Every species can be a model: Reference-quality PacBio genomes from single insects
Sarah B Kingan1, Christine C Lambert1, Liuqi Gu2, Katharina von Wyschetzkii3, Primo Baybayan1, Konrad Lohe4, James R Walters2, Mara KL Lawniczak2, Jonas Kortlach1

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

- HMW DNA (>40 kb fragments)
- Extraction methods vary by sample

Library Protocol Input DNA Genome Size

<table>
<thead>
<tr>
<th>Protocol</th>
<th>DNA (ng)</th>
<th>Genome Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard HiFi Library</td>
<td>15 µg</td>
<td>no limit</td>
</tr>
<tr>
<td>Low DNA Input 1 (10x)</td>
<td>400 ng</td>
<td>&lt;1 Gb</td>
</tr>
<tr>
<td>Low DNA Input 2 (10x)</td>
<td>300 ng</td>
<td>&lt;600 Mb</td>
</tr>
</tbody>
</table>

HiFi Reads
- Accuracy: >99%
- Length: 5-25 kb
- Coverage: 15 to 20-fold

HiFi Reads

- Single Female pupa collected
- DNA extraction with OmniPrep Kit

Monarch (Danaus plexippus)

- Standard HiFi Library Prep with 16 µg DNA

- Trio binning + FALCON

- Falcon + Purge-Dups

Identifying Sex Chromosomes (ZW)

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

Red Admiral (Vanessa atalanta)

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

Mosquito (Anopheles coluzzii)

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

Assembly is fast with HiFi reads
- Sample 1 subreads assembled and compared to HiFi assembly

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>HiFi Read Assembly</th>
<th>Long Read Assembly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage</td>
<td>25-bp</td>
<td>40-bp</td>
</tr>
<tr>
<td>N50 Read Length (NT)</td>
<td>11 Mb (19 kb)</td>
<td>12 kb (22 kb)</td>
</tr>
<tr>
<td>Primary Asm Length</td>
<td>262 Mb</td>
<td>243 Mb</td>
</tr>
<tr>
<td>Primary Contig N50</td>
<td>5.28 Mb</td>
<td>3.95 Mb</td>
</tr>
<tr>
<td>Primary Contigs</td>
<td>449</td>
<td>272</td>
</tr>
<tr>
<td>BUSCO</td>
<td>C:98.7%, D:0.2%</td>
<td>C:98.7%, D:0.2%</td>
</tr>
<tr>
<td>CPU Hours (Consensus + Assembly)</td>
<td>1054</td>
<td>1947</td>
</tr>
</tbody>
</table>

HiFi assemblies capture satellites and other repeats
- 9 Mb of HiFi Read assembly does not map to Long Read assembly
- Primarily to "UNK" (96%) or sex chromosomes (3%, Y, 1%) with 86.5% confidence
- HiFi reads cover repeat sequence at (4bp, L = 367 bp) maps across contig (below)

References and Acknowledgements

All protocols can be found on our website:
www.pacb.com/documentation/


PacBio Workflows:
https://github.com/sfjuan/pacbio_dups

FALCON: https://github.com/PacificBiosciences/pb-assembly

Genomescopes: https://github.com/natashaab/genomescopes

5. The authors wish to thank Michelle Vierra, Greg Concapolon, Maggie Westman, Kristo Siera, Adam Knight, Kristel Roberthaw, Vanessa Lockhart and colleagues.