these recovered sites account for 38.4% of the total heterozygous sites expected in the genome.

Furthermore, we built a *de novo* assembly using the pooled sperm genomic sequences. The sperm assembly spans 109 Mbp with 1.7% gap regions and consists of 32,549 scaffolds. The largest scaffold size is 90,435 bp, whereas the smallest scaffold is 1001 bp. After mapping the processed raw reads for each sperm to the sperm assembly, the aligned reads covered on average 49.4 Mbp ± 11.9 of the assembly. The observation that the average coverage is lower than the total assembly length is mainly because the sperm assembly is based on the pooled sequence of all samples and each whole-genome amplified sperm sample does not contain all sequences in the assembly.
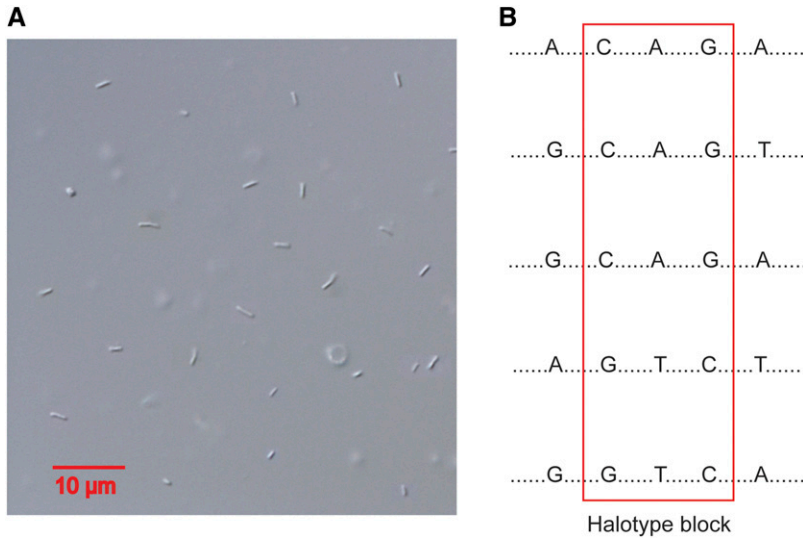
Figure 1 (A) Image of sperm extracted from males of the *D. pulex* PA42 isolate from Portland Arch, Indiana. The sperm is rod shaped, with a length ∼2 μm. (B) A hypothetical example of haplotype block indicated by the red box. Each haplotype is represented by five nucleotide sites. A haplotype block is identified where only two haplotypes occur across the whole set of samples.

Genetic linkage map

For the analyses based on the *Daphnia* reference genomic assembly (Colbourne *et al.* 2011), we selected a total of 1672 marker regions where at least two consecutive SNP loci show the same haplotype in >90 sperm samples; as they contained zero recombination events, such spans serve as single markers in the map construction. The lengths of the marker regions range from 50 to 269,519 bp, with a mean of 22,771 $\pm$ 21,710 (SD). The total length of the marker regions is ∼38.1 Mbp, comprising 19.1% of the ∼200-Mbp *Daphnia* genome. Based on these markers, we constructed a male-specific *Daphnia* linkage map using the software MSTmap (Wu *et al.* 2008). This genetic map consists of 12 linkage groups, corresponding to the 12 chromosomes in *Daphnia* genome (Zaffagnini and Sabelli 1972), and spans a total genetic distance of 1451 Kosambi cM (Figure S1), with an average intermarker distance of 0.87 cM. The number of haplotype blocks on each linkage group ranges between 67 and 202, with a mean of 139. The map distance for each linkage group varies between 81 and 149 Kosambi cM, with a mean of 121 $\pm$ 22 (SD) cM. This map anchors to chromosomes a total of 187 of 5191 scaffolds from the *Daphnia* genome reference assembly (Colbourne *et al.* 2011). These scaffolds encompass 131.9 Mbp of DNA sequence, which is equivalent to 90.0% of the resolved portion of current assembly.

To demonstrate that our approach can be used for organisms without preexisting reference assembly, we built a genetic map based on sperm *de novo* assembly. We recovered 12 linkage groups using 350 haplotype blocks (File S4). The total length of the 350 haplotype blocks is 2.36 Mbp, corresponding to 2.2% of the sperm assembly. The lengths of the haplotype blocks range from 50 to 47,702 bp, with a mean of 6742 bp. The total genetic map distance is 823 Kosambi cM. The map distance for each lineage group ranges between 23 and 137 cM, with a mean of 67 $\pm$ 34 (SD) cM. The average intermarker distance is 2.35 cM. Furthermore, this map anchors in total 343 scaffolds into the 12 linkage groups.

Differences of scaffold assignment between maps of D. pulex and D. arenata

We compared the male-specific genetic map of *D. pulex* with the prior microsatellite map of *D. arenata* to detect possible chromosomal rearrangements. Because all of the microsatellite markers in *D. arenata* have known physical locations in the reference assembly, we were able to compare the differences in anchored locations of scaffolds. Most notably, a large chunk of chromosome II spanning from 0 to 29.9 cM in *D. arenata* maps to chromosome XII in *D. pulex*; this map difference involves seven scaffolds (5, 27, 58, 63, 70, 84, and 86). Furthermore, we detected numerous cases in which segments of one scaffold in the current reference assembly do not form a continuous tract on the genetic map but are split by segments from other scaffolds (Figure S1), which indicates problematic genomic assembly or possible rearrangements. An example of this category of observations is that one portion of scaffold 260,935–1,006,942 bp maps to chromosome XI, while the rest of scaffold 8 is anchored to chromosome IV (File S1). In addition to these cases of split scaffolds, in many of the largest scaffolds (*e.g.*, scaffolds 1, 2, 3) in the current *Daphnia* assembly, physical orders of haplotype blocks do not agree with the genetic map (File S1). For example, scaffold 1: 260,935–1,006,942 bp is mapped on chromosome II with a map distance from 42.68 to 44.35 cM, whereas the physically downstream segment scaffold 1 3806402–3934086 is mapped from 34.34 to 42.11 cM on the same chromosome.

