Supplement

Supplementary texts

PBcR settings

PBcR, from the Celera WGS version 8.3, release candidate 1, was used. We used the following command line options:

```
PBcR -sensitive -libraryname shrimp_pacbio_round2-3-4 -s pacbio.spec -f
astq_linebreak shrimp_pb_data.fastq
```

Where the file pacbio.spec contained the following:

```
asmOvlErrorRate=0.1
asmUtgErrorRate=0.1
asmCnsErrorRate=0.1
asmCgwErrorRate=0.1
asmOBT=1
asmObtErrorRate=0.08
asmObtErrorLimit=4.5
utgGraphErrorRate=0.05
utgMergeErrorRate=0.05
ovlHashBits=24
ovlHashLoad=0.8
utgMergeErrorLimit=5.25
useGrid=1
scriptOnGrid=1
ovlCorrOnGrid=1
frgCorrOnGrid=1
ovlMemory=128
ovlStoreMemory=128000
threads=32
ovlConcurrency=1
cnsConcurrency=32
merylThreads=32
merylMemory=128000
frgCorrThreads = 16
frgCorrBatchSize = 100000
ovlCorrBatchSize = 100000}
```

DBG2OLC settings

We ran DBG2OLC with the following command line options:

```
./DBG2OLC_Linux k 17 KmerCovTh 2 MinOverlap 20 AdaptiveTh 0.002 LD1 0 M
inLen 200 Contigs linebreak platanus_contigs.fa RemoveChimera 1 f pacbi
o_data.fa
```

quickmerge settings

We ran quickmerge with the following command line options:

```
python merge_wrapper.py -pre merged_quivered_shrimp_assemblies -hco 5.0
-c 1.5 -l 1000000 linebreak hybrid_assembly.fasta .self_assembly.fasta
```

BWA settings

BWA and *samtools* command line settings for aligning and filtering reads were as follows:

```
bwa aln -t ${CORES}  $REFPATH $FDATAPATH > ${OUTPATH}.F.sai
bwa aln -t ${CORES}  $REFPATH $RDATAPATH > ${OUTPATH}.R.sai
bwa sampe ${REFPATH} ${OUTPATH}.F.sai ${OUTPATH}.R.sai $FDATAPATH $RDAT
APATH | samtools view -q 20 -bS - | samtools sort - data/bam/$PREFIX
```

picard-tools settings

picard-tools command line settings for deduplication were as follows:

```
java -jar picard.jar MarkDuplicates INPUT=${prefix}.bam OUTPUT=${prefix
}.dedup.bam METRICS_FILE=${prefix}.dedup.metrics.txt REMOVE_DUPLICATES=
true}
```

GATK settings

GATK command line settings for calling SNPs were as follows:

```
java -d64 -Xmx128g -jar GenomeAnalysisTK.jar -T UnifiedGenotyper -nt ${
CORES} -R ${REFPATH} -I merged-realigned-deduped.bam -gt_mode DISCOVERY
-stand_call_conf 30 -stand_emit_conf 10 -o rawSNPS-Q30_v2.vcf
java -d64 -Xmx128g -jar GenomeAnalysisTK.jar -T VariantAnnotator -nt ${
CORES} -R ${REFPATH} -I merged-realigned-deduped.bam -G StandardAnnotat
ion -V:variant,VCF rawSNPS-Q30_v2.vcf -XA SnpEff -o rawSNPS-Q30-annotat
ed_v2.vcf
java -d64 -Xmx128g -jar GenomeAnalysisTK.jar -T UnifiedGenotyper -nt ${
CORES} -R ${REFPATH} -I merged-realigned-deduped.bam -gt_mode DISCOVERY
-glm INDEL -stand_call_conf 30 -stand_emit_conf 10 -o inDels-Q30_v2.vcf
java -d64 -Xmx20g -jar GenomeAnalysisTK.jar -T VariantFiltration -R ${R
EFPATH} -V rawSNPS-Q30-annotated_v2.vcf --mask inDels-Q30_v2.vcf --mask
Extension 5 --maskName InDel --clusterWindowSize 10 --filterExpression
"MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)" --filterName "BadValidation" -
-filterExpression "QUAL < 30.0" --filterName "LowQual" --filterExpressi
on "QD < 5.0" --filterName "LowVQCBD" --filterExpression "FS > 60" --fi
lterName "FisherStrand" -o Q30-SNPs_v2.vcf
cat Q30-SNPs_v2.vcf | grep 'PASS|textasciicircum#' > only-PASS-Q30-SNPs
```

```
_v2.vcf
java -d64 -Xmx20g -jar GenomeAnalysisTK.jar -T VariantFiltration -R ${R
EFPATH} -V inDels-Q30_v2.vcf --clusterWindowSize 10 --filterExpression
"MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1)" --filterName "BadValidation" -
-filterExpression "QUAL < 30.0" --filterName "LowQual" --filterExpressi
on "QD < 5.0" --filterName "LowVQCBD" --filterExpression "FS > 60" --fi
lterName "FisherStrand" -o Q30-INDEL_v2.vcf
cat Q30-INDEL_v2.vcf | grep 'PASS|textasciicircum#' > only-PASS-Q30-IND
EL_v2.vcf
```

Supplementary tables

Cutoff	N50	Contigs	Length	Largest
82	1420491	253	120342282	4450974
83	1450029	251	120059725	4448668
84	1424862	230	119347677	10883472
85	1926101	210	118647318	10769207
86	1886292	183	116223020	7955201
87	219999	864	106905299	1281479
88	34294	1080	29263120	130428
89	9895	20	156585	21076
90	14014	3	22150	14014
91	12650	3	21171	12650
92	12644	3	21161	12644

Supplementary Table 1: Assembly statistics for hybrid genome assemblies using various quality thresholds.

