

Every species can be a model: Reference-quality PacBio genomes from single insects

Sarah B Kingan¹, Christine C Lambert¹, Liuqi Gu², Katharina von Wyszczetki³, Primo Baybayan¹, Konrad Lohse⁴, James R Walters², Mara KL Lawniczak³, Jonas Korf¹
 1. PacBio, Menlo Park, CA, 2. University of Kansas, Lawrence, 3. Wellcome Sanger Institute, Hinxton UK, 4. Edinburgh University, UK.



Insect Assembly

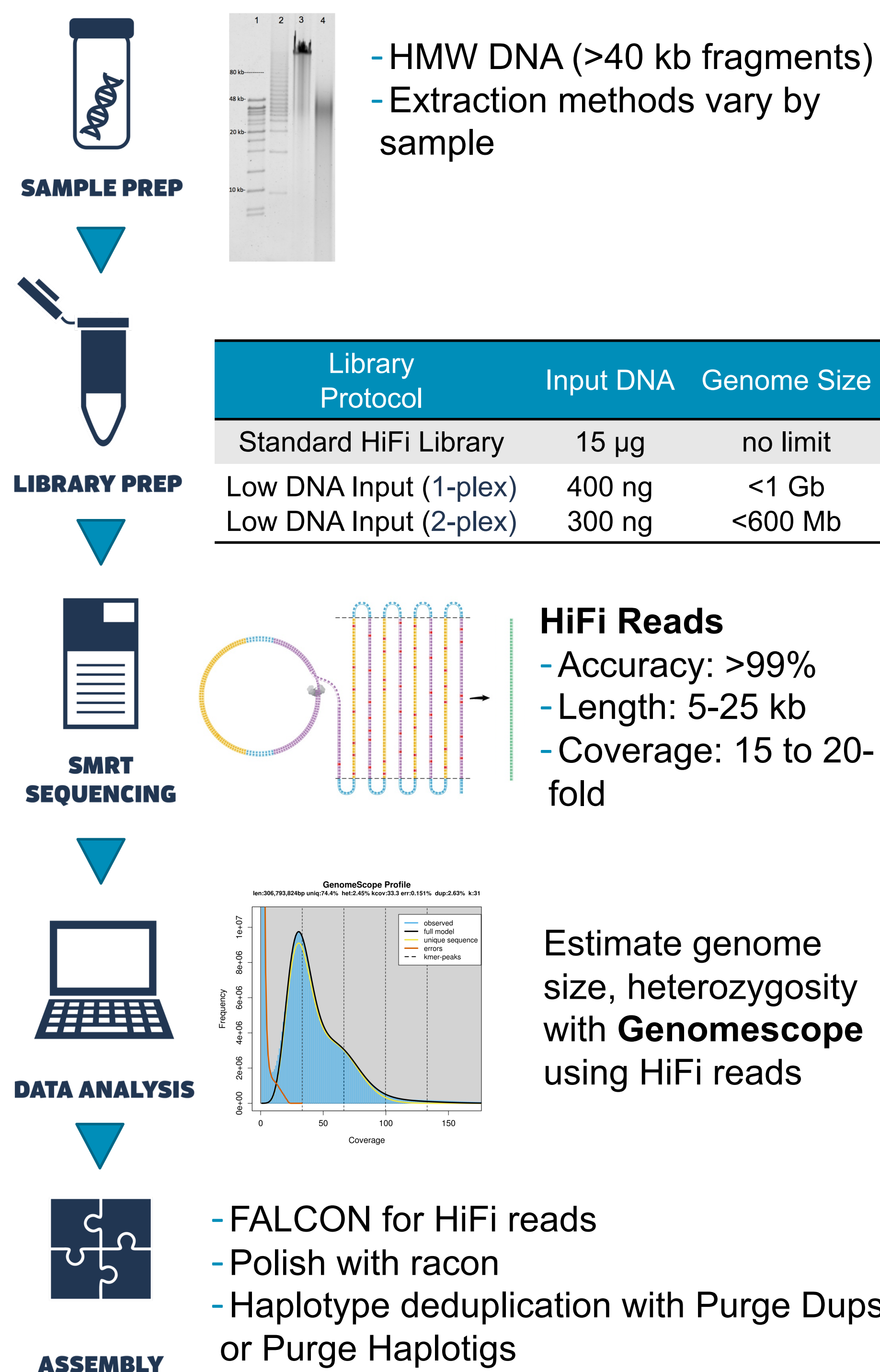
Challenges

- Small body size limits amount of genomic DNA from a single individual
- High heterozygosity samples require diploid-aware *de novo* assembly and curation strategies
- Pooling multiple individuals complicates bioinformatics analysis