Discussion

Combining single-cell isolation based on flow cytometry, whole-genome amplification, and short-read sequencing, we established a rapid experimental workflow for constructing a genetic map in the microcrustacean *D. pulex* using single

sperm. Compared to the extensive time (*e.g.*, weeks or months) required by conventional crossing experiments, this procedure can gather the data needed for a map in a time frame of a few days. Unlike the population-genomic sequencing approach that requires a reference genomic assembly (*e.g.*, Auton *et al.* 2012), the sperm-based method does not necessarily rely on a preexisting reference assembly. Although we did use a *Daphnia* reference assembly to guide the identification of SNP sites in this study, we were also able to use spermwhole-genome sequences to create a *de novo* assembly as an alternative basis for downstream genetic linkage map construction. Our results show that although this map is less marker dense than the one generated with the existing *Daphnia* reference assembly, the correct number of linkage groups is recovered. More importantly, it should be noted that this represents an extreme situation in which there are no previous genomic resources at all. As a result, the quality of the genetic map can be significantly improved by incorporating a few more paired-end and mate-pair sequencing libraries into the *de novo* assembly process. The discussion below focuses only on the genetic map generated using the existing *Daphnia* genomic reference assembly.

Our experimental workflow heavily relies on FACS for isolating single sperm. There are a few other alternative approaches for isolating single cells such as manual micromanipulation and laser dissection (Macaulay and Voet 2014). However, to achieve accurate isolation of single cells in a high-throughput manner, FACS provides a highly reliable platform with advantages that alternative approaches cannot offer. Considering that the small size of *Daphnia* sperm (a length of ∼2–3 μm, Figure 1) makes them indistinguishable from dust particles if they were sorted only by size on the flow cytometer, we stained sperm cells with a Hoechst dye, which binds to double-stranded DNA. Combining sorting by size and wavelength of fluorescence emission on the flow cytometer ensures highly accurate and rapid isolation of single sperm. For example, it takes only ∼1 hr to sort ∼1000 single sperm and the accuracy for single-cell isolation is 100% in our data set (see *Materials and Methods*).

A major concern with whole-genome amplifying an extremely small amount of DNA in a single cell is the uneven amplification on different regions of the genome, which can lead to biased genome-wide coverage (for a recent review see Macaulay and Voet 2014). A major reason that we chose MALBAC in our workflow is that this procedure limits the further amplification of the genomic regions that are amplified early on in the reaction, resulting in much improved coverage across the genome (Zong *et al.* 2012). Although the average breadth of coverage across the sperm samples is only 52.8% of the resolved region in the reference assembly, the mapped scaffolds cover 90.0% of the resolved regions. Therefore, with this efficient amplification procedure, our genetic map achieves the goal of encompassing as much actual genome sequence as possible.

Table 1 Comparison between the sperm-based genetic map for *D. pulex* and the microsatellite genetic map for *D. arenata* (Cristescu *et al.* 2006)

	D. pulex map	*D. arenata* map
Total map distance (Kosambi cM)	1451	1206
No. of markers	1672	185
No. of scaffolds mapped	187	73
Basepairs of genome mapped (Mbp)	131.9	73.9
Average intermarker distance (cM)	0.87	7

The statistics for the *D. arenata* map were compiled from Cristescu *et al.* (2006) and Colbourne *et al.* (2011).

These state-of-the-art technologies greatly aided our generation of an ultradense male-specific genetic map for *D. pulex* based on short-read sequencing, yielding a substantial improvement over the previous microsatellite-based *Daphnia* genetic map (Table 1), which actually required substantially more work. The total genetic map distance of the new map is 1451 Kosambi cM, slightly greater than that of the *D. arenata* map (1206 cM) and similar to *D. magna* (1483 cM; Routtu *et al.* 2014). This map anchors to chromosomes 187 scaffolds (131.9 Mbp), in comparison to 73 anchored scaffolds (73.9 Mb) from the *D. arenata* map. Furthermore, the new map provides much refined estimates of the recombination rate between markers (0.87 *vs.* 7 cM), allowing the mapping of 90% of the *D. pulex* genome to on average <1 cM to the nearest genetic marker. This intermarker distance is also smaller than that in the SNP-based *D. magna* genetic map (1.13 cM; Routtu *et al.* 2014). Therefore, this map provides a framework to examine the role of recombination rate in shaping the various aspects of the genomic architecture of *Daphnia* genome such as patterns of nucleotide substitution and codon usage, which to date have not been explored in detail.

Another goal of this study was to examine whether chromosomal rearrangement is involved in the divergence of *D. pulex* and *D. arenata*. It should be noted that genomic assemblies are not perfect and could contain misassembled regions (*e.g.*, physically distant regions assembled adjacent to each other). Misassembled regions can disrupt the map orders of nearby markers, leading to false conclusions about true genomic rearrangements. In the current map, we observed many cases of split scaffolds and discrepant intrascaffold orderings of markers between genetic and physical maps. Unfortunately, we are not able to distinguish between these assembly errors and true rearrangements for these problematic genomic areas (Figure S1). Nonetheless, given the numerous occurrences of these observations in many different parts of the reference assembly, which was built with only 8.7× coverage of Sanger sequencing reads and with little data from long insert mate-pair libraries (Colbourne *et al.* 2011), our observations of aberrant mappings are more likely to reflect assembly errors than true rearrangements.

The most notable difference between the *D. arenata* and *D. pulex* map involves changes between chromosome II and

chromosome XII. The scaffolds mapped to between 0 and 29.9 cM on chromosome II in *D. arenata* are mapped onto chromosome XII in *D. pulex*. Although this may potentially represent an interesting case of chromosomal translocation, a few observations collectively suggest that this is an error from the previous mapping study. First, this map interval in *D. arenata* was affected by segregation distortion (Cristescu *et al.* 2006), showing severe homozygote deficiency for the markers in this region. The map distance between this region and the closest genetic marker in the *D. arenata* map is 40.6 cM, which indicates its weak genetic linkage with the rest of the chromosome (because 50 cM map distance means free recombination). In fact, while the entire scaffold 5 in the reference assembly is unambiguously mapped to chromosome XII in the current map, the *D. arenata* genetic map shows that this scaffold is split between chromosome II and XII, indicating potential problems in assigning genetic markers to chromosomes.

