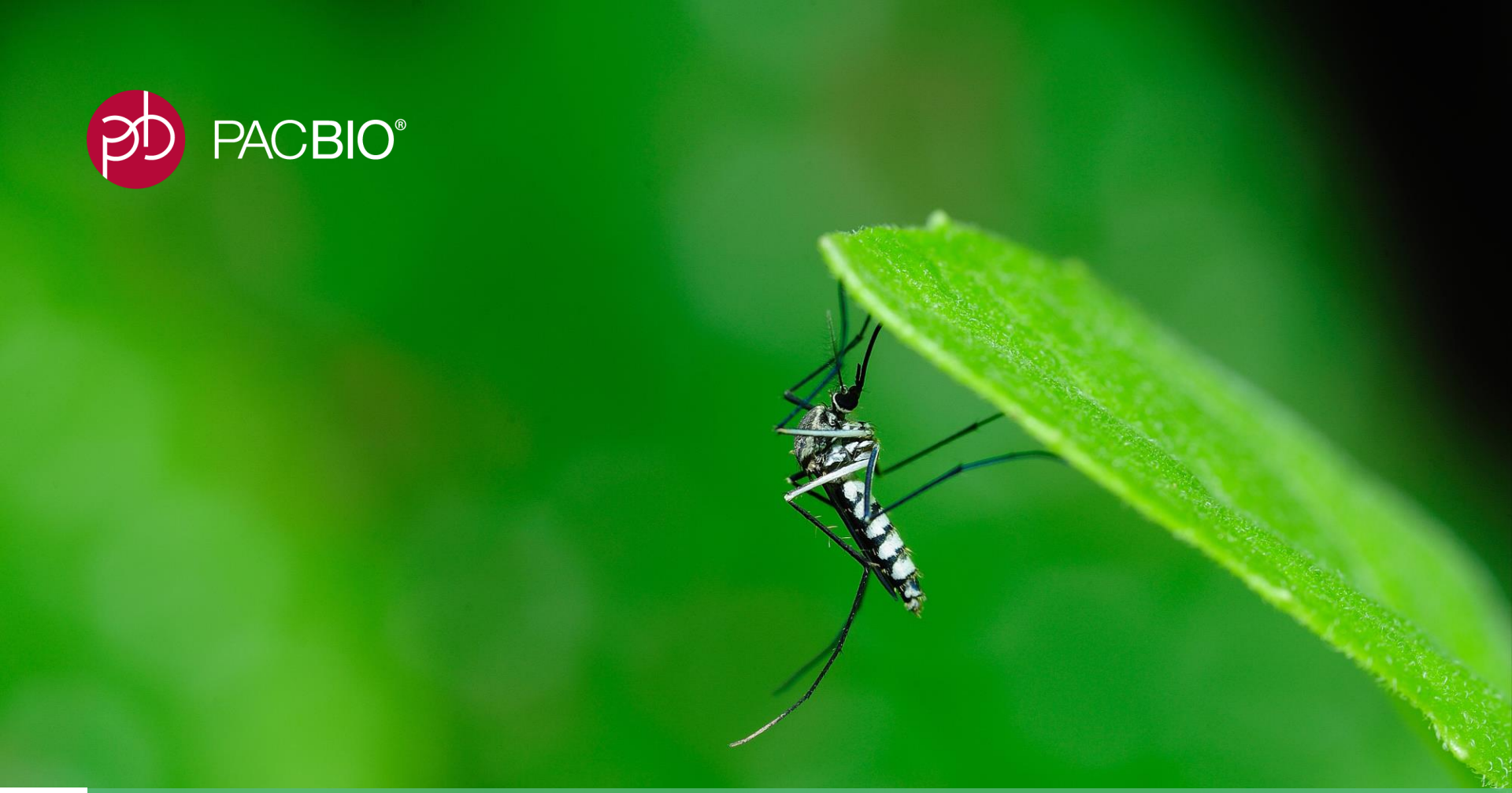


High-Quality *De Novo* Insect Genome Assemblies using PacBio Sequencing

Jonas Korlach, Sarah Kingan

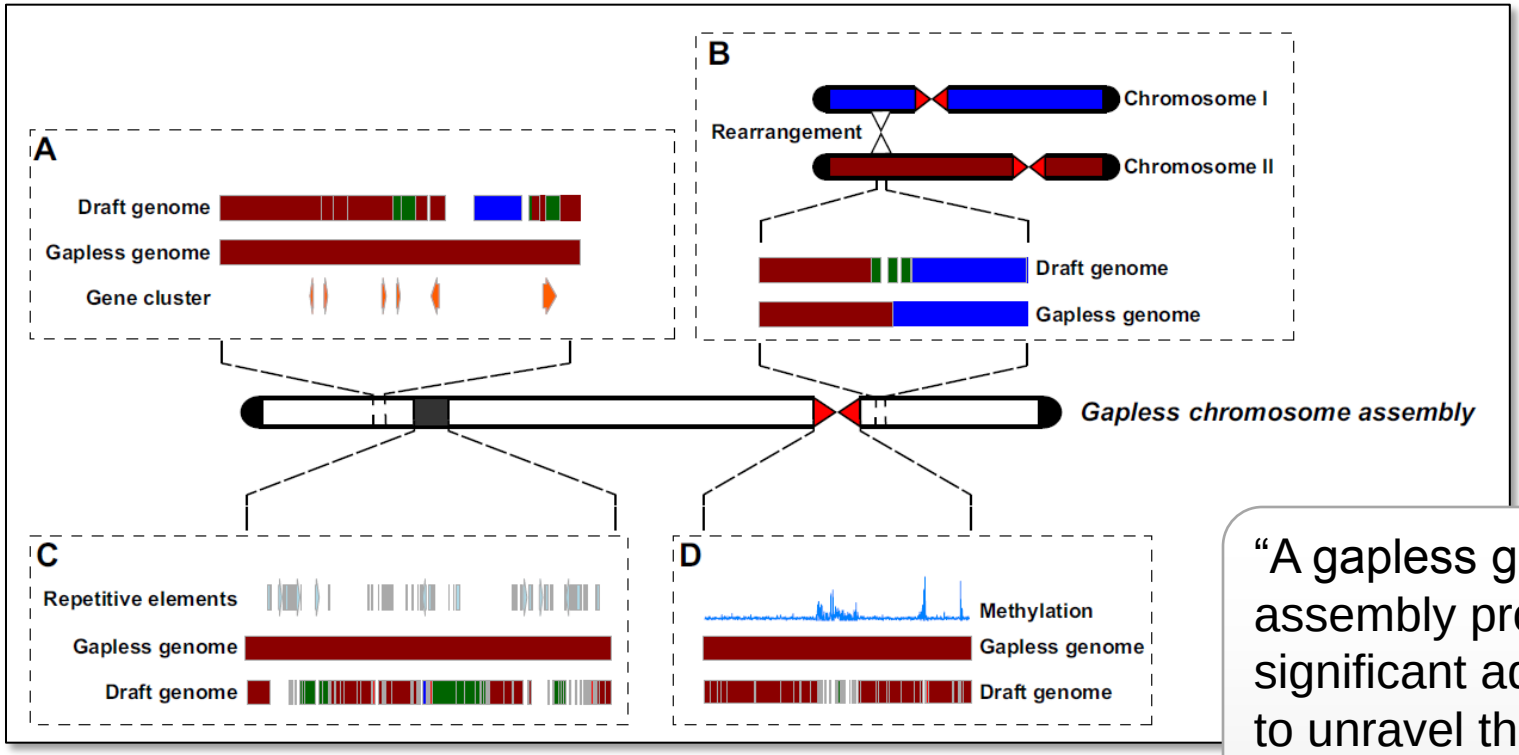
October 5, 2016

DRAFT *vs.* GAPLESS GENOMES

Mind the gap; seven reasons to close fragmented genome assemblies

Bart P.H.J. Thomma [a,*,1], Michael F. Seidl [a], Xiaoqian Shi-Kunne [a], David E. Cook [a], Melvin D. Bolton [b], Jan A.L. van Kan [a], Luigi Faino [a,1]

[a] Laboratory of Phytopathology, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[b] United States Department of Agriculture, Agricultural Research Service, Northern Crop Science Laboratory, Fargo, ND 58102-2765, USA



"A gapless genome assembly provides significant advantages to unravel the complexity of an organism's biology."

Thomma et al. (2015) FG&B 90: 24–30

WHAT IS NEEDED FOR A HIGH-QUALITY GENOME?

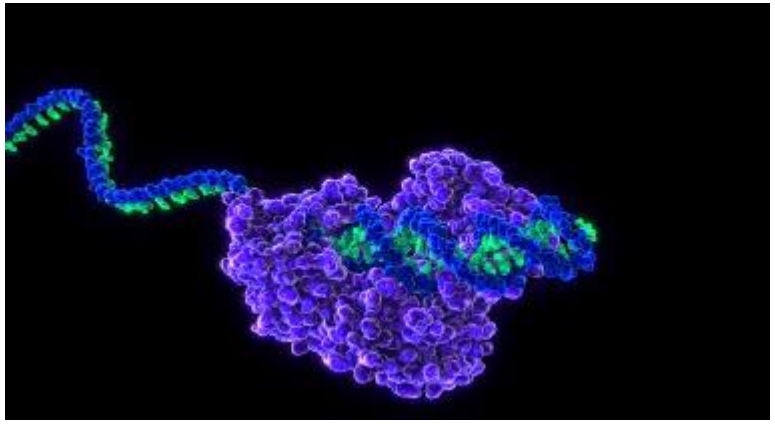
- Long sequence reads
  - Resolve repeats
  - Provides *contiguous* genome assembly
- Uniformity in sequencing coverage
  - Lack of bias (GC%, sequence complexity)
  - Allows sequencing the *entire* genome
- Absence of systematic sequencing errors
  - Random errors wash out in final consensus
  - Provides *accurate* genome sequence

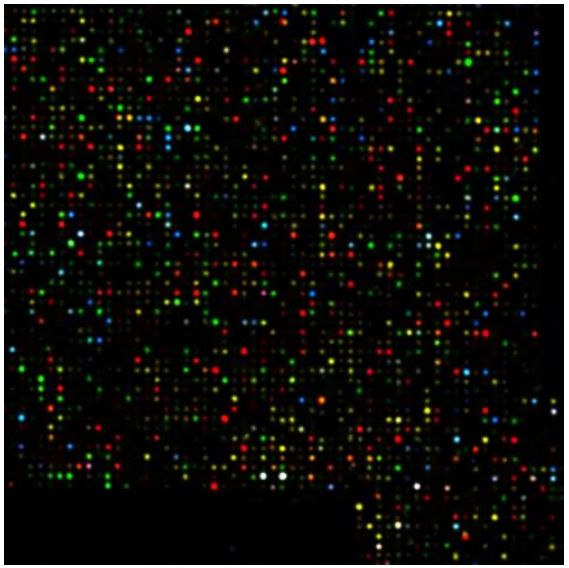
WHAT IS NEEDED FOR A HIGH-QUALITY GENOME?

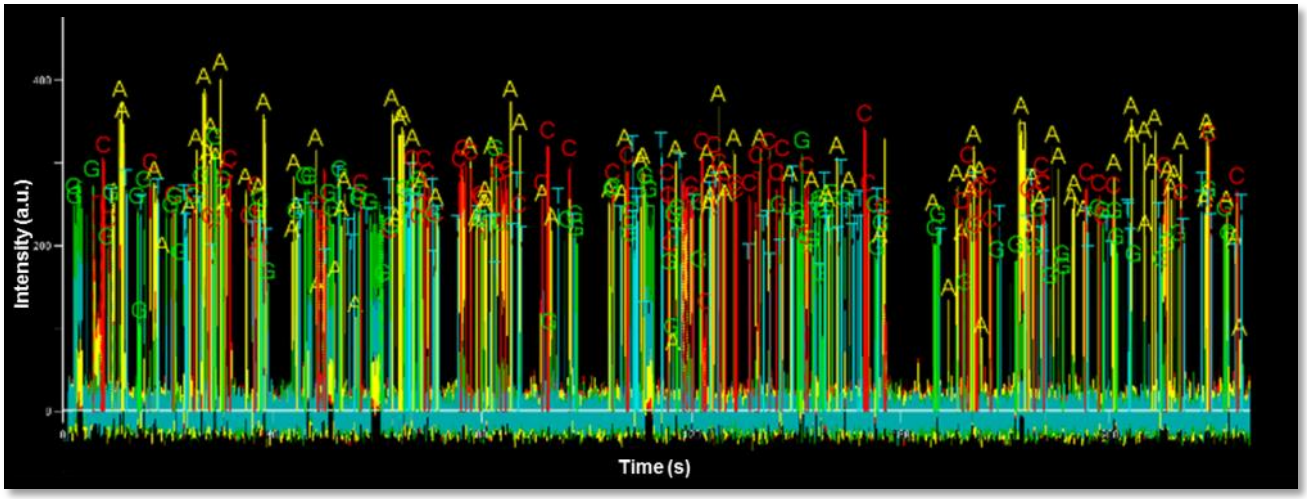
	PacBio[1]
- Long sequence reads	average >10,000 bp
- Resolve repeats	
- Provides *contiguous* genome assembly	
- Uniformity in sequencing coverage	least bias of any technology
- Lack of bias (GC%, sequence complexity)	
- Allows sequencing the *entire* genome	
- Absence of systematic sequencing errors	>99.999%
- Random errors wash out in final consensus	
- Provides *accurate* genome sequence	