GO type	GO term	Description	P-value	FDR q-value
Function	GO:0042302	structural constituent of cuticle	2.44E-06	5.32E-03
Function	GO:0008061	chitin binding	3.10E-06	3.38E-03
Function	GO:0008010	structural constituent of chitin-based larval cuticle	5.31E-06	3.87E-03
Function	GO:0005214	structural constituent of chitin-based cuticle	2.34E-05	1.28E-02
Function	GO:0004180	carboxypeptidase activity	2.52E-05	1.10E-02
Function	GO:0004099	chitin deacetylase activity	8.21E-05	2.99E-02

Function	GO:0016490	structural constituent of peritrophic membrane	1.93E-04	6.03E-02
Function	GO:0070026	nitric oxide binding	4.34E-04	1.19E-01
Function	GO:0070025	carbon monoxide binding	4.34E-04	1.05E-01
Function	GO:0019826	oxygen sensor activity	4.34E-04	9.49E-02
Function	GO:0030594	neurotransmitter receptor activity	8.06E-04	1.60E-01
Function	GO:0008094	DNA-dependent ATPase activity	9.34E-04	1.70E-01
Component	GO:0000796	condensin complex	4.70E-06	4.54E-03
Component	GO:0005576	extracellular region	1.77E-05	8.55E-03
Component	GO:0044815	DNA packaging complex	2.36E-05	7.60E-03
Component	GO:0008074	guanylate cyclase complex soluble	4.34E-04	
Component	GO:0044421	extracellular region part	9.80E-04	1.89E-01
Process	GO:0006022	aminoglycan metabolic process	3.79E-05	2.12E-01
Process	GO:1903046	meiotic cell cycle process	4.48E-05	1.25E-01
Process	GO:0006030	chitin metabolic process	5.61E-05	1.05E-01
Process	GO:1901071	glucosamine-containing compound metabolic process	1.08E-04	1.51E-01
Process	GO:0006040	amino sugar metabolic process	1.32E-04	1.47E-01
Process	GO:1903047	mitotic cell cycle process	1.98E-04	1.84E-01
Process	GO:0040003	chitin-based cuticle development	2.63E-04	2.10E-01
Process	GO:0007366	periodic partitioning by pair rule gene	3.96E-04	2.76E-01
Process	GO:0006721	terpenoid metabolic process	5.05E-04	3.13E-01
Process	GO:0007512	adult heart development	6.52E-04	3.64E-01
Process	GO:000737	cephalic furrow formation	6.52E-	3.31E-

	6		04	01
Process	GO:0042335	cuticle development	9.47E-04	4.40E-01

Supplementary Table2: GO terms for differentially expressed genes between males and hermaphrodites.

Supplementary figures

Supplementary Figure 1: A bootstrap consensus maximum likelihood tree, generated using 300 bootstrap iterations in MEGA, comparing the predicted translations of putative clam shrimp HOX genes to D. melanogaster *HOX genes. Clam shrimp genes are assumed to be orthologous to* D. melanogaster *genes that are sister to them.*

Supplementary Figure 2: A bootstrap consensus maximum likelihood tree, generated using 300 bootstrap iterations in MEGA, comparing the CDS sequences putative clam shrimp HOX genes to D. melanogaster *HOX genes. Because of the degree of divergence between sequences, this tree had to be created using an approach that allows the usage of information at sites that are missing in some sequences. Thus, it is less trustworthy than the protein tree, and was not used to call orthologs. The exception to this is the two unannotated genes: their genomic sequences were included in this tree, and they were found to be orthologous to* D. melanogaster Scr*.*

*Supplementary Figure 3: A set of line diagrams indicating the mapping of a set of markers onto the putative sex chromosome using a variety of methods. The methods are as follow, from top to bottom: 1. BLAST alignment results of the markers onto the genome assembly contigs that we hypothesize make up the sex chromosome. 2. A linkage map of the sex chromosome from a hermaphrodite cross, as performed in Weeks 2010. 3. A linkage map of the sex chromosome from a male * hermaphrodite cross, as performed in Weeks 2010. The length of the hermaphrodite linkage map is scaled to the assembly contig lengths according to the genome-wide recombination rate calculated in this paper. Note that the total map distance in hermaphrodites resembles the total physical distance in the assembly, but the total map distance in males is much larger.*

Supplementary Figure 4: A histogram indicating the number of polymorphic loci across the genome in the inbred JT4(4)5-L strain. The JT4(4)5-L inbred strain Illumina sequencing data was aligned to the final reference and SNPs were called with GATK. Plotted here are histogram bins (width of 100kb) containing counts of polymorphic loci. SNPs with estimated frequencies less than 5% or greater than 95% were excluded to avoid calling errors as SNPs. Some contigs seem to contain a small amount of residual heterozygosity, while others are nearly free of polymorphism. The regions of residual heterozygosity are generally associated with the small contigs, perhaps explaining their inability to assemble to larger sizes, or perhaps indicating that these

smaller contigs harbor middle repetitive DNA consistent with their being heterochromatic.

Supplementary references

1. Weeks SC, Benvenuto C, Sanderson TF, Duff RJ. Sex chromosome evolution in the clam shrimp, Eulimnadia texana. Journal of Evolutionary Biology. 2010;23:1100–1106.