Because of the high heterozygosity in the *Daphnia* genome and the great number of heterozygous sites recovered from single-sperm whole-genome sequencing, our data set offers a great opportunity to locate genomic intervals containing the breakpoints for crossover events and gene-conversion tracts. However, accomplishing such a task requires a reference assembly that embodies the correct physical order of all nucleotide sites in the genome. Because of the great number of potentially problematic assembled regions in the current *Daphnia* assembly, we are working on a *de novo* assembly for the PA42 *D. pulex* isolate using the sperm sequencing strategy in combination with a range of sequencing libraries with different insert sizes. In fact, there is a growing interest in using genome sequences of mapping panels to facilitate *de novo* genome assembly and alleviate problems of falsely assembling the two divergent alleles of the same locus into paralogous loci (Hahn *et al.* 2014). With the new genome assembly, we will hopefully gain sufficient power to reveal the genomic location of crossover and gene-conversion events and characterize the possible genetic elements that control the occurrence of recombination events.

Acknowledgments

Literature cited

Auton, A., A. Fledel-Alon, S. Pfeifer, O. Venn, L. Ségurel *et al.*, 2012 A fine-scale chimpanzee genetic map from population sequencing. Science 336: 193–198.

Auton, A., Y. R. Li, J. Kidd, K. Oliveira, J. Nadel *et al.*, 2013 Genetic recombination is targeted towards gene promoter regions in dogs. PLoS Genet. 9: e1003984.

Bachtrog, D., and B. Charlesworth, 2002 Reduced adaptation of a non-recombining neo-Y chromosome. Nature 416: 323–326.

Bauer, E., M. Falque, H. Walter, C. Bauland, C. Camisan *et al.*, 2013 Intraspecific variation of recombination rate in maize. Genome Biol. 14: R103.

Brandvain, Y., and G. Coop, 2012 Scrambling eggs: meiotic drive and the evolution of female recombination rates. Genetics 190: 709–723.

Chakravarti, A., 1991 A graphical representation of genetic and physical maps: the Marey Map. Genomics 11: 219–222.

Choi, K. H., X. H. Zhao, K. A. Kelly, O. Venn, J. D. Higgins *et al.*, 2013 *Arabidopsis* meiotic crossover hot spots overlap with H2A. Z nucleosomes at gene promoters. Nat. Genet. 45: 1327–1336.

Colbourne, J. K., T. J. Crease, L. J. Weider, P. D. N. Hebert, F. Dufresne *et al.*, 1998 Phylogenetics and evolution of a circumarctic species complex (Cladocera: *Daphnia pulex*). Biol. J. Linn. Soc. Lond. 65: 347–365.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker *et al.*, 2011 The ecoresponsive genome of *Daphnia pulex*. Science 331: 555–561.

Comeron, J. M., R. Ratnappan, and S. Bailin, 2012 The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 8: e1002905.

Coop, G., and M. Przeworski, 2007 An evolutionary view of human recombination. Nat. Rev. Genet. 8: 23–34.

Cristescu, M. E., J. K. Colbourne, J. Radivojac, and M. Lynch, 2006 A microsatellite-based genetic linkage map of the waterflea, *Daphnia pulex*: on the prospect of crustacean genomics. Genomics 88: 415–430.

Cui, X. F., H. H. Li, T. M. Goradia, K. Lange, H. H. Kazazian *et al.*, 1989 Single-sperm typing: determination of genetic distance between the G-gamma globin and parathyroid hormone loci by using the polymerase chain reaction and allele-specific oligomers. Proc. Natl. Acad. Sci. USA 86: 9389–9393.

Duret, L., and P. F. Arndt, 2008 The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet. 4: e1000071.

Hahn, M. W., S. V. Zhang, and L. C. Moyle, 2014 Sequencing, assembling, and correcting draft genomes using recombinant populations. G3 Genes Genomes Genetics 4: 669–679.

Hussin, J. G., A. Hodgkinson, Y. Idaghdour, J.-C. Grenier, J.-P. Goulet *et al.*, 2015 Recombination affects accumulation of damaging and disease-associated mutations in human populations. Nat. Genet. 47: 400–404.

International Cassava Genetic Map Consortium (ICGMC), 2015 High-resolution linkage map and chromosome-scale genome assembly for Cassava (*Manihot esculenta* Crantz) from 10 populations. G3 Genes Genomes Genetics 5: 133–144.

Kajitani, R., K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura *et al.*, 2014 Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. 24: 1384–1395.

Kawakami, T., L. Smeds, N. Backstrom, A. Husby, A. Qvarnstrom *et al.*, 2014 A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol. Ecol. 23: 4035–4058.

Kong, A., G. Thorleifsson, D. F. Gudbjartsson, G. Masson, A. Sigurdsson *et al.*, 2010 Fine-scale recombination rate differences between sexes, populations and individuals. Nature 467: 1099–1103.

Langley, C. H., B. P. Lazzaro, W. Phillips, E. Heikkinen, and J. M. Braverman, 2000 Linkage disequilibria and the site frequency

spectra in the su(s) and su(w(a)) regions of the *Drosophila melanogaster* X chromosome. Genetics 156: 1837–1852.

Li, H. H., U. B. Gyllensten, X. F. Cui, R. K. Saiki, H. A. Erlich *et al.*, 1988 Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335: 414–417.

Lu, S., C. Zong, W. Fan, M. Yang, J. Li *et al.*, 2012 Probing meiotic recombination and aneuploidy of single sperm cells by whole-genome sequencing. Science 338: 1627–1630.

Lynch, M., and B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.

Lynch, M., A. Seyfert, B. Eads, and E. Williams, 2008 Localization of the genetic determinants of meiosis suppression in *Daphnia pulex*. Genetics 180: 317–327.

Lynch, M., S. Xu, T. Maruki, X. Q. Jiang, P. Pfaffelhuber *et al.*, 2014 Genome-wide linkage-disequilibrium profiles from single individuals. Genetics 198: 269–281.