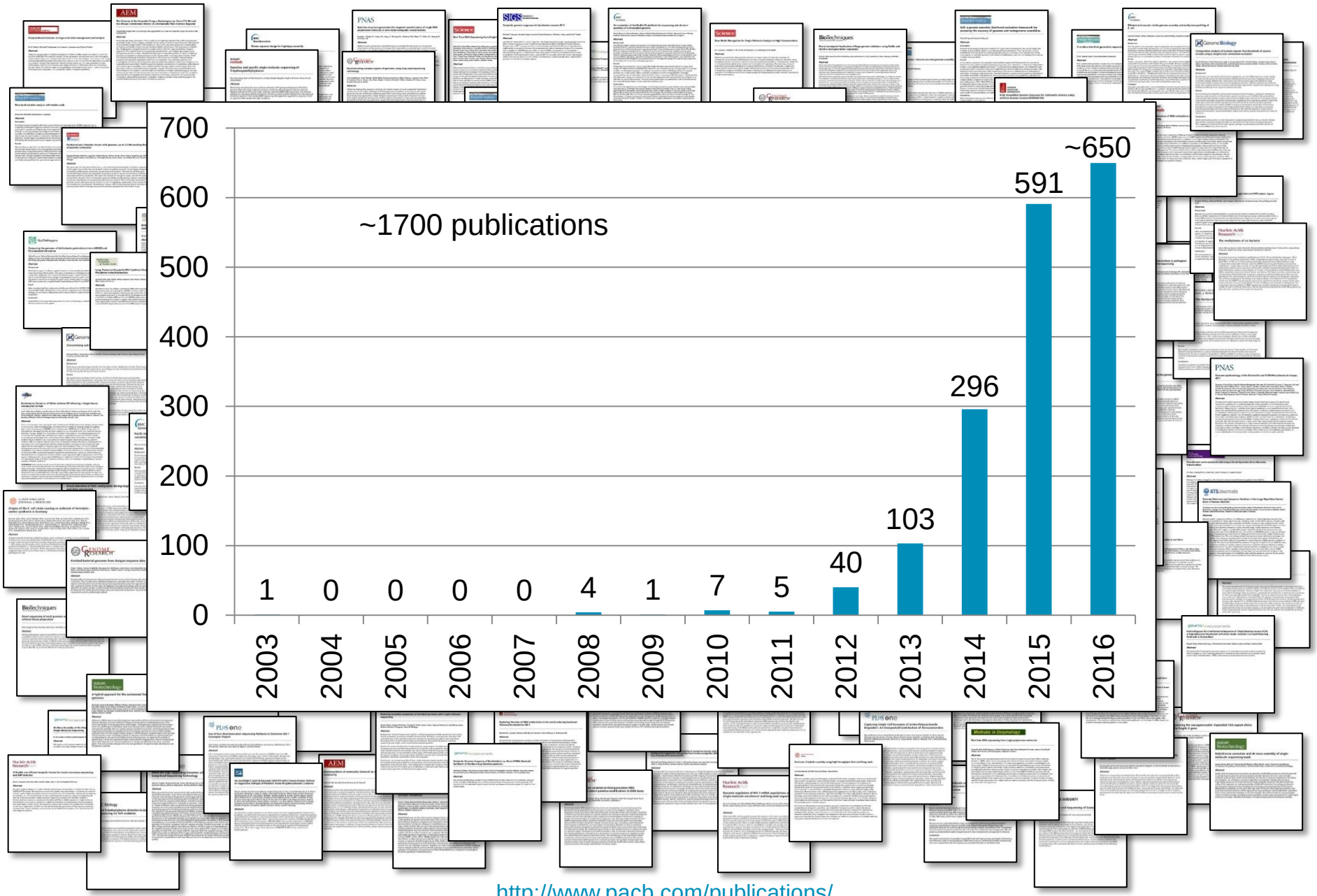
[1] http://www.pacb.com/wp-content/uploads/2015/09/Revolutionize-Genomics-with-SMRT-Sequencing.pdf

SINGLE MOLECULE, REAL-TIME (SMRT) DNA SEQUENCING









http://www.pacb.com/publications/

HIGH-QUALITY PACBIO GENOME ASSEMBLIES

UTRICULARIA GIBBA

GOAT

COW

BOS INDICUS (INDIAN CATTLE)

CABERNET SAUVIGNON

WHITE TAILED DEER

PIG

CASSAVA

HUMAN

COFFEE ARABICA

SHEEP

'ALALA

MAIZE

COTTON

ICE PLANT

HUMMINGBIRD

MEDICAGO TRUNCATULA

C. ELEGANS

SEABASS

QUINOA

RICE

ARABIDOPSIS

DROSOPHILA

MOSQUITO

SESAME

VERTICILLIUM DAHLIAE

NEOFUSICOCCUM. PARVUM

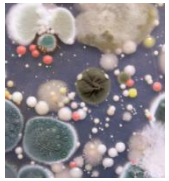
SPINACH

EUTYPA LATA

BACTERIA

YEAST

SUNFLOWER



OROPETIUM THOMAEUM

#1MbCtgClub

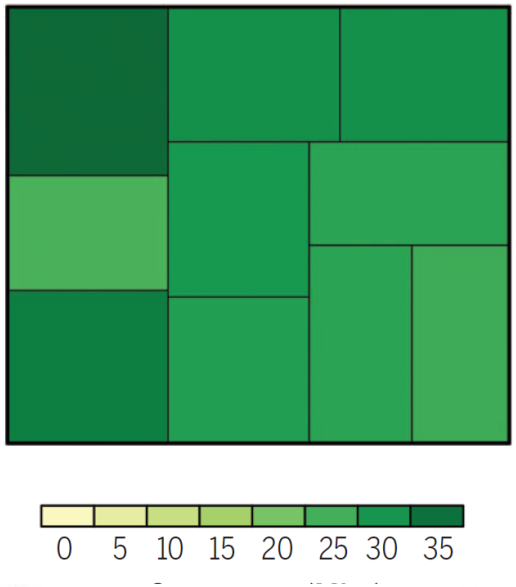
NEW GORILLA GENOME

Science, vol. 352, issue 6281.

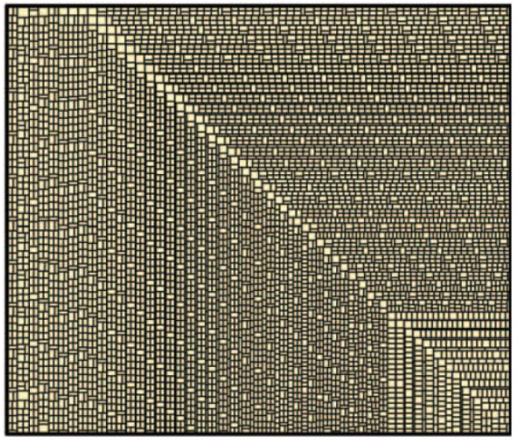
- >150fold greater contiguity
- Closure of >93% of previous gaps
- >148 Mb of new sequence
- Correction of previous errors
- 86% of structural variants are new
- Corrected previous estimates related to evolution

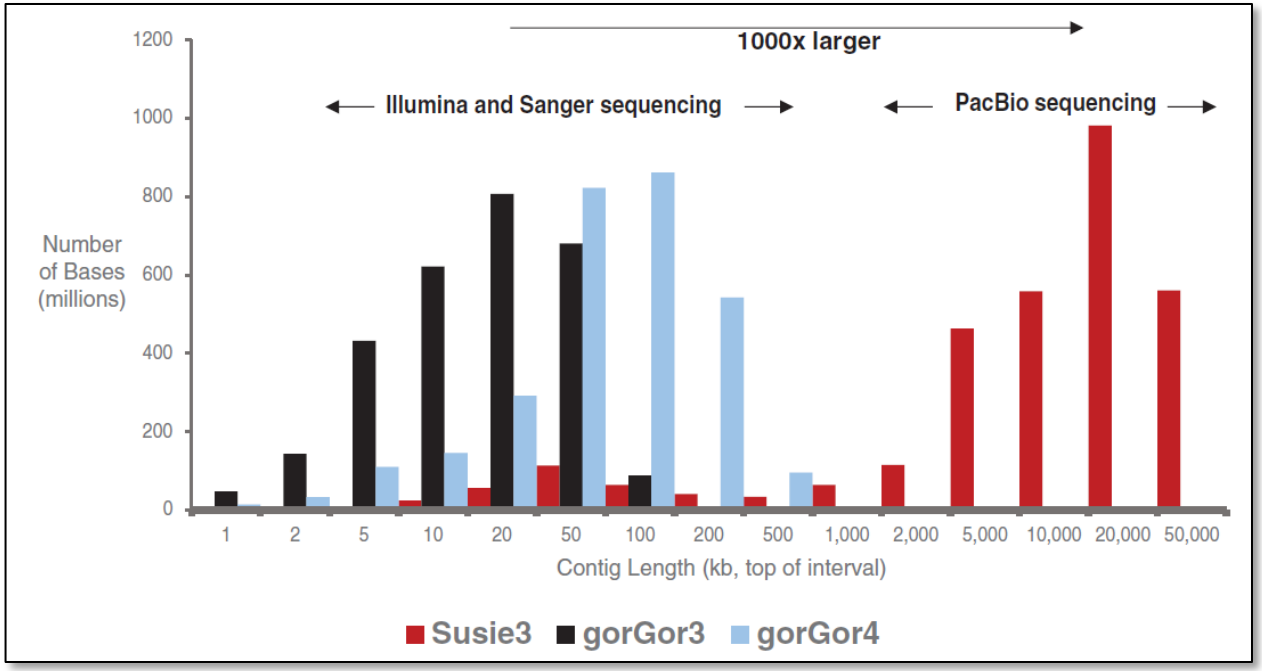
COMPARISON WITH PREVIOUS GORILLA ASSEMBLIES

B Long-read assembly (Susie3)



C Short-read assembly (gorGor3)





Gordon et al. (2016) Science 352: aae0344-1-7.

PACBIO PUBLICATIONS IN INSECT RESEARCH

- Bats may eat diurnal flies that rest on wind turbines
- Best practices in insect genome sequencing: What works and what doesn't.
- Birth of a new gene on the Y chromosome of Drosophila melanogaster.
- Buzz off, that's my bee!
- Comparative genome analysis of Wolbachia strain wAu
- Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge.
- Complete genome of Serratia sp. strain FGI 94, a strain associated with leaf-cutter ant fungus gardens.
- Complete genome sequence of Burkholderia sp. strain RPE64, bacterial symbiont of the bean bug Riptortus pedestris.
- Complete genome sequence of Endomicrobium proavitum, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum Elusimicrobia).
- Differing patterns of selection and geospatial genetic diversity within two leading Plasmodium vivax candidate vaccine antigens.
- DNA methylation on N6-adenine in C. elegans.
- Evolution of mosquito preference for humans linked to an odorant receptor.
- Genome analysis of a major urban malaria vector mosquito, Anopheles stephensi.
- Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia.
- Genome of Cnaphalocrocis medinalis granulovirus, the first Crambidae-infecting betabaculovirus isolated from rice leaffolder to sequenced.
- Genome sequence of the Drosophila melanogaster male-killing Spiroplasma strain MSRO endosymbiont.
- Genomes of 'Candidatus Liberibacter solanacearum' Haplotype A from New Zealand and the United States Suggest Significant Genome Plasticity in the Species.
- Genomics and host specialization of honey bee and bumble bee gut symbionts.
- Gut symbionts from distinct hosts exhibit genotoxic activity via divergent colibactin biosynthetic pathways.
- Insights into the preservation of the homomorphic sex-determining chromosome of Aedes aegypti from the discovery of a male-biased gene tightly linked to the M-locus.
- Landscape of standing variation for tandem duplications in Drosophila yakuba and Drosophila simulans.
- Large-scale mitogenomics enables insights into Schizophora (Diptera) radiation and population diversity.
- Long-read single molecule sequencing to resolve tandem gene copies: The Mst77Y region on the Drosophila melanogaster Y chromosome.
- Major improvements to the Heliconius melpomene genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution.
- Microplitis demolitor bracovirus proviral loci and clustered replication genes exhibit distinct DNA amplification patterns during replication.
- PacBio full-length transcriptome profiling of insect mitochondrial gene expression.
- Paenibacillus larvae-directed bacteriophage HB10c2 and its application in American Foulbrood-affected honey bee larvae.
- Quantitative profiling of Drosophila melanogaster Dscam1 isoforms reveals no changes in splicing after bacterial exposure.
- Selections that isolate recombinant mitochondrial genomes in animals.
- Single molecule long read sequencing resolves the detailed structure of complex satellite DNA loci in Drosophila melanogaster
- Site-specific genetic engineering of the Anopheles gambiae Y chromosome.
- Structural changes following the reversal of a Y chromosome to an autosome in Drosophila pseudoobscura
- Structure of the type IV secretion system in different strains of Anaplasma phagocytophilum.
- The bacterial microbiome of Dermacentor andersoni ticks influences pathogen susceptibility.
- The genome and methylome of a beetle with complex social behavior, Nicrophorus vespilloides (Coleoptera: Silphidae).
- The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera.
- The industrial melanism mutation in British peppered moths is a transposable element.
- The mitochondrial genome of a Texas outbreak strain of the cattle tick, Rhipicephalus (Boophilus) microplus, derived from whole genome sequencing Pacific Biosciences and Illumina reads.
- Unique transposon landscapes are pervasive across Drosophila melanogaster genomes.
- Variation and evolution in the glutamine-rich repeat region of Drosophila argonaute-2.
- Whole-genome sequence of Bacillus sp. SDLI1, isolated from the social bee Scaptotrigona depilis.
- Whole-genome sequence of Burkholderia sp. strain RPE67, a bacterial gut symbiont of the bean bug Riptortus pedestris.
- Whole-genome sequence of Erysipelothrix larvae LV19(T) (=KCTC 33523(T)), a useful strain for arsenic detoxification, from the larval gut of the rhinoceros beetle, Trypoxylus dichotomus.
- Whole-genome sequence of Serratia symbiotica strain CWBI-2.3T, a free-living symbiont of the black bean aphid Aphis fabae.

http://www.pacb.com/publications/

DROSOPHILA SPECIES ASSEMBLIES

ARTICLES

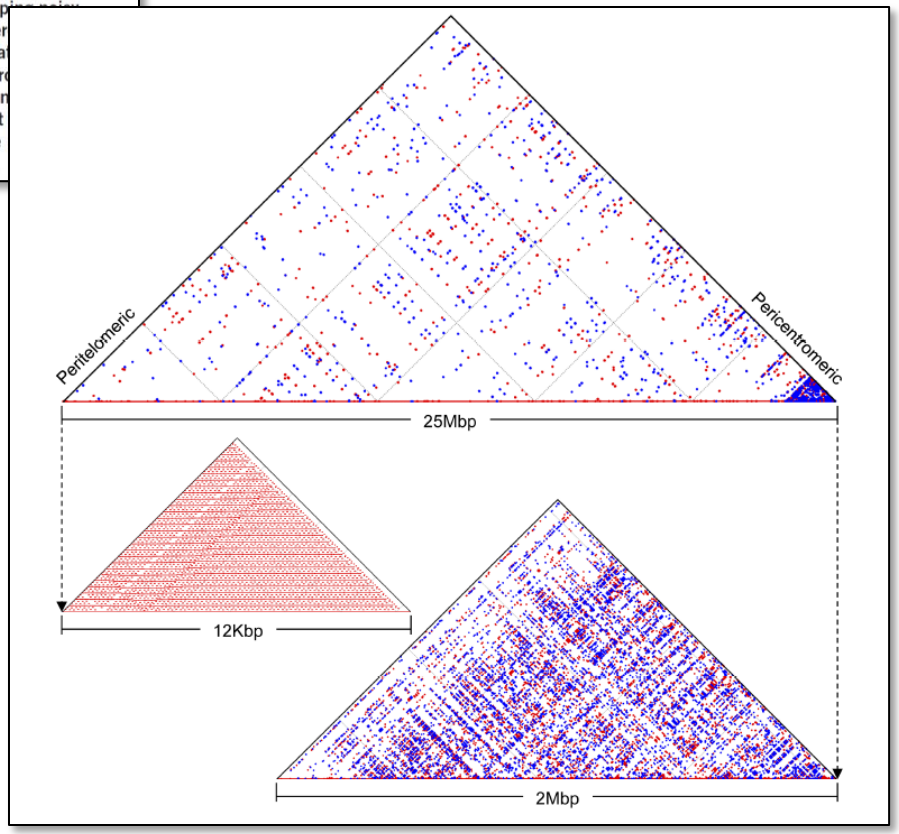
nature biotechnology

Assembling large genomes with single-molecule sequencing and locality-sensitive hashing

Konstantin Berlin[1–3,6], Sergey Koren[4,6], Chen-Shan Chin[5], James P Drake[5], Jane M Landolin[5] & Adam M Phillippy[4]

Long-read, single-molecule real-time (SMRT) sequencing is routinely used to finish microbial genomes, but available assembly methods have not scaled well to larger genomes. We introduce the MinHash Alignment Process (MHAP) for overlapp long reads using probabilistic, locality-sensitive hashing. Integrating MHAP with the Celera Assembler enabled refer *de novo* assemblies of *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Drosophila melanogaster* and a human hyda cell line (CHM1) from SMRT sequencing. The resulting assemblies are highly continuous, include fully resolved chr arms and close persistent gaps in these reference genomes. Our assembly of *D. melanogaster* revealed previously u heterochromatic and telomeric transition sequences, and we assembled low-complexity sequences from CHM1 that in the human GRCh38 reference. Using MHAP and the Celera Assembler, single-molecule sequencing can produce near-complete eukaryotic assemblies that are 99.99% accurate when compared with available reference genomes.