Solutions

- PacBio libraries generated from single insects enabled by Low DNA Input Protocol
- Two samples (genome <600 Mb) can be multiplexed on one SMRT Cell 8M on Sequel II System
- Genome assembly with HiFi reads is computationally efficient
- Assemblies from HiFi reads are more complete than other technologies

PacBio Workflows



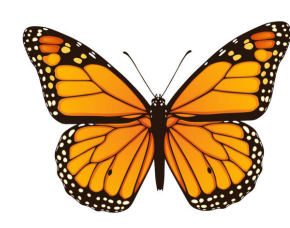
References and Acknowledgements

All protocols can be found on our website:
 - www.pacbio.com/documentation/
 BUSCO: Waterhouse RM et al. (2017) *Mol Biol Evol*, 35(3):543-548
 Purge Dups: https://github.com/dfguan/purge_dups
 FALCON: <https://github.com/PacificBiosciences/pb-assembly>
 Canu: <https://github.com/marbl/canu>
 Purge haplotigs: https://bitbucket.org/mroachawri/purge_haplotigs/
 Racon: <https://github.com/lcb-sci/racon>
 Genomescope: <https://github.com/schatzlab/genomescope>

- Mongue A. J. et al. (2017) G3: 7:3281
- Killick R., and Eckley, I. (2014). J of Stat Software, 58:1t
- Kingan S., et al. (2019) Genes, 10:62
- Vitková M., et al. (2005) Chrom Res. 13:145

The authors wish to thank Michelle Vierra, Greg Concepcion, Maggie Weitzman, Kristin Mars, Nick Sisneros, Adam Knight, Kristin Robertshaw, Pamela Bentley Mills

Monarch (*Danaus plexippus*)