Macaulay, I. C., and T. Voet, 2014 Single cell genomics: advances and future perspectives. PLoS Genet. 10: e1004126.

Malkova, A., J. Swanson, M. German, J. H. McCusker, E. A. Housworth *et al.*, 2004 Gene conversion and crossing over along the 405-kb left arm of *Saccharomyces cerevisiae* chromosome VII. Genetics 168: 49–63.

Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. Nature 454: 479–485.

McVean, G. A. T., S. R. Myers, S. Hunt, P. Deloukas, D. R. Bentley *et al.*, 2004 The fine-scale structure of recombination rate variation in the human genome. Science 304: 581–584.

Morrell, P. L., D. M. Toleno, K. E. Lundy, and M. T. Clegg, 2006 Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. Genetics 173: 1705–1723.

Myers, S., L. Bottolo, C. Freeman, G. McVean, and P. Donnelly, 2005 A fine-scale map of recombination rates and hotspots across the human genome. Science 310: 321–324.

Myers, S., R. Bowden, A. Tumian, R. E. Bontrop, C. Freeman *et al.*, 2010 Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327: 876–879.

Nachman, M. W., 2001 Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 17: 481–485.

Omilian, A. R., and M. Lynch, 2009 Patterns of intraspecific DNA variation in the *Daphnia* nuclear genome. Genetics 182: 325–336.

Pan, X. H., A. E. Urban, D. Palejev, V. Schulz, F. Grubert *et al.*, 2008 A procedure for highly specific, sensitive, and unbiased whole-genome amplification. Proc. Natl. Acad. Sci. USA 105: 15499–15504.

Rezvoy, C., D. Charif, L. Gueguen, and G. A. B. Marais, 2007 MareyMap: an R-based tool with graphical interface for estimating recombination rates. Bioinformatics 23: 2188–2189.

Routtu, J., M. D. Hall, B. Albere, C. Beisel, R. D. Bergeron *et al.*, 2014 An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. BMC Genomics 15: 1033.

Stumpf, M. P. H., and G. A. T. McVean, 2003 Estimating recombination rates from population-genetic data. Nat. Rev. Genet. 4: 959–968.

Tucker, A. E., M. S. Ackerman, B. D. Eads, S. Xu, and M. Lynch, 2013 Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. Proc. Natl. Acad. Sci. USA 110: 15740–15745.

Wang, J. B., H. C. Fan, B. Behr, and S. R. Quake, 2012 Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. Cell 150: 402–412.

Wu, Y. H., P. R. Bhat, T. J. Close, and S. Lonardi, 2008 Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. 4: e1000212.

Yang, S. H., Y. Yuan, L. Wang, J. Li, W. Wang *et al.*, 2012 Great majority of recombination events in *Arabidopsis* are gene conversion events. Proc. Natl. Acad. Sci. USA 109: 20992–20997.

Zaffagnini, F., and B. Sabelli, 1972 Karyologic observations on the maturation of the summer and winter eggs of *Daphnia pulex* and *Daphnia middendorffiana*. Chromosoma 36: 193–203.

Zong, C., S. Lu, A. R. Chapman, and X. S. Xie, 2012 Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. Science 338: 1622–1626.

Communicating editor: J. Shendure

GENETICS

A Male-Specific Genetic Map of the Microcrustacean *Daphnia pulex* Based on Single-Sperm Whole-Genome Sequencing

Sen Xu, Matthew S. Ackerman, Hongan Long, Lydia Bright, Ken Spitze, Jordan S. Ramsdell, W. Kelley Thomas, and Michael Lynch

Supplementary Figure

Figure S1 The male genetic linkage map based on 1672 haplotype blocks derived from 104 single sperm genomes. The map distance for each block is on the left hand side of each linkage group, whereas the marker name is listed on the right hand side, with the number before the underscore designating the scaffold and the number after representing a randomly assigned ranking number (see supplementary data). The red squares next to marker names indicate possibly mis-assembled genomic regions in the *Daphnia arenata* reference genome. Some map positions consist of many haplotype blocks and these blocks are shown in the top left corner.

Chromosome 1
141 cM

S: 42_9,42_10,42_11,42_12,42_13,42_14,42_15,42_16,
42_17,42_18,42_19,42_20,42_21,42_22

0	61_1
2.69	61_2
3.23	162_1,61_3 ■
5.94	61_4
8.64	61_5
13.27	61_6
13.81	61_7
16.36	61_8
16.9	3_1,61_10,61_11,61_12,61_13,61_9
17.43	3_2
21.62	3_3,3_4
23.63	3_5
25.18	3_6
25.69	3_7
29.33	3_8
30.06	3_9
31.68	3_10,3_11
33.28	3_12
33.88	3_13
38.49	3_14
39.01	3_15,3_16
39.55	3_17
43.14	3_18
45.72	3_19,3_20
48.3	3_21,3_22,3_23
49.84	3_24
51.39	3_25
52.94	3_26,3_27
56.94	3_28,3_29,3_30
58	3_31
59.92	3_32
68.57	3_33
69.27	3_34
69.86	3_36,3_35
74.52	3_37,3_38
77.1	3_39
79.17	3_40
83.4	3_41,3_42
85.03	3_43
86.63	3_44,3_45,3_46,3_47
88.24	3_48
88.77	3_49
89.31	3_50
93.94	3_51
95.5	3_52
98.64	211_1,211_2
101.74	211_3,3_53 ■
106.56	106_1
108.94	106_2
111.28	106_3
113.78	106_4
115.54	106_5
116.08	207_1
117.63	53_3,53_1,53_2
118.17	53_4
119.17	53_5
120.3	53_6,53_7,53_8,53_9
123.94	53_10,53_11
128.69	42_1,53_12,53_13,53_14,53_15,53_16,53_17,53_18,53_19
129.36	53_20
130.98	42_4,42_5,42_6,42_7,42_2,42_3
132.57	42_8
133.16	S
133.71	42_23
135.26	130_1
135.86	130_2,42_24,42_25 ■
137.5	130_3,130_4
140.08	66_1
140.6	130_5
141.22	130_6,66_2,66_3,66_4,66_5 ■