- Highly contiguous, including complete chromosome arms
- Filled persistent gaps in version 5 reference
- Revealed previously unknown heterochromatic and telomeric transition sequences



Single-contig assembly of *D. melanogaster* chromosome arm 3L

NEW INSIGHTS FROM PACBIO *DROSOPHILA* ASSEMBLIES

Long-Read Single Molecule Sequencing to Resolve Tandem Gene Copies: The *Mst77Y* Region on the *Drosophila melanogaster* Y Chromosome

Flavia J. Krsticevic,* Carlos G. Schrago,† and A. Bernardo Carvalho†,1
*Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas, CONICET, Ocampo y Esmeralda, S2000EZP Rosario, Argentina, and †Departamento de Genética, Universidade Federal do Rio de Janeiro, 21941-971, Rio de Janeiro, Brazil

Nucleic Acids Research, 2015 1
doi: 10.1093/nar/gkv1193

Unique transposon landscapes are pervasive across *Drosophila melanogaster* genomes

Reazur Rahman[1], Gung-wei Chirn[1], Abhay Kanodia[1], Yuliya A. Sytnikova[1], Björn Brembs[2], Casey M. Bergman[3] and Nelson C. Lau[1,*]

[1]Department of Biology and Rosenstiel Basic Medical Science Research Center, Brandeis University, Waltham, MA 02454, USA, [2]Institute of Zoology, Universität Regensburg, Regensburg, Germany and [3]Faculty of Life Sciences, University of Manchester, Manchester M21 0RG, UK

PNAS

Birth of a new gene on the Y chromosome of *Drosophila melanogaster*

Antonio Bernardo Carvalho[a,1], Beatriz Vicoso[a,b], Claudia A. M. Russo[a], Bonnielin Swenor[c], and Andrew G. Clark[c,1]

[a]Departamento de Genética, Universidade Federal do Rio de Janeiro, Caixa Postal 68011, 21941-971, Rio de Janeiro, Brazil; [b]Institute of Science and Technology Austria, A-3400 Klosterneuburg, Austria; and [c]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853

Variation and Evolution in the Glutamine-Rich Repeat Region of *Drosophila* Argonaute-2

William H. Palmer[1] and Darren J. Obbard
Institute of Evolutionary Biology and Centre for Infection, Evolution and Immunity, University of Edinburgh, EH9 3FL UK

Structural changes following the reversal of a Y chromosome to an autosome in *Drosophila pseudoobscura*

Ching-Ho Chang, Amanda M Larracuente
doi: http://dx.doi.org/10.1101/058412

Single molecule long read sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*

Daniel Emerson Khost, Danna G Eickbush, Amanda M Larracuente
doi: http://dx.doi.org/10.1101/054155

Krsticevic et al. doi: 10.1534/g3.115.017277; Carvalho et al. www.pnas.org/cgi/doi/10.1073/pnas.1516543112; Palmer and Obbard doi: 10.1534/g3.116.031880;

OVERVIEW

I. PacBio workflow

- Sample prep
- Sequencing
- Genome assembly

II. *Aedes aegypti* project

- Generating the PacBio assembly
- Quality and completeness
  - Assembly statistics
  - Phasing of allelic haplotypes
  - Full-length, base-pair resolution of insecticide-resistance gene

SIMPLE SCALABLE WORKFLOW

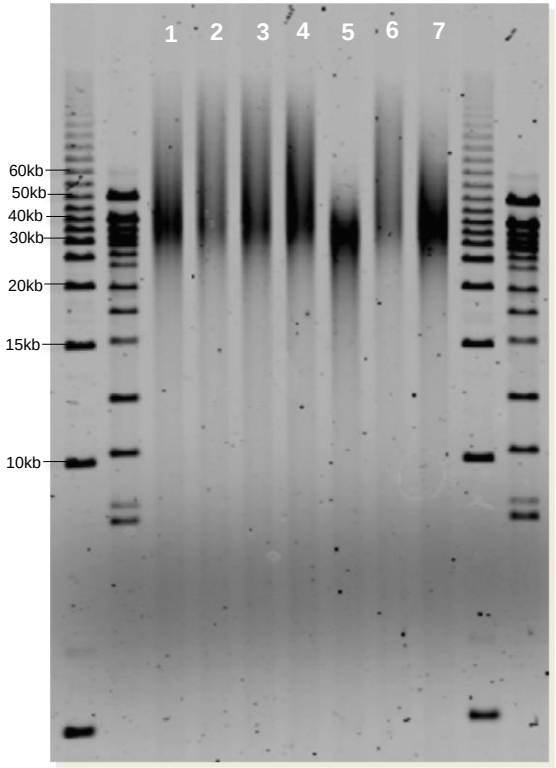
- SMRT Sequencing workflow is the same for all PacBio systems
  - Amplicons, RNA or **large genomes**

LIBRARY CONSTRUCTION → PRIMER ANNEALING AND POLY BINDING → LOADING → INSTRUMENT RUN → PRIMARY ANALYSIS → SECONDARY ANALYSIS

IMPORTANCE OF HIGH QUALITY TEMPLATE DNA

HMW DNA → Long Reads

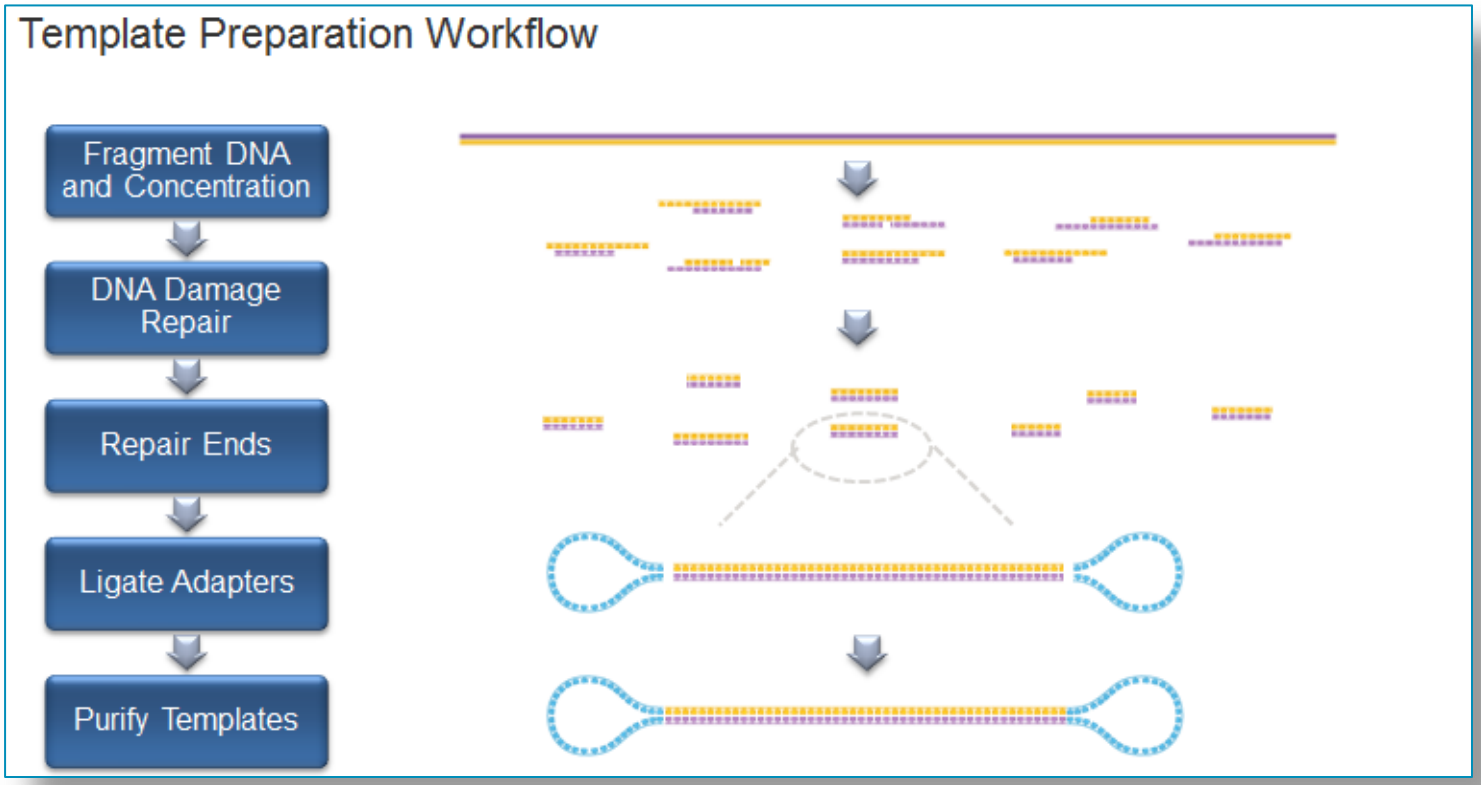
- DNA shearing
  - Megaruptor
  - Covaris g-TUBE
  - Needle shearing
- 10 ug for 20 kb library
- Multiple cells per library



DNA sheared with Megaruptor (Diagenode)

SMRTbell LIBRARY CONSTRUCTION

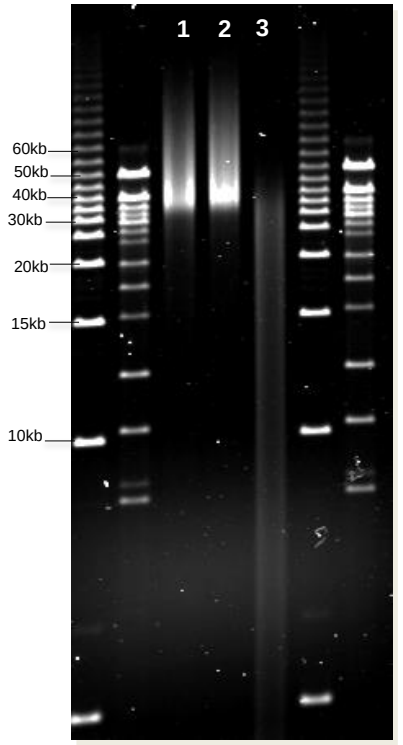
LIBRARY CONSTRUCTION | PRIMER ANNEALING AND POLY BINDING | LOADING | INSTRUMENT RUN | PRIMARY ANALYSIS | SECONDARY ANALYSIS



Library construction takes approximately 3-4 hours

SIZE SELECTION OF SMRTbell LIBRARY

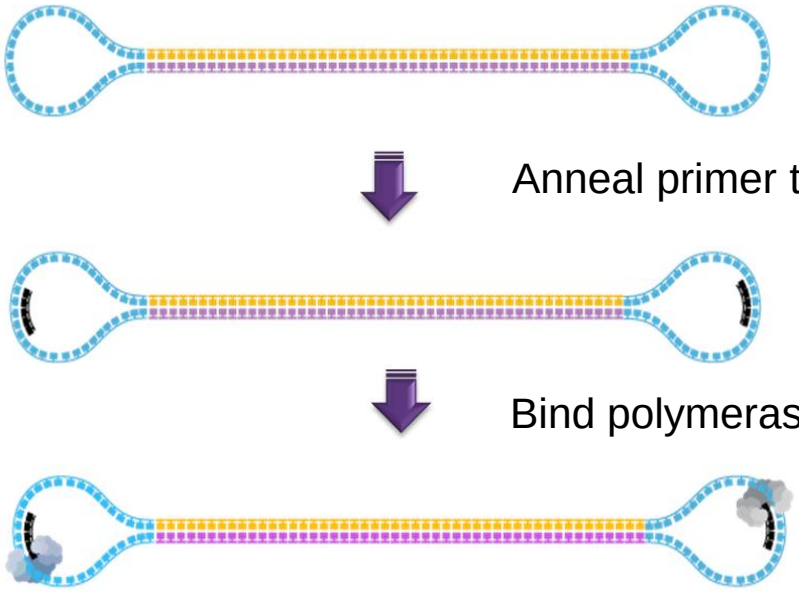
- Removes smaller library fragments
- Enables reads up to 60 kb long
- 20 – 30 kb size selection
- BluePippin or ELF (SageScience)



30 kb SMRTbell Library

ANNEAL PRIMER AND BIND POLYMERASE TO SMRTbell TEMPLATES

LIBRARY CONSTRUCTION | **PRIMER ANNEALING AND POLY BINDING** | LOADING | INSTRUMENT RUN | PRIMARY ANALYSIS | SECONDARY ANALYSIS



Anneal primer to both ends of the SMRTbell template

Bind polymerase to both ends of the SMRTbell template

PACBIO RS II AND SEQUEL SYSTEMS

	PacBio RS II	Sequel
Machine Launch	2013	2015
Number ZMW per Cell	**150,000**	**1,000,000**
Output per Cell	0.5 – 1 Gb	~ 5 Gb
Input DNA per Cell	~ 250 ng	~ 250 ng
N50 Read Length	15 – 20 kb	~15 kb
Consensus accuracy	QV50	QV50