SAMPLE PREP

- Single Female pupa collected
- DNA extraction with OmniPrep Kit

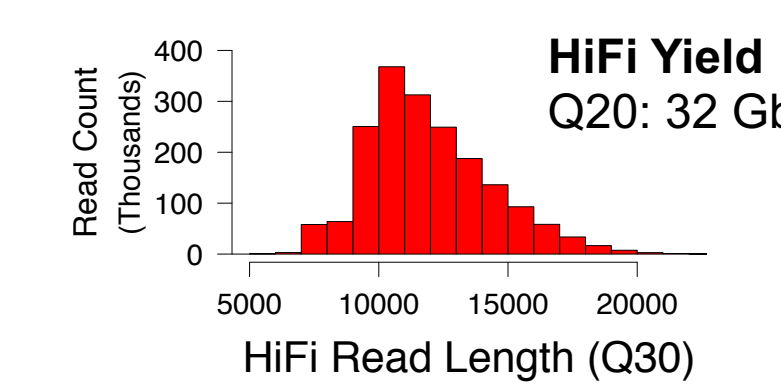


LIBRARY PREP

- Standard HiFi Library Prep with 16 µg DNA

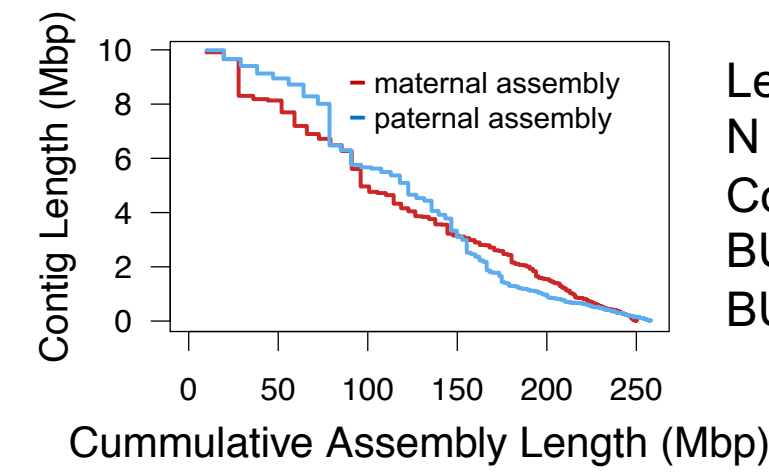


SMRT SEQUENCING

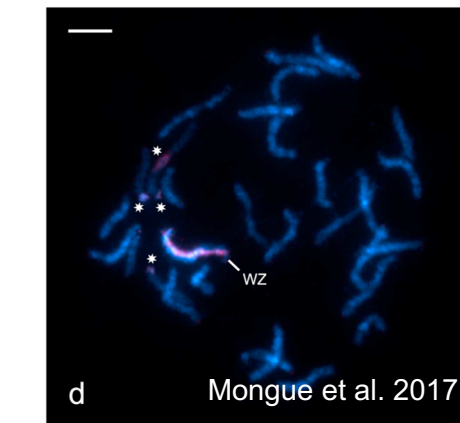
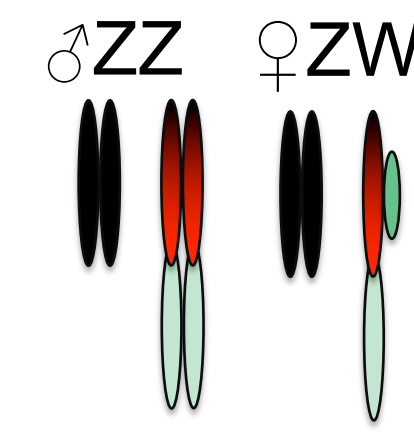


ASSEMBLY

- Trio binning + FALCON

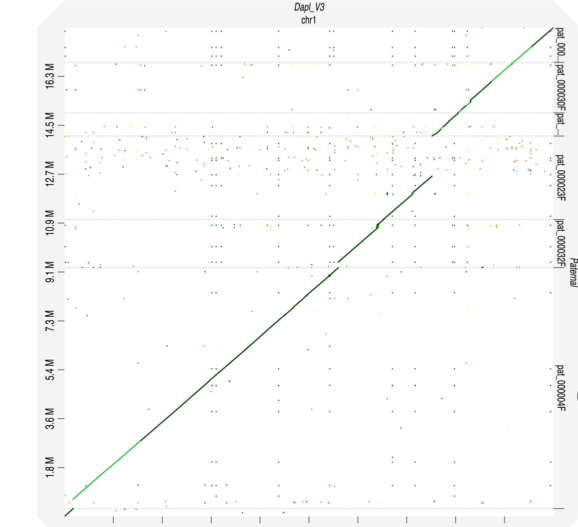


Identifying Sex Chromosomes (ZW)