S. Xu *et al.*   3 SI

S. Xu *et al.*

Chromosome 2
150 cM

$: 112_1,113_1,115_1,148_1,170_1,206_1,217_1,228_1,
237_1,237_2,266_1,30_16,45_1,45_2,73_1,73_2,73_3,
78_3,78_4,78_1,78_2

$$: 1_36,1_37,1_38,1_39,1_40,1_41,1_42,1_43,1_44,
1_45,1_46,1_47,1_48,1_49,1_50,1_51

$$$: 1_83,1_84,1_85,1_86,1_87,1_88,1_89,1_90,1_91,
1_92,1_93,1_94,1_95,1_96,1_97,1_82,19_1,19_2

$$$$: 19_8,19_9,74_14,74_15,74_16,74_17,74_18,
74_19,74_20,74_21,74_11,74_12,74_13,74_10

Position	Marker
0	135_1
0.99	135_2
8.07	30_1
8.6	30_3,30_2
10.24	30_4
12	30_5
15.64	30_6,30_7,30_8,30_9
16.8	30_10
19.73	30_11
20.27	30_12
21.82	30_13
26.48	30_14,30_15
29.09	$
30.66	134_1,73_4
32.22	71_1
33.77	1_1,1_2,1_3,115_2,115_3,115_4,115_5,115_6,115_7
34.34	1_4
36.93	1_5
37.45	1_6,1_7
42.11	1_8
42.68	1_9,1_10
44.35	1_11,1_12,1_13,1_14
45.25	1_15
50.06	1_17,1_18,1_19,1_20*,1_16
50.83	1_21
51.46	1_22
53.13	1_23
53.87	1_24,1_25,1_26,1_27,1_28,1_29
55.6	1_30,1_31
57.22	1_32,1_33
58.79	1_34
59.31	1_35
59.82	$$
60.34	1_52
63.95	1_53
65.51	1_54,1_55,1_56
68.11	1_57,1_58,1_59
69.67	1_60
71.22	1_61,1_62,1_63
71.73	1_64
72.91	1_65
73.8	1_66
74.32	1_67
74.83	1_68,1_69,1_70,1_71
76.38	1_72
78.92	1_73
79.48	1_74,1_75,1_76,1_77
82.06	1_78,1_79
84.64	1_80,1_81
86.19	$$$
87.73	19_5,19_6,19_3,19_7,19_4
90.31	74_1,74_2
91.86	74_3,74_4,74_5,74_6
92.37	74_7
93.98	74_8
94.6	74_9
95.22	$$$$
95.89	19_10
99.8	19_11
103.06	19_12,19_13
105.99	19_14
108.07	19_15
110.2	19_16,19_17
112.34	19_18,19_19,19_20
114.39	19_21
117.13	19_22,19_23
118.48	19_24
120.46	19_25,19_26
122.22	19_27,19_28
122.94	19_29
126.99	19_30
129.99	19_31,19_32
133.04	90_1
134.16	90_2
136.04	90_3
148.17	90_4
149.69	90_5

Chromosome 3
116 cM

Position	Marker
0	2_1
1.07	2_2
8.28	2_3
9.28	2_6,2_4,2_7,2_5
11.27	2_8
13.2	2_9
14.97	2_10,2_11,2_12
16.52	2_13
18.06	2_14,2_15
20.32	2_16
22.19	2_17
24.06	2_18,2_19,21_1
26.03	21_2
27.94	62_1,62_2,62_3
29.79	62_4,62_5,62_6
31.52	62_7
33.06	62_8
34.61	62_11,62_12,62_9,62_10
35.13	62_13
37.72	97_1
39.33	97_2
41.13	97_3
42.93	97_4
43.59	97_5,97_6,97_7
50.13	97_8
51.35	97_9
55.34	97_10
56.21	2_20,2_21 ■
57.15	97_11
64.79	2_22,2_23
66.35	2_24,2_25
68.93	2_26,2_29,2_27,2_28
70.48	2_30
72.03	2_31
72.54	2_32,2_33,2_34
76.16	2_35,2_36,2_37
77.7	2_38
78.22	2_39,2_40,2_41
80.8	2_42,2_43,2_44,2_45
86.33	2_46
87.39	2_47
91.67	2_48
92.25	2_49,2_50
93.83	2_51
94.35	2_52,2_53,2_54,2_55,2_56,2_57
94.87	2_58
98.55	2_60,2_59
100.19	2_61,2_62,2_63
101.78	26_2
103.37	21_3,21_4,26_3 ■
105.37	16_1
107.42	16_2,16_3,16_4
109.46	16_5
110.61	16_6
111.74	$
112.91	178_3
113.42	178_4
115.57	178_5,2_66,44_8 ■

Chromosome 4
82 cM

Left axis labels (cM positions):
0
1.38
2.27
5.01
5.58
6.1
6.63
10.54
13.96
15.69
17.88
20.03
24.77
25.28
25.84
29.34
29.94
32.52
33.04
37.95
39.72
41.51
44.19
44.92
46.64
47.31
50.82
53.4
55.14
58.56
66.34
66.88
69.76
70.84
75.61
76.17
77.74
79.31
80.86
81.38
81.89

Right axis labels (markers):
163_1
163_2
163_3
163_4,186_3,186_1,186_2,198_5,198_1,198_2,198_3,198_4
43_1
198_6
198_7
43_2
43_3,43_4
43_5
43_6
43_7
43_8
43_13,43_9,43_10,43_11,43_14,43_12
43_15
43_16
43_20,43_18,43_19,43_17
8_2
285_1
8_3
8_4
8_5,8_6
8_7
8_8,8_9
8_10,8_11,8_12,8_13
8_14
8_15,8_16,8_17,8_18,8_19
8_21,8_20
8_23,8_22
8_24,8_25,8_26
11_1,11_2
11_3
11_4,11_5,11_6
11_7
11_8,11_9,11_10
11_11
11_12
11_13,11_14
11_15
11_16
11_17,11_18,11_19,31_1,31_2,31_3