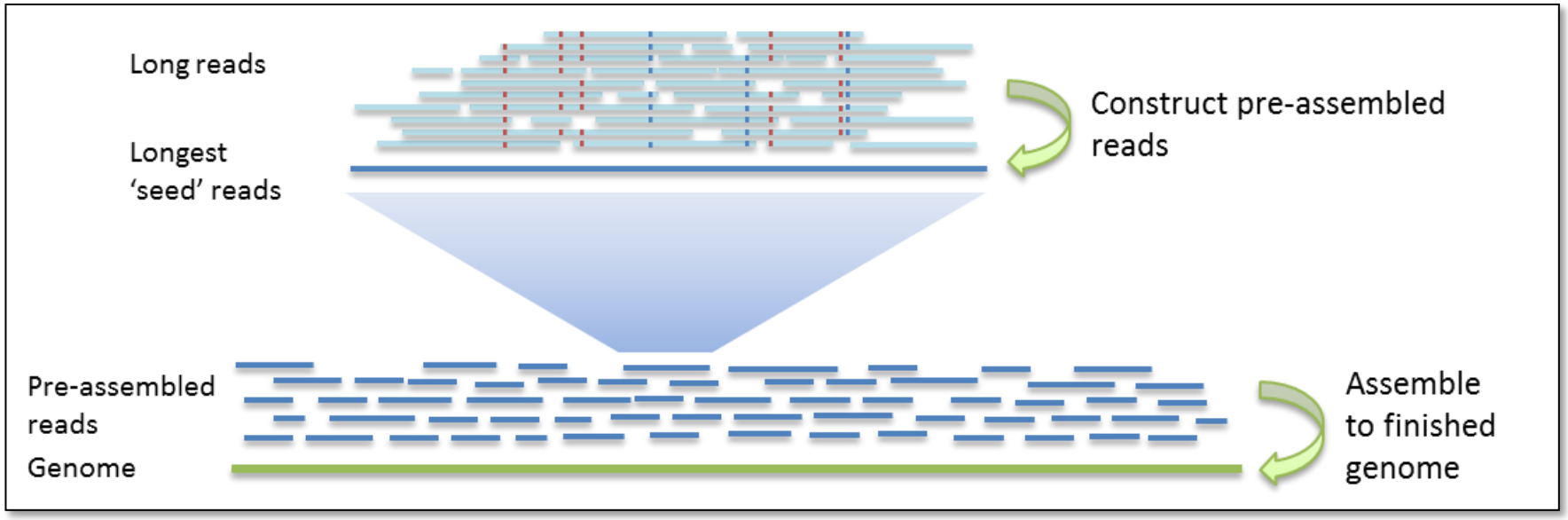
LARGE GENOME SEQUENCING WITH SEQUEL

LARGE GENOME SEQUENCING WITH SEQUEL

Genome Length	120 Mb
Total BP	10.8 Gb (90-fold)
Number Cells	2
Read N50	16.4 kb
Input per Cell	0.25 ug

Assembly Length	122.9 Mb
Number Contigs	238
Contig N50	10.4 Mb
Max Contig	15.0 Mb

HIERARCHICAL GENOME ASSEMBLY PROCESS (HGAP)



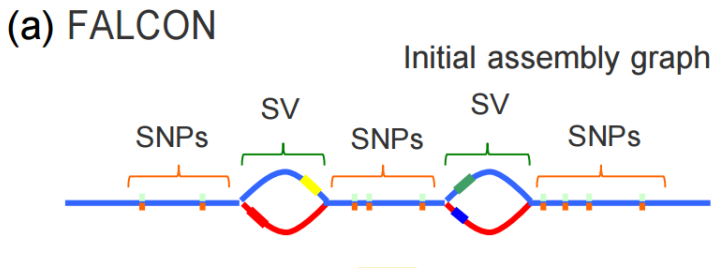
Chin et al. (2013) Nature Methods 10, 563–569

DIPLOID ASSEMBLY WITH FALCON-UNZIP

Jason Chin, PacBio

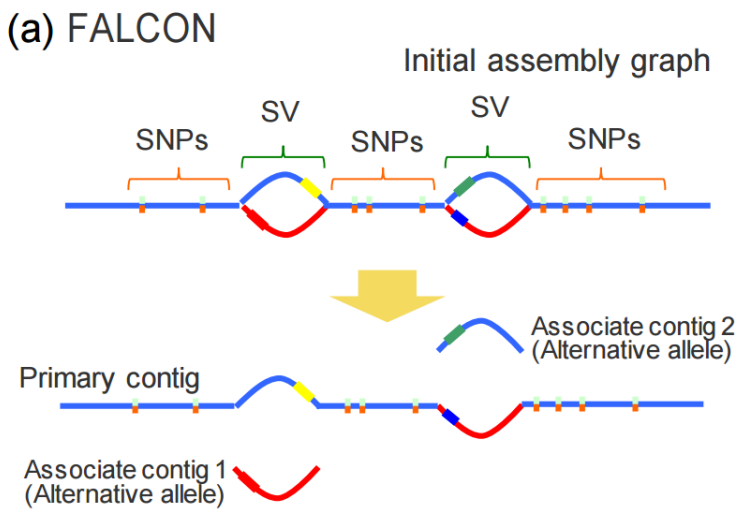
Chin et al. 2016 Nature Methods, accepted. http://biorxiv.org/content/early/2016/06/03/056887

DIPLOID ASSEMBLY WITH FALCON-UNZIP



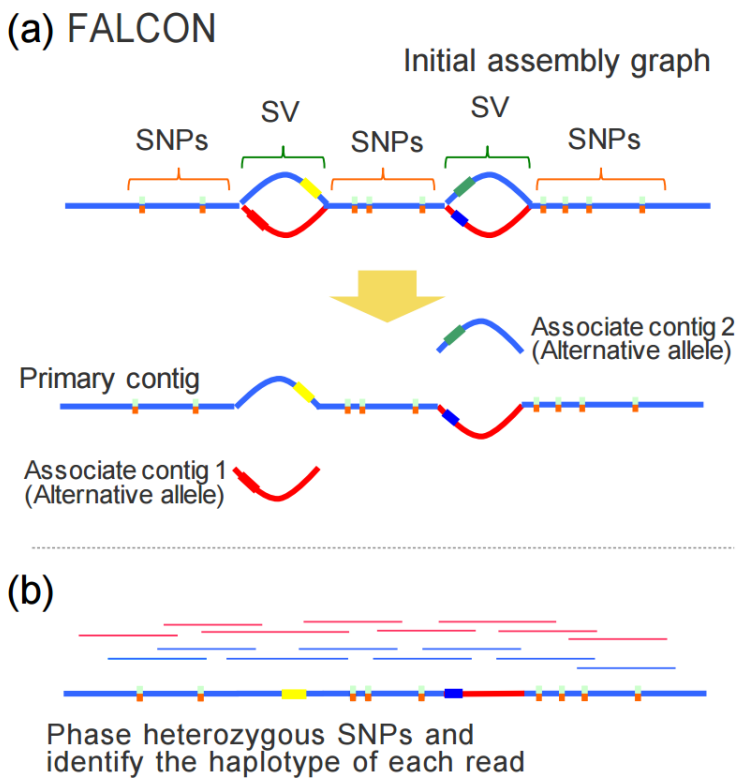
Chin et al. 2016 Nature Methods, accepted. http://biorxiv.org/content/early/2016/06/03/056887

DIPLOID ASSEMBLY WITH FALCON-UNZIP



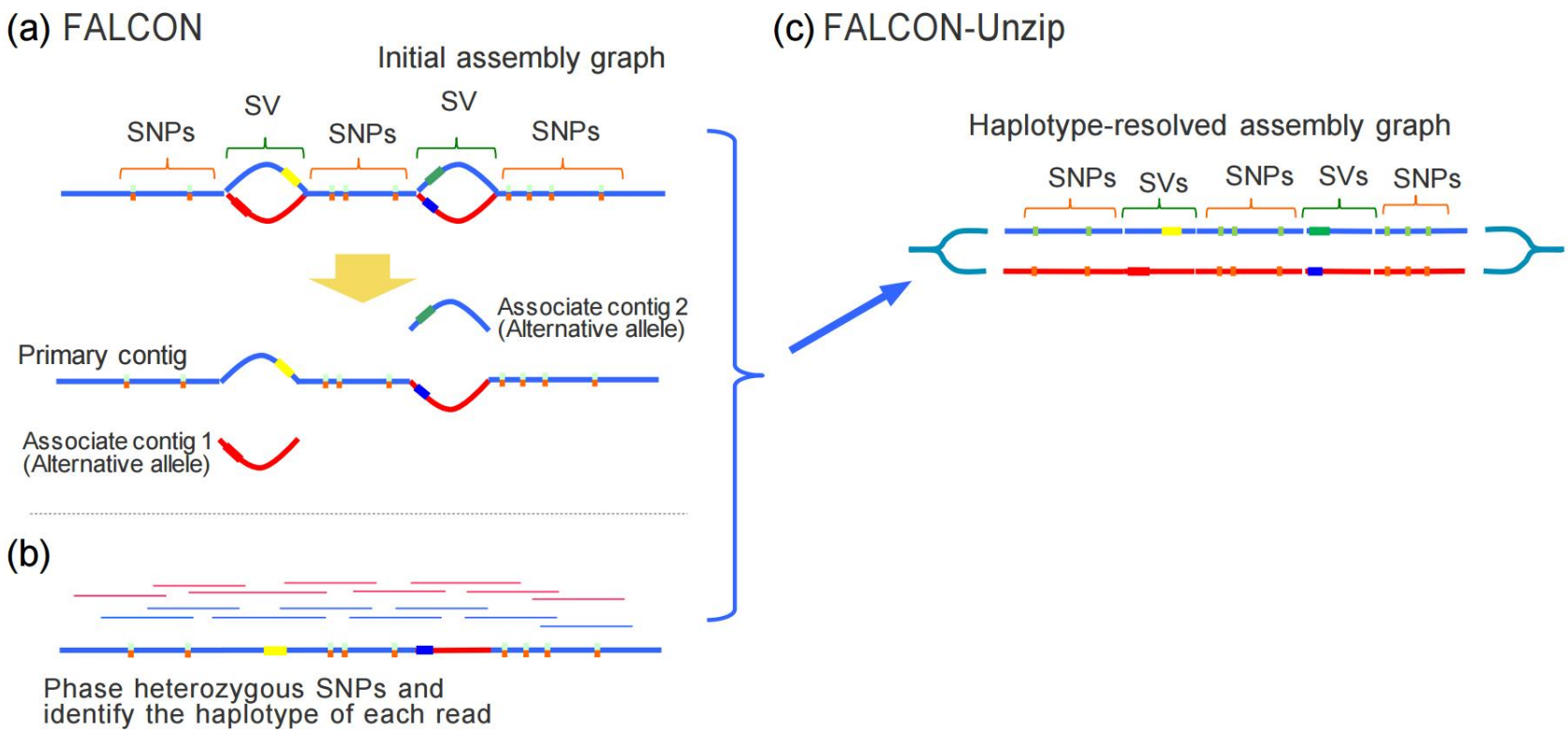
Chin et al. 2016 Nature Methods, accepted. http://biorxiv.org/content/early/2016/06/03/056887

DIPLOID ASSEMBLY WITH FALCON-UNZIP



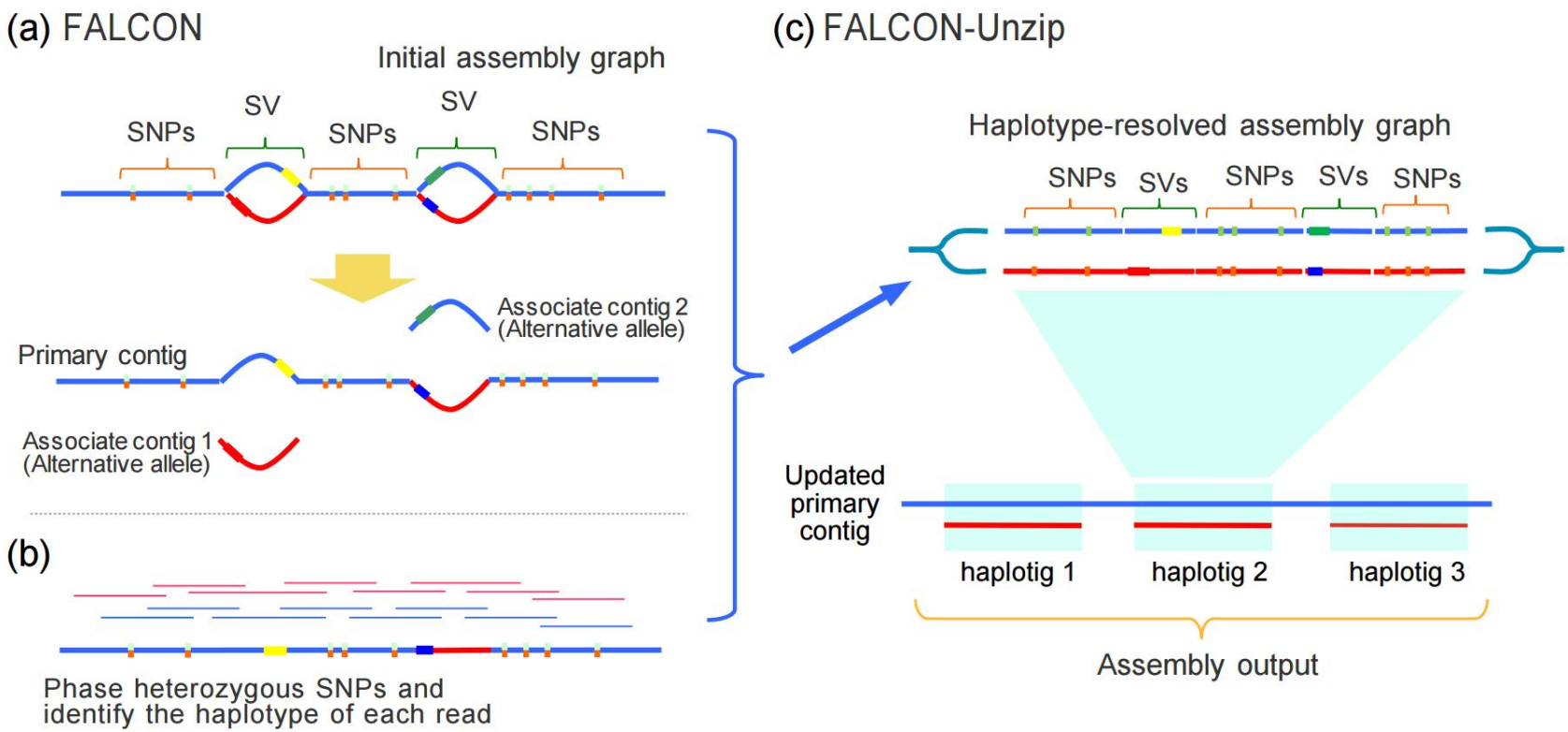
Chin et al. 2016 Nature Methods, accepted. http://biorxiv.org/content/early/2016/06/03/056887

DIPLOID ASSEMBLY WITH FALCON-UNZIP



Chin et al. 2016 Nature Methods, accepted. http://biorxiv.org/content/early/2016/06/03/056887