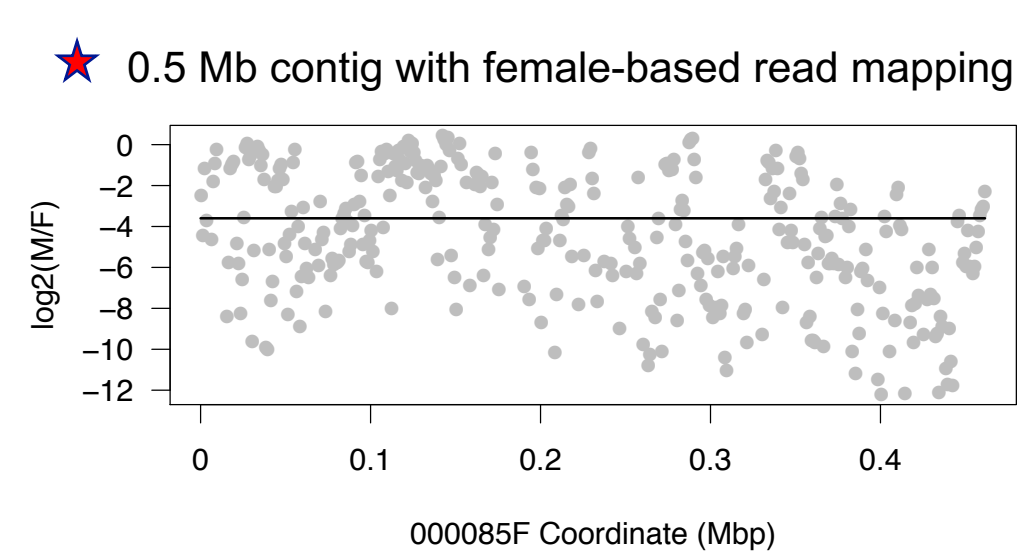


- Neo-Z chrom resulted from fusion of Ancestral Z and autosome¹
- W chromosome absent from previous reference but cytogenetic analysis consistent with fused Neo- and Anc-W

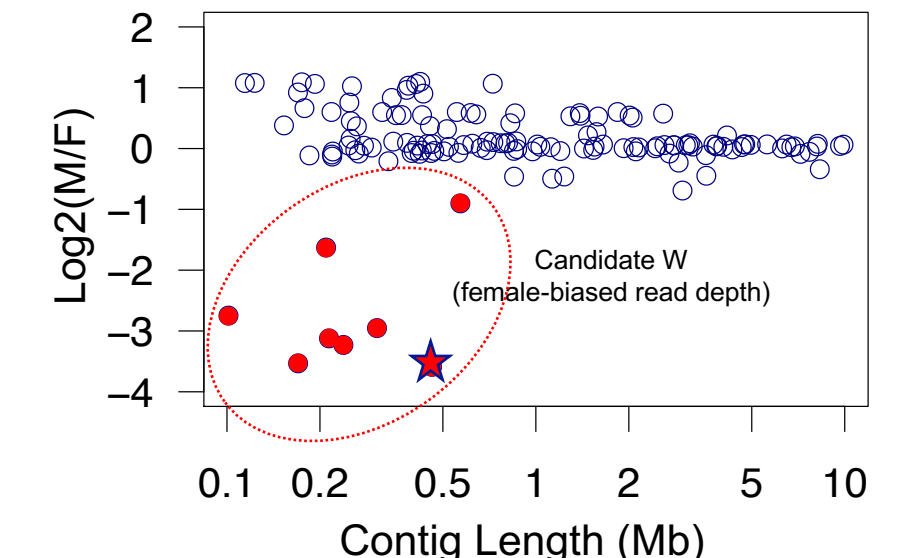
Dapl v3 Z Chrom



- Paternal assembly spans the 16Mb Neo- and Anc-Z with 7 contigs (left)
- Candidate W contigs identified by Illumina read depth analysis² (below)



Sex-Ratio of Illumina Read Depth for Maternal Assembly



Red Admiral (*Vanessa atalanta*)



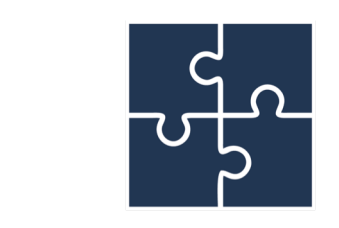
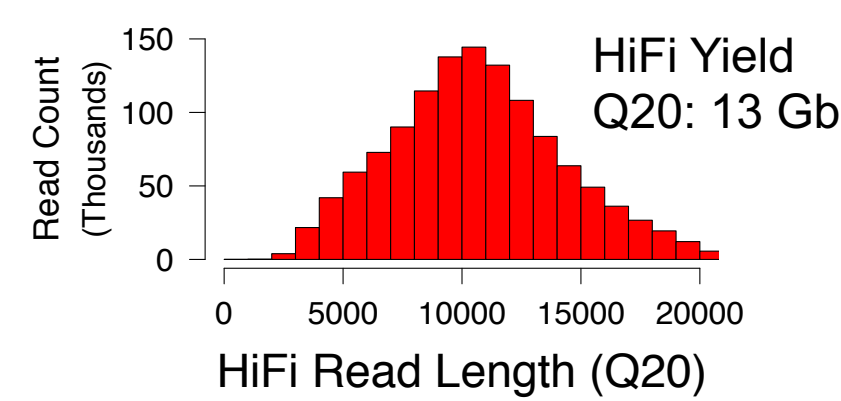
LIBRARY PREP