$: 12_23,12_24,12_25,12_26,12_27,12_28,160_1,38_1,
38_2,38_3,38_4,38_5,38_6,38_7,38_8,38_9,38_10,
38_11

$$: 125_1,125_2,125_3,157_4,157_3,272_1,307_1,331_1,
331_2,37_2,37_3,37_4,37_5,40_8,40_9,40_10,49_8

Chromosome 5
133 cM

cM	Marker
0	20_3,20_4
5.42	20_5
6.21	20_6
7	20_7,20_8,20_9
10.35	20_10
12.94	89_1
17.59	89_2
18.17	89_3
18.73	89_4
21.46	89_5,89_6
26.63	89_7
29.67	20_11
32.19	20_12
34.54	20_13
38.13	20_15,20_14,89_8
38.86	89_9
40.48	89_10
41.01	89_11
41.69	89_12
45.27	89_13,89_14
45.85	89_15
49.46	131_2,131_1
53.11	131_3,131_4
67.44	39_1
68.19	15_1,15_2
69.94	15_3
71.75	15_5,15_4
74.72	15_6,15_7,15_8,15_9,15_10,15_11
77.66	15_12,15_13,15_14
79.42	15_15,15_16
81.26	15_17
84.27	39_3,39_2,39_4,39_5,39_6
86.38	39_7,39_8,39_9
87.93	39_10,39_11,39_12
89.48	39_13,39_14
91.02	15_18
92.17	15_19,15_20,15_21,15_22
94.19	15_23,15_24,15_25,15_26,15_27,15_28
95.81	12_1
96.36	12_2,12_3,12_4,12_5,15_29
96.92	12_6
98.7	12_7,12_8,12_9,12_10,12_11,12_12
101.22	12_13
102.86	12_14,12_15
104.55	12_16,12_17
106.18	12_18
106.76	12_19
108.33	12_20,12_21
109.88	12_22
110.39	$
111.94	37_1
112.46	157_1
112.98	157_2
113.49	$$
114.01	37_6
115.55	213_1,275_1,98_1
119.17	213_2
120.72	60_1
121.33	60_5,60_3,60_2,60_4
122.8	60_6
129.52	60_7
130.26	60_8
132.62	60_9

Chromosome 6
119 cM

$: 128_3,128_4,128_5,128_6,128_7,128_1,128_2,230_1,32_27, 32_28,32_17,32_18,32_19,32_20,32_21,32_22,32_23,32_24,32_25, 32_26,45_4,45_3,57_3,57_4,92_2,92_1,92_3,92_4

Position	Marker
0	80_1
0.52	80_2,80_3,80_4
1.09	80_5
4.7	80_6
6.5	80_7
8.31	80_8
9.92	80_9
13	23_1
17	23_2
17.96	23_3
18.98	23_4
22.76	23_5,23_6
25.35	28_10,28_1,28_2,28_3,28_4,28_5,28_6,28_7,28_8,28_9
25.9	28_11
28.45	28_12
30	28_13
30.52	28_14
31.04	28_15,28_16,28_17
31.56	28_18
35.17	28_19
37.77	28_20,28_21
41.42	187_1
42.08	187_2,187_3,187_4
42.67	23_7
43.2	187_5
45.81	23_8,23_9,23_10,23_11
47.38	23_12
48.96	123_1
52.58	123_2
54.21	23_13
55.78	123_3
57.34	123_4
57.86	123_5
58.38	123_6,123_7,23_14
60.96	23_15,23_16
62.54	23_19,23_20,23_21,23_17,23_18
63.11	23_22
64.69	23_23
66.23	23_24,23_25,23_26,23_27,23_28
67.78	23_29,23_30
68.31	299_1
68.85	23_31
75.54	47_1
76.69	47_2,47_3
78.69	47_4
80.26	47_5
80.86	47_6
83.52	47_7,47_8,47_9,47_10,47_11,47_12
84.03	47_13
85.58	47_14,47_15,47_16,47_17
87.14	47_18
88.7	47_20,47_21,47_22,47_23,47_19
90.32	138_1,138_2
94.11	138_7,138_3,138_8,138_4,138_5,138_6
95.85	138_10,138_9
98.54	191_1
101.12	191_3,191_2
102.75	191_4
105.03	32_1,32_2,32_3
108.95	32_4
111.44	32_5,32_6,32_7,32_8
113.86	32_9,32_10,32_11,32_12,32_13,32_14,32_15
114.39	32_16
115.96	$
116.51	230_2
117.04	32_29
118.59	128_8,57_5,57_6,57_7

$: 101_1,101_2,101_3,101_4,133_2,133_3,133_4,133_1,
190_1,190_2,40_3,40_4,40_5,40_6,40_1,40_2,48_2,48_1,
48_3,48_4,48_5,48_6,48_7,54_1,54_2,54_3,54_4,55_2,
55_3,55_4,55_5,55_6,55_7,68_1,82_1,82_1,82_2,93_1,93_2

$$: 129_1,129_2,204_1,220_1,220_2,220_3,220_4,250_2,46_1,
46_2,55_11,55_9,55_10,65_3,65_4,65_5,65_6,65_7,65_8,
68_3,91_2,91_3,91_4

Chromosome 7
139 cM

Left positions	Right labels
0	87_1
0.52	87_2,87_3,87_4
6.22	87_5,87_6
9.89	145_1
16.67	145_2
17.19	4_1
18.74	4_3,4_2
22.28	4_4
22.87	4_5,4_6
26.49	4_7,4_8,4_9
28.04	4_10,4_11
28.59	4_12
33.28	4_13
34.12	4_14,4_15
35.87	4_16
37.51	4_17
39.1	4_21,4_19,4_18,4_20
40.7	4_25,4_28,4_26,4_22,4_27,4_24,4_23
41.29	4_29
41.83	4_30
42.35	4_31
49.16	4_32,4_33
53.96	4_34
54.67	4_35
57.76	4_36,4_37
59.97	4_38
61.06	4_39,4_40
63.15	4_41,4_42
65.26	4_43
68.05	4_44
70.33	4_45,4_46,4_47,4_48
72.19	4_49
74.09	4_50,4_51
76.1	4_52,4_53,4_54,4_55
77.09	4_56
79.12	257_1,4_57
81.33	55_1
82.52	$
83.5	40_7
84.36	101_5
84.87	48_8
85.39	55_8
87.97	250_1,65_2
89.51	68_2
90.33	91_1
91.27	$$
95.34	91_5
96.32	46_3,91_6,91_7,91_8
97.35	46_4
102.5	18_1
103.59	18_2
110.72	18_3,18_4
121.72	18_5
123.19	18_6
124.81	18_7
125.33	18_8
126.87	18_9,18_10
128.42	18_11
131	18_12
132.55	18_13
133.84	18_14,18_15,18_16,18_17
138.75	18_18,18_19,18_20