DIPLOID ASSEMBLY WITH FALCON-UNZIP

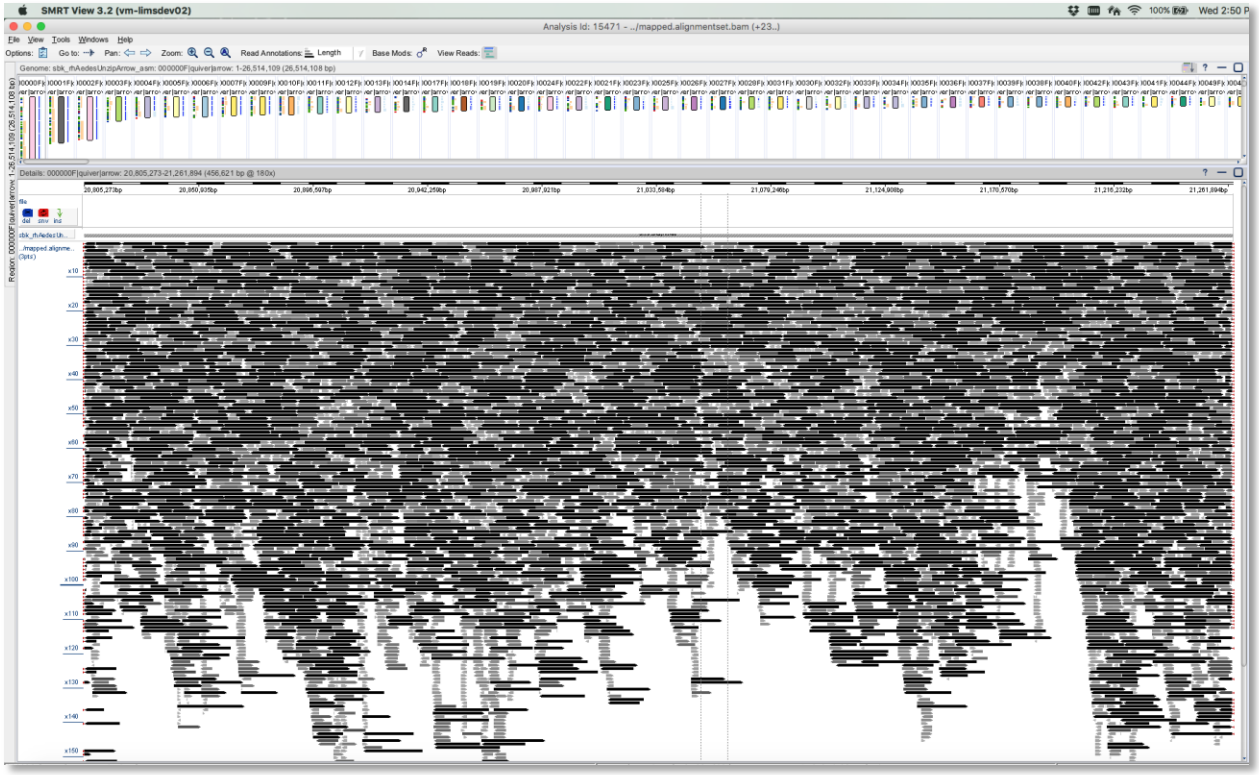


Chin et al. 2016 Nature Methods, accepted. http://biorxiv.org/content/early/2016/06/03/056887

GENOME POLISHING

Quiver and Arrow Algorithms

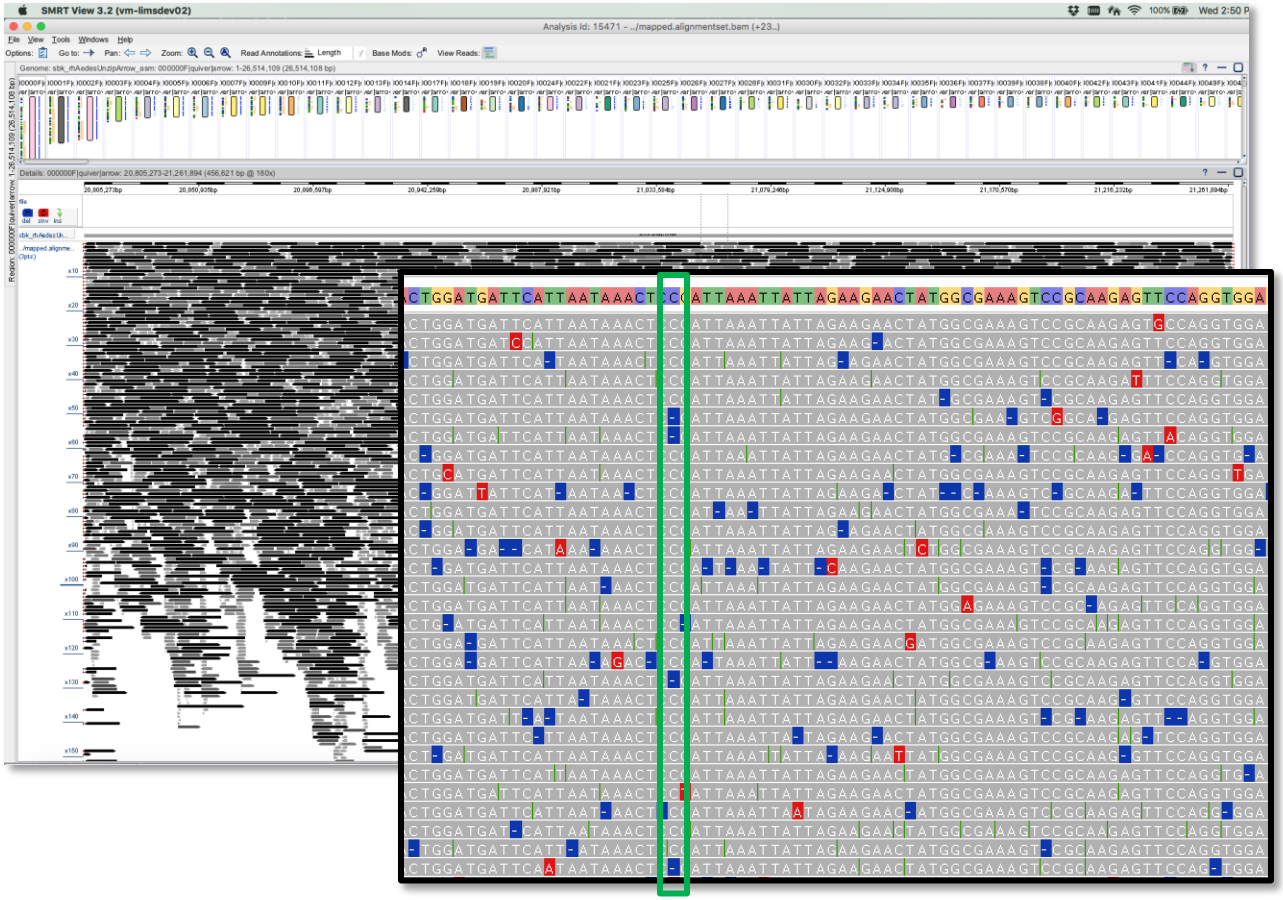
- Map raw reads back to genome sequence
- Compute consensus base and base quality
- Hidden Markov Model, trained on sequencing chemistry characteristics



GENOME POLISHING

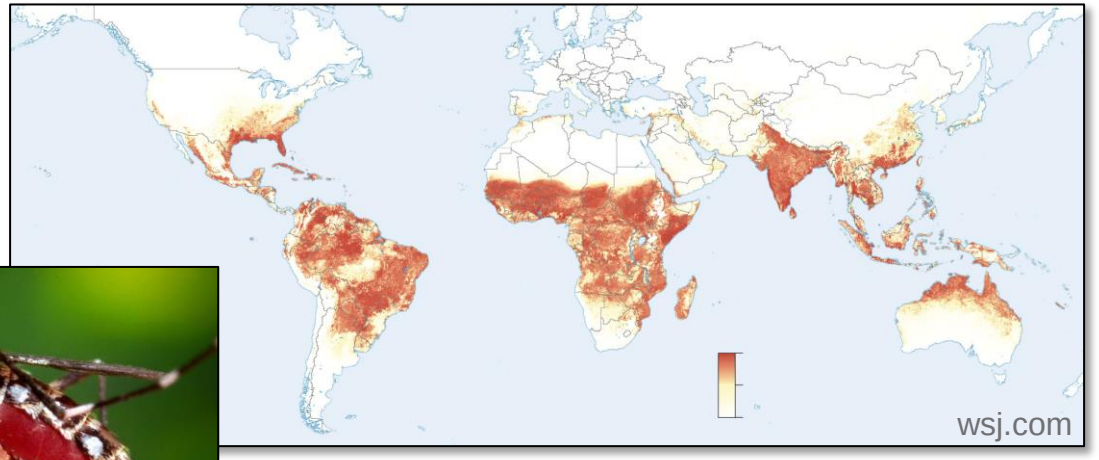
Quiver and Arrow Algorithms

- Map raw reads back to genome sequence
- Compute consensus base and base quality
- Hidden Markov Model, trained on sequencing chemistry characteristics



AEDES AEGYPTI GENOME WORKING GROUP (AGWG)

- Led by Drs. Leslie Vosshall and Ben Matthews, HHMI, Rockefeller University
- Dozens of academic researchers and private companies

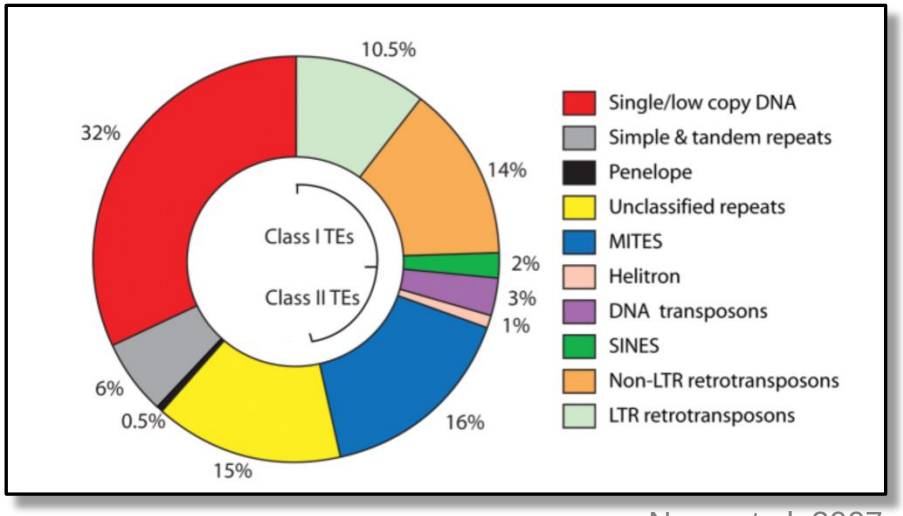


- Vector for yellow fever, Zika, dengue and other diseases
- Tropical and semi-tropical

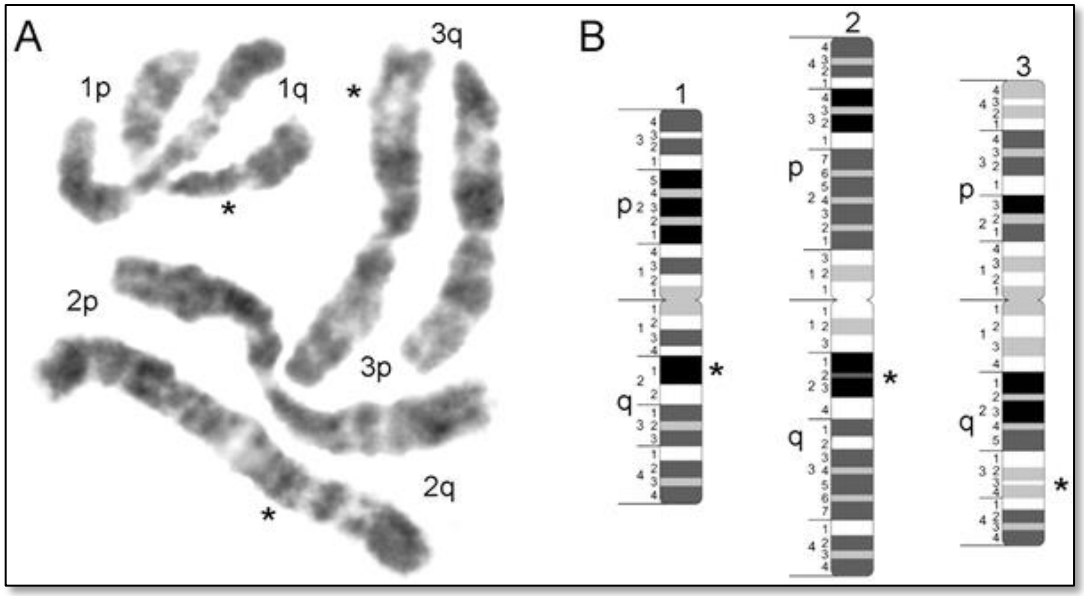
- Ancestral subspecies: *Aedes aegypti formosus* from Africa
- Anthropophilic *A. a. aegypti* dispersed by humans
- High selective pressure due to eradication efforts

AEDES AEGYPTI GENOME

- 3 metacentric chromosomes
  - Homomorphic sex chromosomes
- 1.3 Gb
  - 5 times the length of *Anopheles gambiae*
- Highly repetitive



Nene et al. 2007



Timoshevskiy et al. 2013

CURRENT REFERENCE

Nene et al. 2007

- Published in June, 2007
- Sanger-based assembly of Liverpool inbred strain
- Collaborative effort of private and public entities
- Remains fragmented and low quality

AEDES AEGYPTI PACBIO ASSEMBLY

Overview

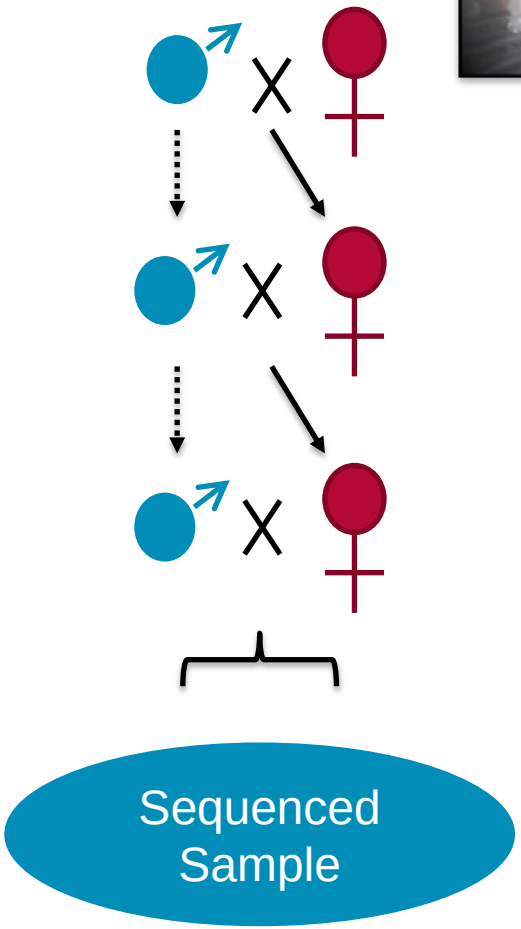
- Biological Sample
- Library Preparation
- Sequencing
- Diploid Genome Assembly
- Assessment of Genome Quality and Completeness
- Long-range Haplotype Resolution
- Full-length, Base-Pair Resolution of Insecticide Resistance Gene