- Single Female collected
- DNA extraction with "10X modified" protocol³

- Low DNA Input Library Prep

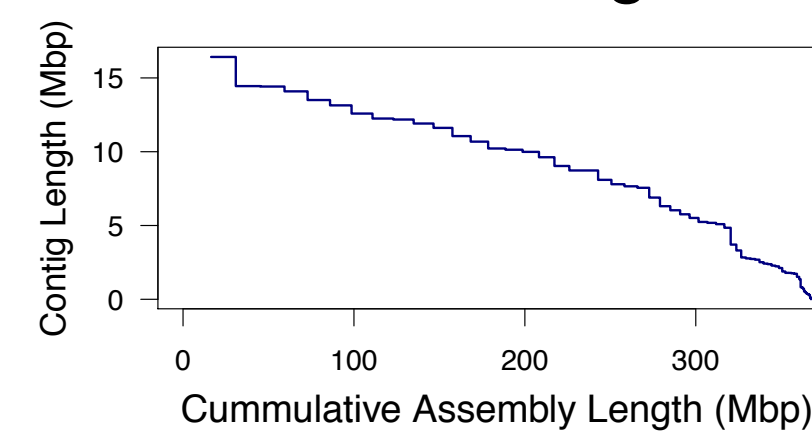


SMRT SEQUENCING

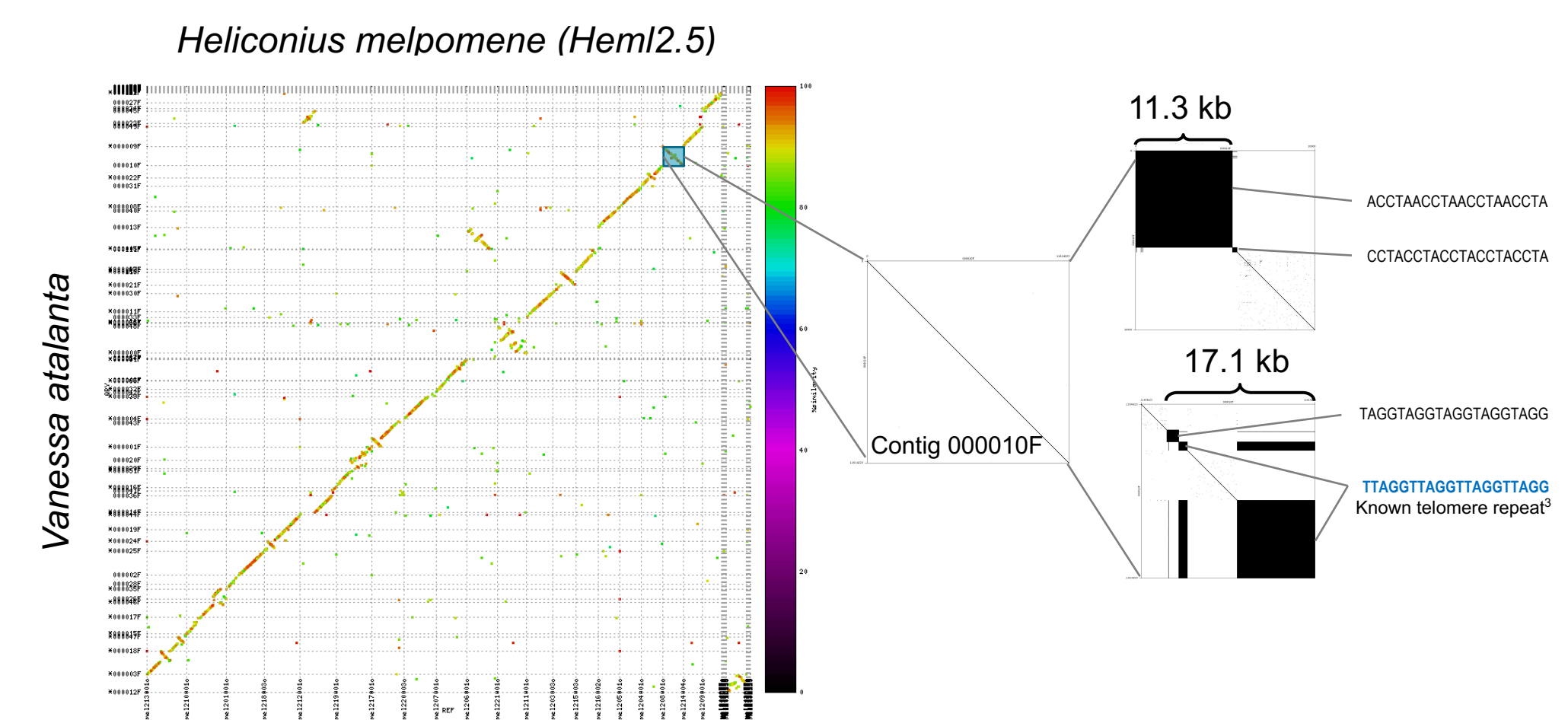