$: 109_1,109_2,126_1,169_2,169_3,169_1,199_1,199_2,20_2,
20_1,22_2,22_3,22_4,226_5,226_3,226_4,247_1,291_2,34_2,
52_6,52_5,52_7,57_1,57_2,76_22,76_23,76_24,76_25,77_4,
77_5,77_6,77_7,77_8,77_9,77_10,77_11,77_12,77_13,77_14,
77_15,94_3,94_4,94_5

Chromosome 8
113 cM

Position	Marker
0	25_1,25_2
4.85	
6.58	25_3,25_4
8.25	25_5
	25_6
10.84	25_7
11.36	25_8,25_9
17.27	25_10
18.17	25_11
22.6	25_12
26.22	25_13
26.75	25_14
30.29	25_15
30.91	25_16
31.48	25_17
39.64	25_18,25_19,25_20
41.03	25_21
47.81	25_22
49.71	25_23,259_1,259_2,7_2,7_3,7_1
50.93	259_3
53.44	7_4
54.99	7_5,7_6,7_7,7_8,7_9,7_10
56.63	7_11,7_12,7_13
58.28	7_14,7_15
59.89	7_22,7_23,7_24,7_25,7_16,7_17,7_18,7_19,7_20,7_21
63.56	7_26,7_27,7_28,7_29
65.22	7_30,7_31
68.72	7_32,7_33,7_34
70.27	7_35
72.36	7_36,7_37
74.44	7_38
75.99	7_39
78.57	83_1 ■
80.12	7_40,7_41,7_42,7_43,83_3,83_4,83_2
81.67	83_5,83_6,83_7,83_8,83_9
83.22	83_10,83_11,83_12,83_13,83_14
85.99	83_15
87.77	76_4,76_1,76_5,76_2,76_3,83_16
88.35	76_6
89.89	76_7
91.47	76_8
95.12	76_9,76_10
96.67	76_11,76_12
98.25	76_13
99.83	76_14,76_15
101.38	76_16,76_17,76_18
103	76_19
104.61	33_1
105.13	226_1,291_1,52_2,52_1,52_3,52_4,76_20,76_21,77_1,77_2,850_1,94_1 ■
106.68	183_1 ■
107.2	33_2
107.71	226_2
108.23	34_1
108.74	94_2
109.29	22_2
109.87	22_1 ■
110.46	$ ■
111.05	76_26 ■
112.66	52_8,77_16,77_17,77_18

S. Xu *et al.*

Chromosome 9
119 cM

$: 104_1,107_2,107_3,107_4,107_5,107_1,118_1,118_2,
139_1,14_7,14_2,14_3,14_4,14_5,14_6,172_1,
289_1,41_2,41_3,41_4,41_5,41_6,41_7,95_1,95_2,
99_2

Position	Marker
0	9_1
1.53	9_2,9_3
6.15	9_4
13.65	9_5
14.95	9_6,9_7,9_8
17.3	9_9,9_10
21.63	9_11
22.8	9_12
32.19	9_13
33.73	9_14,9_15
38.25	9_16
40.38	9_17
41.34	9_18
42.34	9_19,9_20,9_21,9_22,9_23
44.34	9_24,9_25,9_26,99_1
45.21	9_27
49.21	14_1,41_1
51.12	$ ■
52.02	41_8
52.85	99_3
53.62	99_4
54.48	41_9
56.24	107_6,118_3,14_9,14_8
57.95	118_4 ■
58.59	14_10
60.17	14_11,14_12,14_13
60.85	14_14
66.65	13_1,13_2
68.2	13_3,13_4,13_5,13_6,13_7
69.75	13_8
70.27	13_9
72.85	13_10,13_11
75.43	13_12
75.94	13_13
88.2	13_14
89.88	13_15
90.43	13_16,13_17
91	13_18
95.56	13_19
96.12	13_21,13_20
96.68	13_22
100.34	13_23
100.85	13_24
103.43	13_25
106.01	51_1
107.85	51_2
109.11	51_3,51_4
117.96	51_5
118.48	51_6

$: 102_2,102_1,103_1,103_2,152_1,152_2,35_2,
35_3,35_4,35_1,35_5,35_6,35_7,56_2,56_3,
56_4,56_5,56_6,56_7,56_8,56_9,56_10,56_1,
88_6,88_7,88_8,88_3,88_9,88_4,88_5

$$: 105_1,120_1,176_1,223_2,223_1,26_1,49_4,49_5,
49_6,49_7,49_2,49_3,64_1,64_2,64_3,64_4,64_5,64_6,
65_1,79_2,79_3,79_4,79_5