BIOLOGICAL SAMPLE

Swift, NYTimes, May 2016

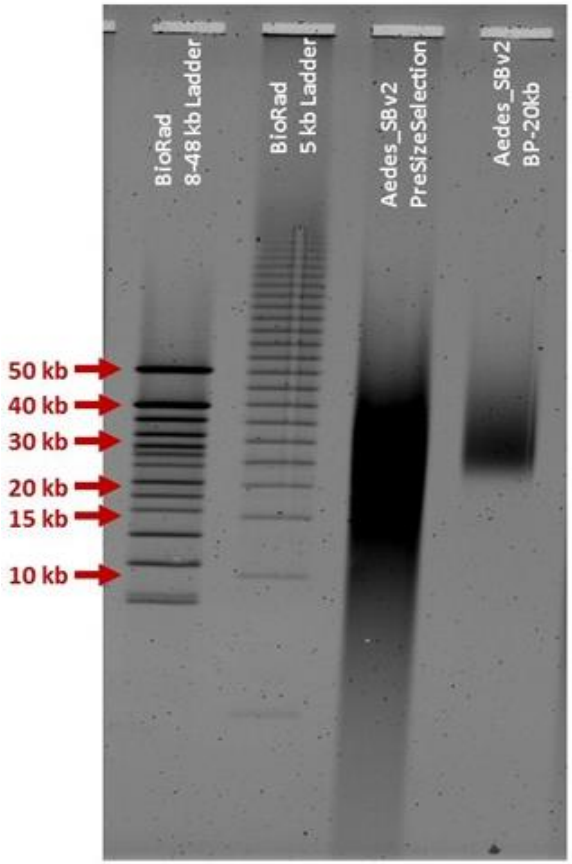
Ben Matthews, Rockefeller University

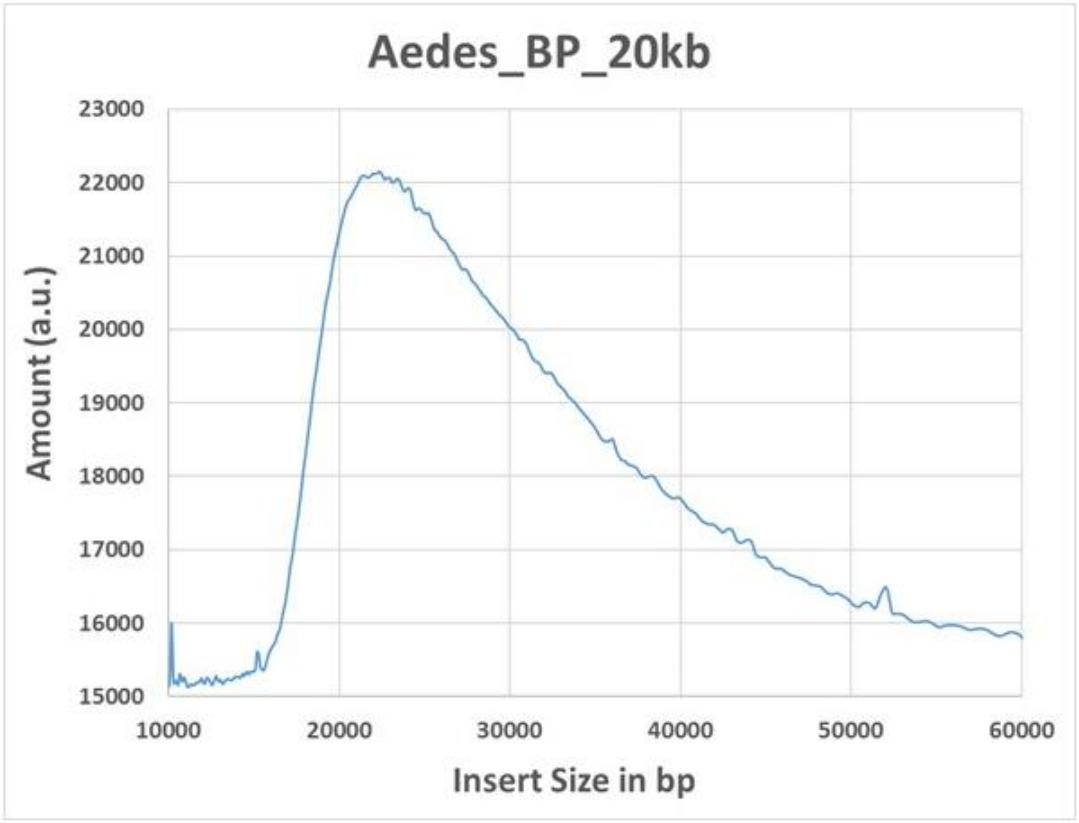
- Inbred (12 generation) strain matching current reference (“Liverpool”) is lost.
- New inbred strain generated:
  - Single male and female from founder strain
  - Male mated to single direct female descendent for 4 generations.
  - Ploidy ≤ 4N
- 80 animals (male pupae)
- HMW genomic DNA extracted with MagAttract Kit, Qiagen



LONG-INSERT LIBRARIES

- Three libraries generated by Paul Peluso (PacBio)
- Size selection of SMRTbell library using BluePippin and ELF

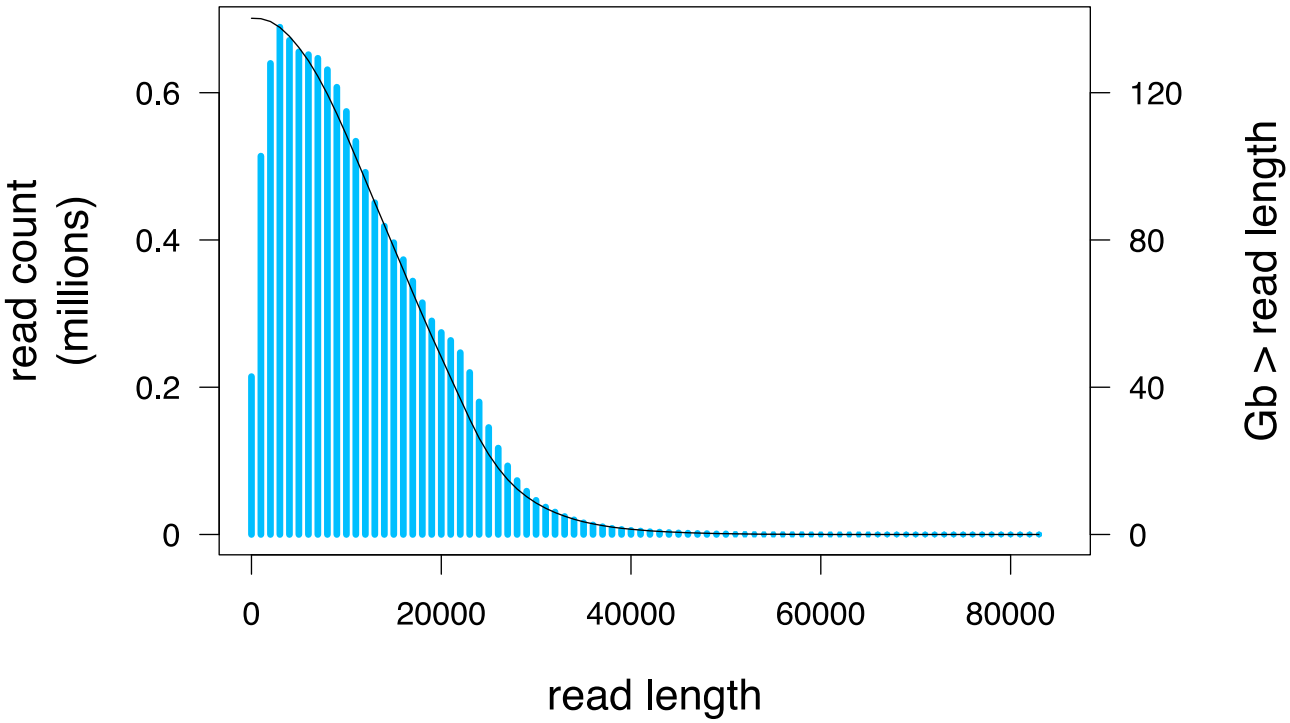




TOTAL RAW READ DATA

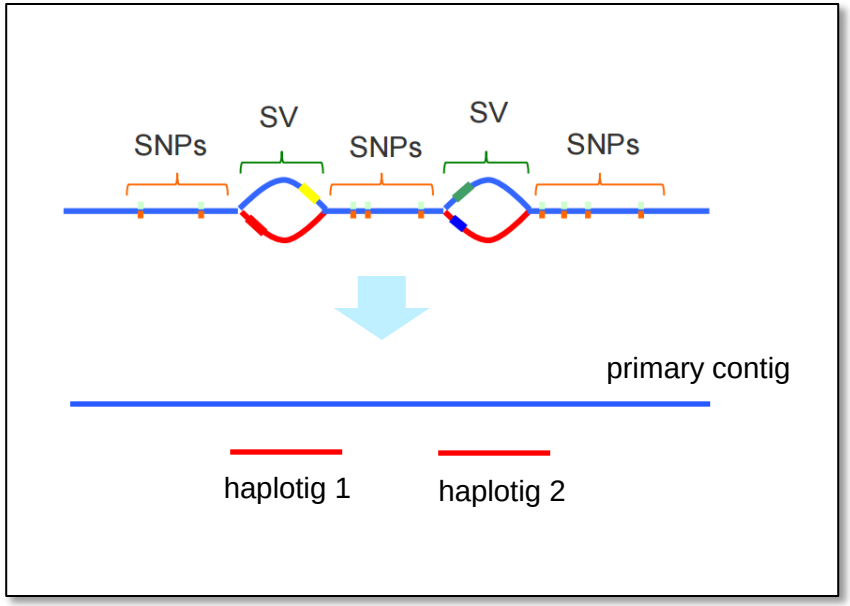
PacBio RS II

Total SMRT Cells	Total Bases	Genome Coverage	Subread N50
177	140 Gb	~108-fold	17 kb



DIPLOID ASSEMBLY

- Falcon-UNZIP assembly using p-reads longer than 19 kb
- Polishing with Arrow to QV40
- Identification of additional haplotigs using coverage and gene annotation



	Primary Contigs	Alternate Haplotigs
Total Length	1.45 Gb	594 Mb
Number of Contigs	3,462	4,328
Contig N50	1.43 Mb	382 kb
Longest Contig	26.5 Mb	5.42 Mb

PACBIO ASSEMBLY: IMPROVED QUALITY AND CONTIGUITY

Assembly	**L3.31 reference[1]**	**PacBio**
Genome Size	1.38 Gb	1.45 Gb
Number of Contigs	36,204	3,462
Improvement factor		**10**
Contig N50	0.083 Mb	1.43 Mb
Improvement factor		**17**
CEGMA Genes	91%	92%
CEGMA Genes with frame shift errors	6 (2.7%)	4 (1.7%)

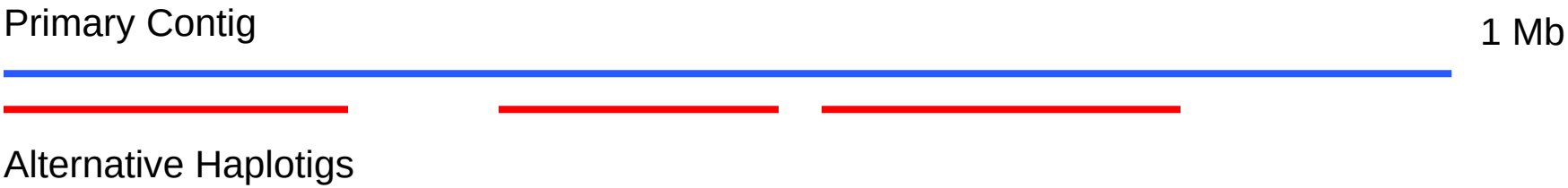
[1] ftp://ftp.ensemblgenomes.org/pub/metazoa/release-31/fasta/aedes_aegypti/dna/

PACBIO ASSEMBLY: IMPROVED QUALITY AND CONTIGUITY

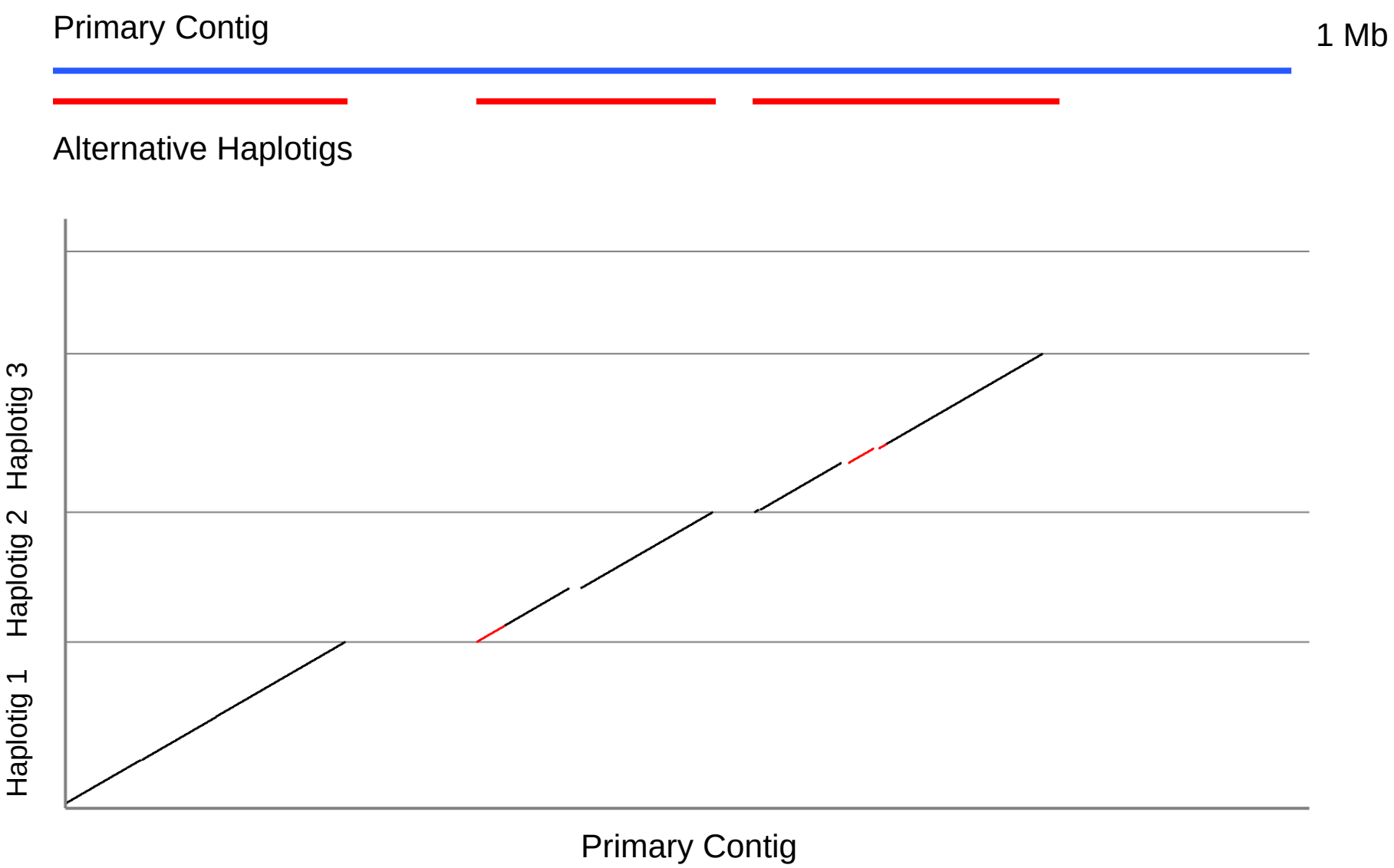
Assembly	**2015 Illumina[1]**	**PacBio**
Genome Size	0.74 Gb	1.45 Gb
Number of Contigs	961,292	3,462
Improvement factor		**278**
Contig N50	0.001 Mb	1.43 Mb
Improvement factor		**1430**
CEGMA Genes	72%	92%
CEGMA Genes with frame shift errors	7 (3.9%)	4 (1.7%)