ASSEMBLY

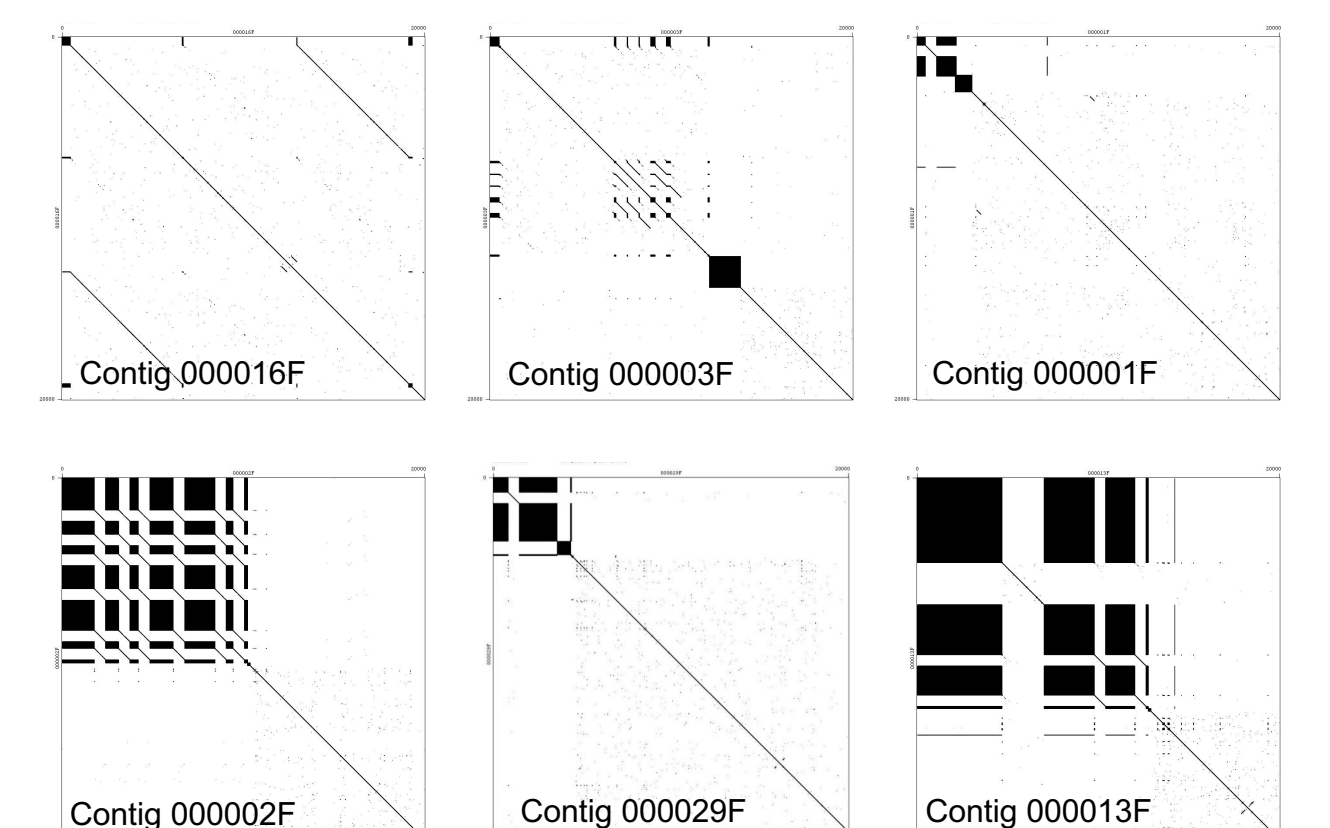
- FALCON + Purge-Dups



- Single-contig chromosomes with resolved telomere sequences revealed by alignment to *H. melpomene*