Chromosome 10
139 cM

Position	Marker
0	36_1,36_2,36_3,36_4,36_5
1.36	36_6
3.69	36_7
8.35	36_8,6_1,6_2
8.88	6_3
13.67	6_4
14.47	6_5
18.3	6_6
20.91	6_7
21.44	6_8
27.4	6_9
28.58	6_10
35.7	6_11
36.23	59_1,6_12,6_13,6_14,6_15 ■
37.77	6_16
38.29	6_17,6_18
41.22	6_19
42.12	6_20
50.91	6_21,6_22,6_23
52.6	6_24
54.54	6_25
57.35	6_26
58.33	6_27,88_1
62.41	88_2
63.4	$ ■
64.38	88_10
65.22	56_11
66.78	79_1
67.3	49_1
68.09	$$ ■
70.08	100_1,182_2,182_3,182_1,79_6,79_7
71.96	100_2,100_3,100_4,100_5,29_1
72.49	100_6
74.03	29_3,29_4,29_2
74.56	29_5
78.23	29_6
79.86	29_7
82.46	29_8
82.98	29_9,29_10
84.52	29_11
86.08	121_1
87.63	121_2
88.15	121_3
89.74	121_4
91.38	121_5
94.03	121_6
95.58	121_7,121_8,17_1,17_2,17_3
97.13	17_4
98.68	17_5
99.2	17_6,17_7
101.83	17_8
103.48	17_9
105.1	17_10
106.65	17_12,17_11
108.28	17_13
109.95	17_14
111.65	17_16,17_17,17_15
114.33	17_18
118.03	17_19,17_20
119.66	17_21
120.42	17_22,17_23,17_24,17_25,17_26,17_27,17_28
123.36	332_1
126.16	72_1
128.06	72_2
129.61	72_3,72_4
132.22	72_5
132.8	72_6,72_7,72_8
135.44	72_9
138.05	173_2,173_3,173_1,72_10 ■
138.57	72_11

S. Xu *et al.*

Chromosome 11
83 cM

Position	Markers
0	67_1
0.52	67_2
3.12	10_1
6.8	10_2
11.91	10_3
12.93	10_4
13.97	10_5
23.28	10_6
33.35	10_7
33.87	10_8,10_9,10_10
35.69	10_11,10_12,10_13
39.05	10_14
40.6	10_15
43.16	10_16
44.73	24_1,24_2
47.39	24_3,24_4
49.09	10_18,10_19,10_20,10_17,24_5,24_6,24_7,743_1
51.59	24_8
53.13	24_9,24_10
54.68	24_11,24_12
55.22	24_13
56.84	10_21,24_14,24_15,24_16
57.42	24_17
57.99	24_18
59.57	24_19,24_20
61.19	24_21
65.8	24_22
66.32	24_23
68.92	24_24
69.55	24_25
70.21	24_26,24_27
76.93	280_1,75_1,8_1
77.45	280_2
78.44	75_2
79.54	75_3
81.52	111_4,111_5,111_1,111_6,111_7,111_2,111_3,75_4,75_5
82.64	111_8

S. Xu *et al.*

Chromosome 12
121 cM

$: 146_2,146_3,146_1,158_1,158_2,161_2,161_3,161_1,
27_6,27_2,27_3,27_4,27_5,58_7,58_8,58_9,58_5,58_6,
59_4,59_5,59_6,59_7,63_14,84_2,84_1

0
0.52
4.16
7.8
12.57
16.63
17.4
20.28
21.27
22.23
25.02
26.82
27.75
32
33.24
37.04
39.62
42.2
45.81
47.85
49.01
51.69
55.3
57.85
58.4
59.07
60.46
62.01
63.55
65.1
66.66
67.42
69.81
71.46
73.1
75.45
77.92
81.69
83.24
84.79
85.42
87.22
88.14
90.14
90.94
92.6
94.14
94.92
95.84
96.75
98.03
100.34
102.25
106.21
106.82
111.48
116.27
117.93
118.45
119.07
120.57

142_1
142_2,142_3,142_4
142_5
142_6
5_1
5_2
5_3
5_4
5_5,5_6,5_7,5_8,5_9,5_10
5_11
5_12
5_13
5_14
5_15,5_16
5_17
5_18
5_19,5_20
5_21
5_23,5_22
5_24,5_25
5_26
5_27
5_28,5_29
5_30
5_31,5_32,5_33,5_34,5_35
5_36
5_37
5_38
5_39,5_40
5_41,5_42,5_43
5_44
5_45,5_46,5_47,5_48,5_49,5_50,5_51
5_53,5_54,5_52
142_7
142_8
142_9,5_55
5_56,63_1
63_2
63_3
63_4
63_5
63_6,63_7,63_8,63_9,63_10
63_11
63_12
63_13
27_1,58_1,58_3,58_2
58_4
59_2
59_3
$
58_10
161_4,731_1
59_8
86_1
58_11,96_1
86_2
621_1,70_1
70_2
70_3
70_4,70_5,70_6
70_7

Files S1-S4

Available for download at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.179028/-/DC1

File S1. The raw text output file for the genetic map of *D. pulex*.

File S2. Python script phasingHaplotype.py to generate haplotype blocks and input file for MSTmap software. For usage and input data format, please see beginning of the file.

File S3. MareyMap database created by combining the genetic map data and the current *Daphnia* reference assembly.

File S4. The raw text output file for the genetic map based on sperm *de novo* assembly.

File S5

Supplementary material

Method for haplotype block reconstruction

The first step in the construction of our genetic map is to remove sites where very few of the sperm have SNP calls. When very few sperm samples were sequenced at a site, it becomes difficult to detect recombination events between a pair of sites since they may be no sperm sequenced at both sites. Once these sites are removed, the remaining polymorphic sites that are adjacent to each other in the reference assembly are clustered into blocks that show no evidence of recombination, and then these blocks are phased in order to minimize the number of recombination events. Whether a recombination event occurs between two sites can be easily checked by counting the number of haplotypes that occur at each site. If only two haplotypes occur, then there is no evidence that a recombination event occurred between these sites. Then they are joined together into a longer haplotype block. Each haplotype block is extended to include all adjacent sites for which there is evidence of only two haplotypes. When a third haplotype is detected, a new block is created and the procedure is repeated. Once all the haplotype blocks are established, they are phased to minimize the number of recombination events occurring between adjacent haplotype blocks. This can be easily accomplished since each haplotype block has only two potential phases, and we assume that each haplotype block is in the phases that gives a lower number of recombination events with the upstream block. Since haplotype blocks are constructed under the assumption that sites adjacent to each other in the reference assembly are separated by very little genetic distance, it is not expected a large number of recombination events occurring between adjacent blocks. When this does occur, we can mask the downstream haplotype block which has an excessive number of recombination events and

S. Xu *et al.*

ignore this haplotype block for phasing purposes. Finally, the adjacent haplotype blocks that are

not affected by recombination events need to be placed into the same parental or maternal phases

so that the MST software, which assumes that the parental and maternal identity of markers are

known, can be used to construct a genetic map.