[1] ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001014885.1_ASM101488v1

LONG-RANGE HAPLOTYPE RESOLUTION



LONG-RANGE HAPLOTYPE RESOLUTION

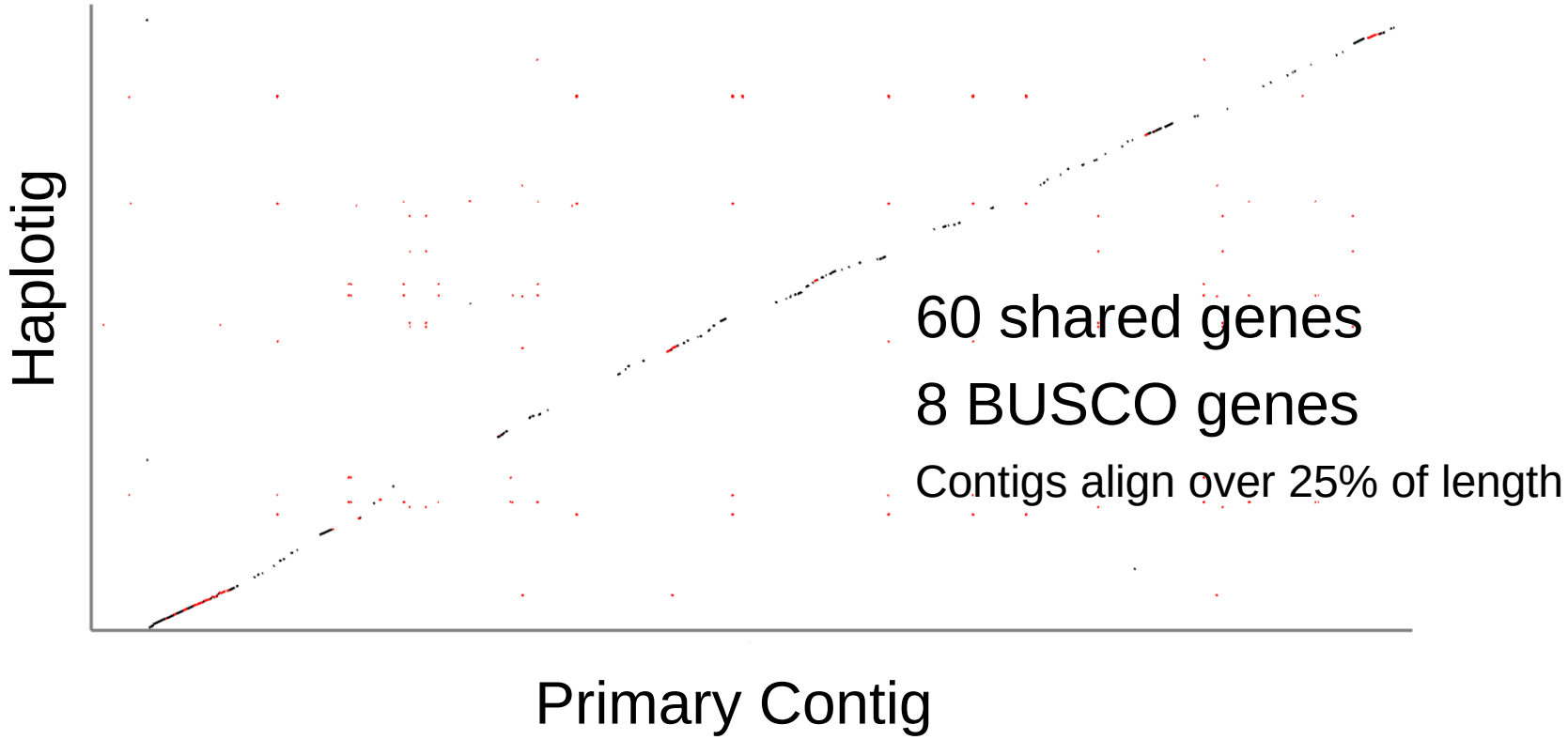


LONG-RANGE HAPLOTYPE RESOLUTION

Primary Contig

4 Mb

Alternative Haplotig



EXAMPLE - VOLTAGE-GATED SODIUM CHANNEL

- Target for insecticides, mutations likely involved in insecticide resistance[1]

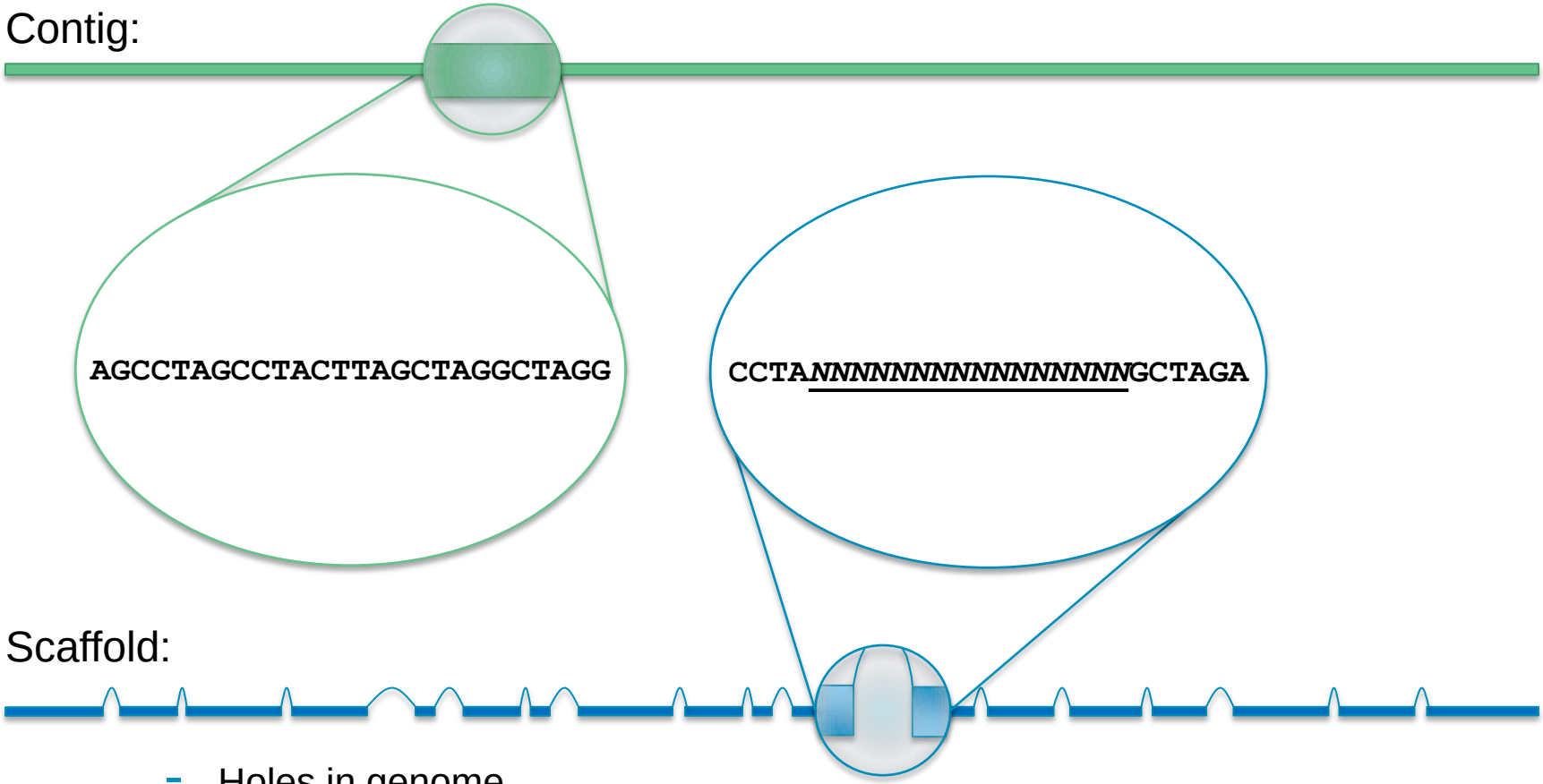
	PacBio assembly	**L3.31 Reference[2]**	**2015 Illumina assembly[3]**
Location	Contig 88 (2.87 Mb)	Scaffold 186 (1.96 Mb)	Scaffold 76061 (0.011 Mb)
Gene Status	**Complete** ~500 kb gene size 2109 a.a. protein	**Incomplete** 11% of protein is missing (N-terminal 225 a.a.)	**Incomplete** 84% of protein is missing (N-terminal 1766 a.a.)

[1] Kawada et al. (2016) Discovery of Point Mutations in the Voltage-Gated Sodium Channel from African Aedes aegypti Populations: Potential Phylogenetic Reasons for Gene Introgression. PLoS Negl Trop Dis 10: e0004780

[2] ftp://ftp.ensemblgenomes.org/pub/metazoa/release-31/fasta/aedes_aegypti/dna/

[3] ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_001014885.1_ASM101488v1

CONTIGs *vs.* SCAFFOLDs



- Holes in genome
- Missing/incorrect information
- Knowledge gaps

Genomes vs. GenNNNes: The Difference between Contigs and Scaffolds in Genome Assemblies

SUMMARY

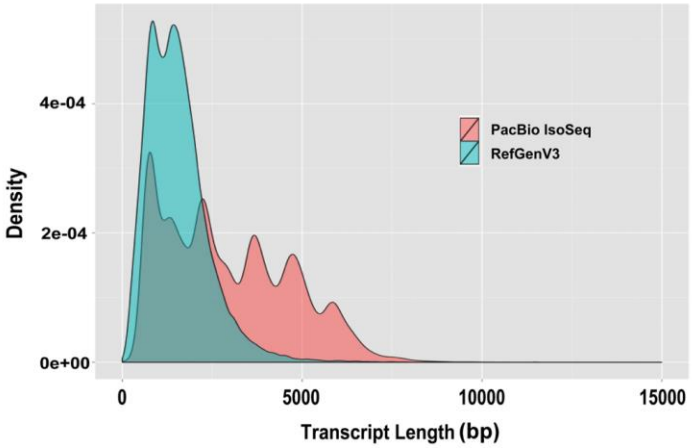
Advantages of Diploid Assembly with PacBio

- More accurate assembly of non-model or non-laboratory organisms
  - Non-inbred
  - Heterozygous
  - Pooled individuals
- Phasing of long-range allelic haplotypes
- More complete base-pair resolution of genome
- Higher genome contiguity and completeness
- Estimate of variation in parental chromosomes
  - Spatial patterns of variation across the genome
  - Copy number/structural variation

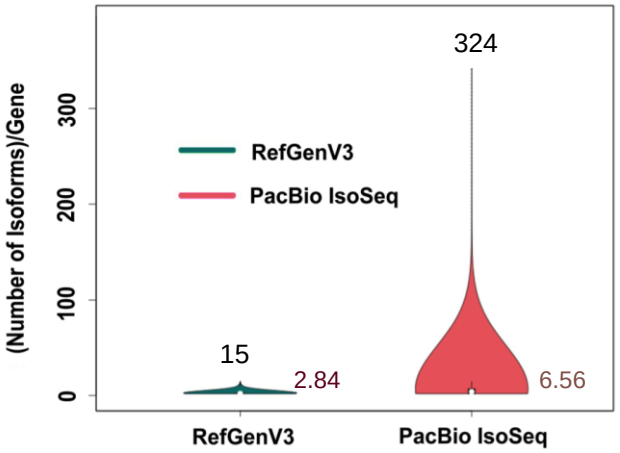
GENOME ANNOTATION

Iso-Seq Method for Full-Length *de novo* Transcript Assembly

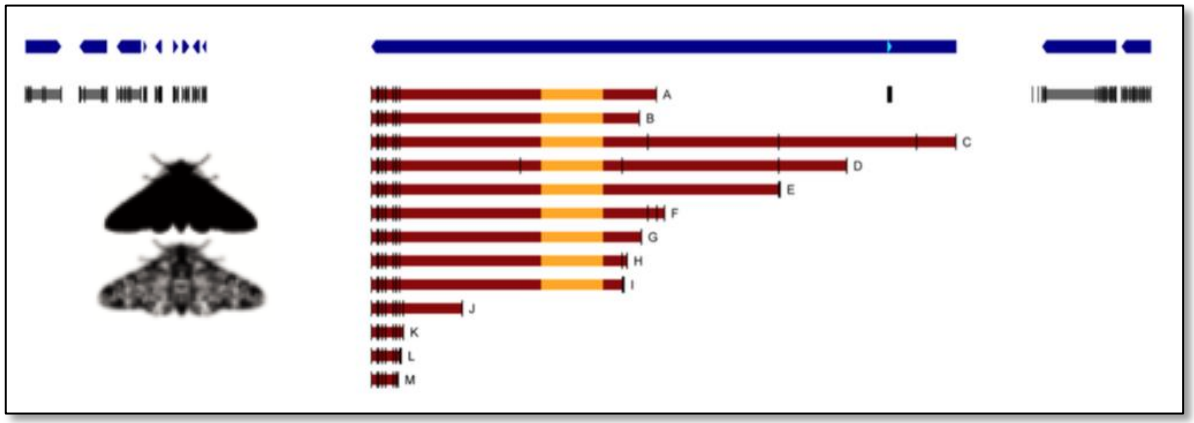
- Longer transcripts



- More numerous isoforms



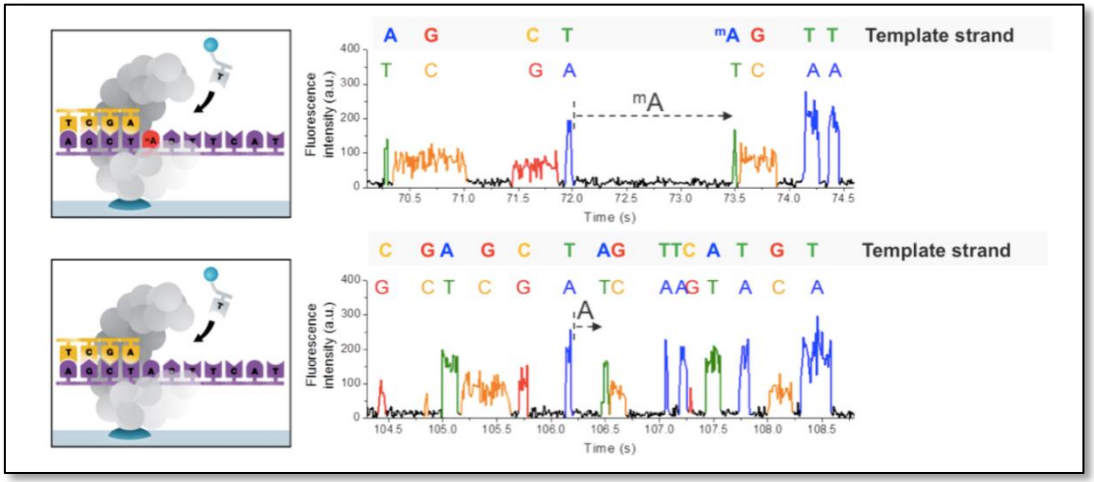
- Peppered moth ”carbonaria” morph due to TE insertion in *cortex* gene