- Diversity in telomere structure



Mosquito (*Anopheles coluzzii*)



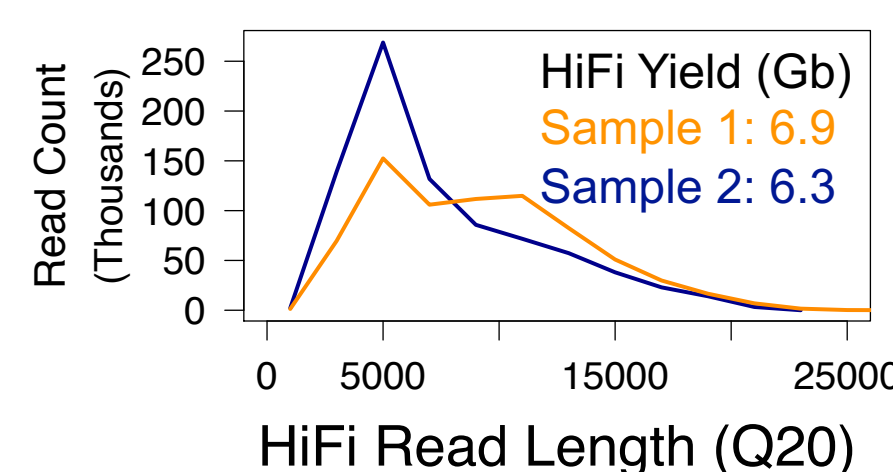
LIBRARY PREP

- Single Females collected
- DNA extraction with "10X modified" protocol³

- Low DNA Input Prep (multiplex 2 samples)
- 230 ng input DNA per sample barcode

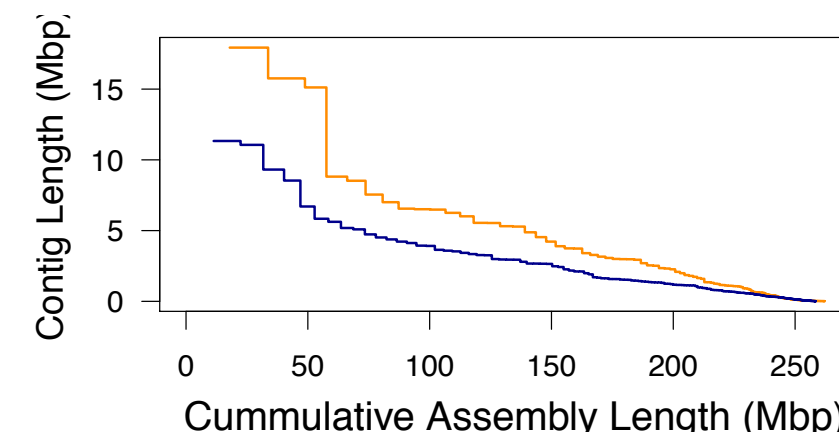


SMRT SEQUENCING



ASSEMBLY

- FALCON + Purge-Dups



Assembly is fast with HiFi reads

- Sample 1 subreads assembled and compared to HiFi assembly

Sample 1	HiFi Read Assembly	Long Read Assembly
Coverage	25-fold	40-fold
N50 Read Length (N5)	11 kb (19 kb)	12 kb (22 kb)
Primary Asm Length	262 Mb	243 Mb
Primary Contig N50	5.28 Mb	3.86 Mb
Primary Contigs	465	212
BUSCO	C:98.7%, D:0.1% F:0.6%, M:0.7%	C:98.7, D:0.2% F:0.6%, M:0.7%
CPU Hours (Consensus + Assembly)	1604	1947

HiFi assemblies capture satellites and other repeats

- 9 Mb of HiFi Read assembly does not map to Long Read assembly
- Primarily map to "UNKN" (96%) or sex chromosomes (3% Y, 1% X)
- A known satellite repeat (AgX367, L = 367 bp) maps across contig (below)