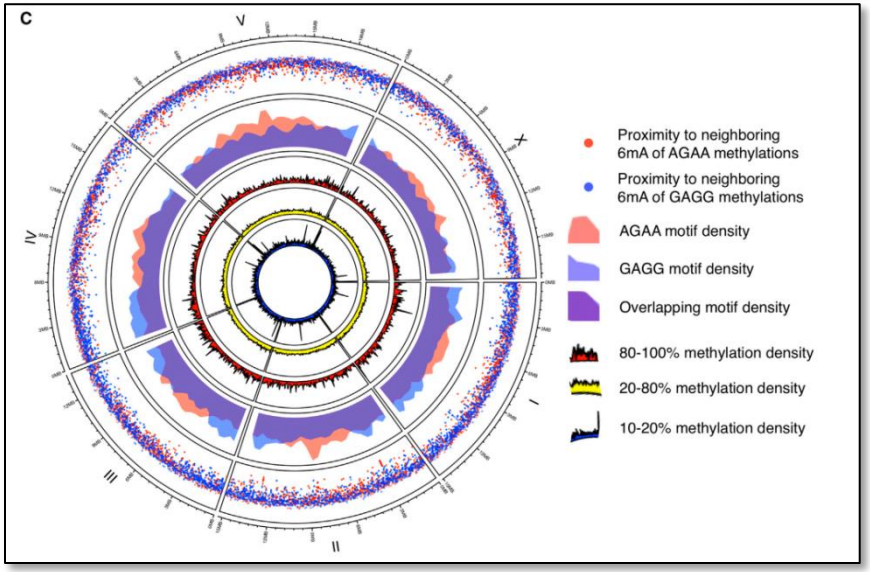


Wang et al. (2016) Nature Communications 7: 11708
van't Hof et al. (2016) Nature, 534: 102

EPIGENOME CHARACTERIZATION

- Detection of base modifications
- Common mods: 5mC, 4mC, 6mA
- Detected through polymerase kinetics



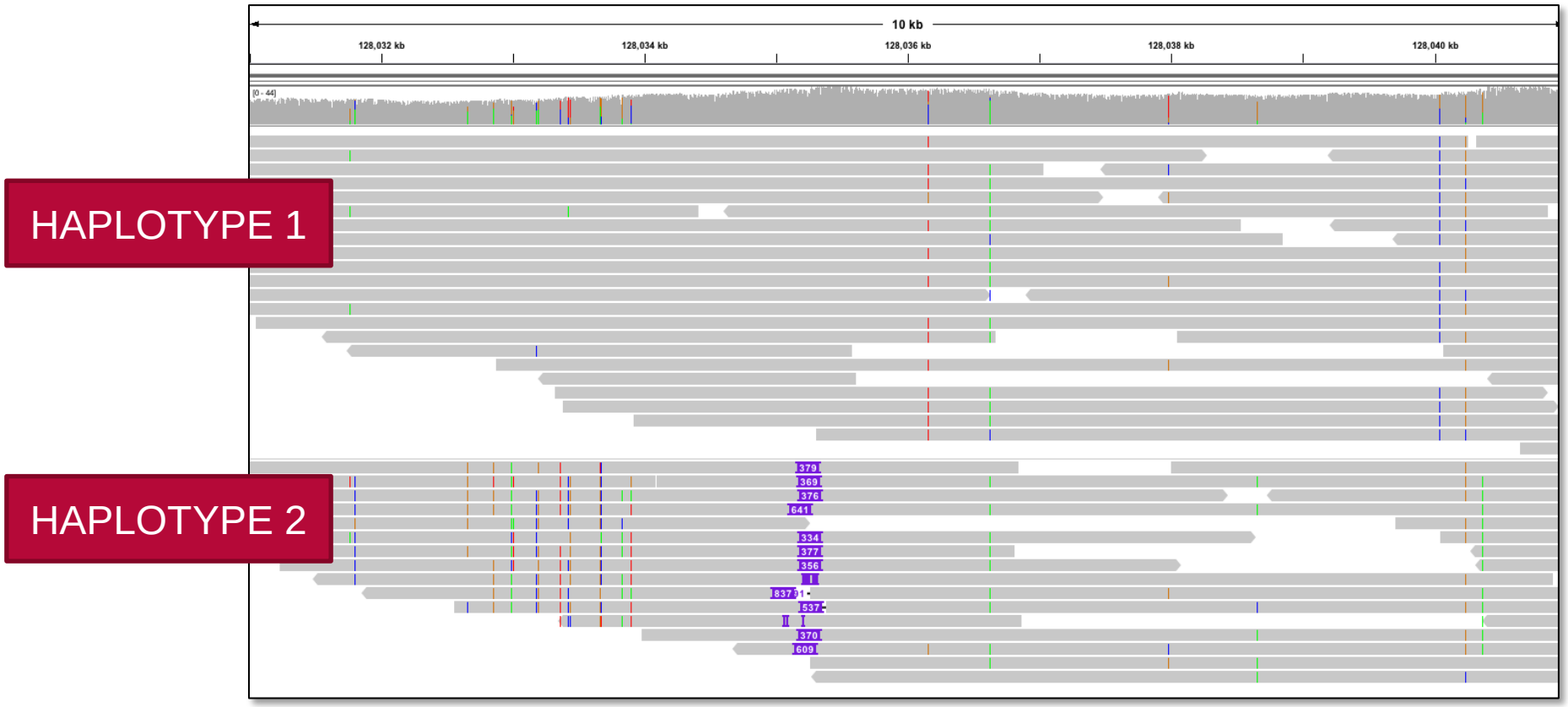


- Relationship to gene expression, host-pathogen interactions, sequence motifs, DNA-damage/repair

Flusberg et al. (2010) Nat Methods 7: 461–465; Greer et al. (2015) Cell 161: 1–11

TOOLS: IGV FOR VARIANT VISUALIZATION

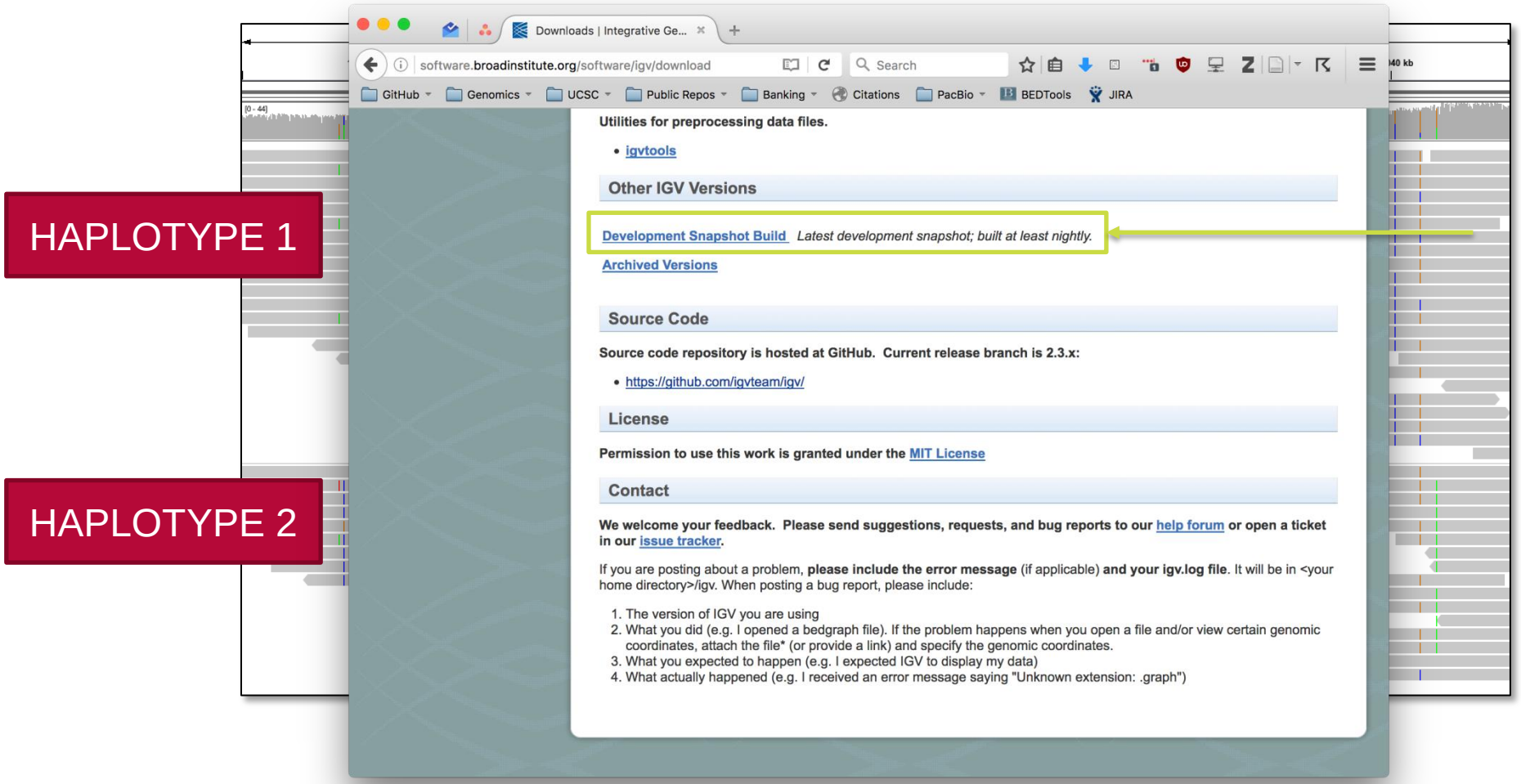
Single Nucleotide Variants and Haplotype Resolution



Aaron Wenger, PacBio

TOOLS: IGV FOR VARIANT VISUALIZATION

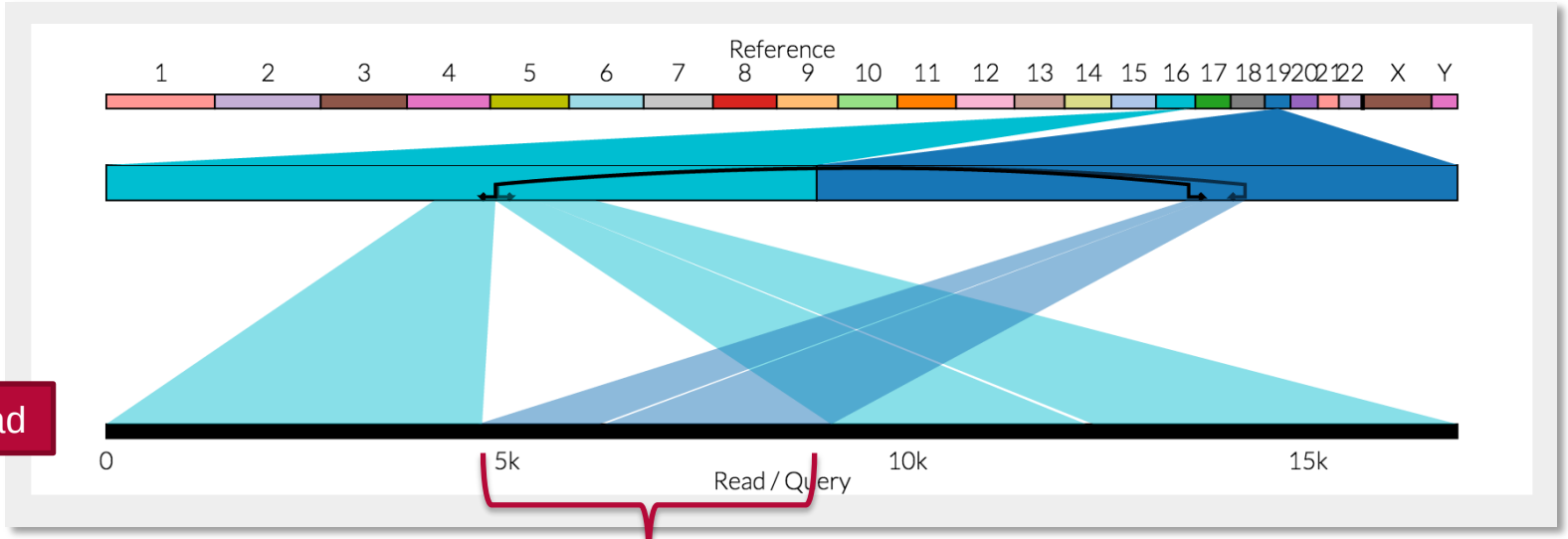
Single Nucleotide Variants and Haplotype Resolution



Aaron Wenger, PacBio

GENOME ANNOTATION

Structural Variation Visualized in Ribbon



Single Read

Insertion of Region from chr19 into chr16

Maria Nattestad (CSHL), http://marianattestad.com/research/

SUMMARY

Advantages of Diploid Assembly with PacBio

- More accurate assembly of non-model or non-laboratory organisms
  - Non-inbred
  - Heterozygous
  - Pooled individuals
- Phasing of long-range allelic haplotypes
- More complete base-pair resolution of genome
- Higher genome contiguity and completeness
- Estimate of variation in parental chromosomes
  - Spatial patterns of variation across the genome
  - Copy number/structural variation