

## Bioinformatics Workshop Agenda – June 27, 2017

Hosted by the Genomics Resource Center,  
University of Maryland School of Medicine



THE LEADER IN LONG-READ SEQUENCING

 PACBIO®

**8:30 - 8:55 a.m.** Registration and Continental Breakfast

**9:00 - 9:10 a.m.** Welcome and Introduction

Roberto Lleras, Manager, Field Applications Scientist, Bioinformatics, PacBio

**9:10 - 9:55 a.m.** Introduction to SMRTLink 5.0

Minor Variant Detection with Juliet

Roberto Lleras, Manager, Field Applications Scientist, Bioinformatics, PacBio

Structural Variant Detection with PBSV

Aaron Wenger, Ph.D., Staff Scientist, PacBio

**10:00 - 11:15 a.m.** Concurrent Breakout Sessions

**SESSION I: So, I Have a Diploid Assembly...Now What?**

Understanding, Curating, and Analyzing Your Diploid Genome Assembly

Sarah Kingan, Ph.D., Senior Scientist, Bioinformatics, PacBio

Chromosome-scale De Novo Assembly of Mammalian Genomes Using Chromatin Interaction Data

Jay Ghurye, Ph.D. Candidate, Department of Computer Science, University of Maryland

SESSION I Discussion

**SESSION II: Mini-training Session: Best Practices in Multiplexing with PacBio**

Best Practice for Interpreting Demultiplexed Output

Carmen Guarco, Ph.D., Scientist, Field Applications Support, Bioinformatics, PacBio

Downstream Applications: Minor Variant Calling, Microbial Assembly, CCS2, and Iso-Seq

Roberto Lleras, Manager, Field Applications Scientist, Bioinformatics, PacBio

SESSION II Discussion

**11:15 - 11:30 a.m.** Coffee Break

Breakout Sessions Wrap-up

Keynote: Accurate Detection of Complex Structural Variation

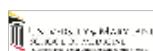
Fritz J. Sedlazeck, Ph.D., Lead Scientific Programmer, Human Genome Sequencing Center, Baylor College of Medicine

Open Forum for User Questions, Comments and Feedback on SMRTLink

Closing Remarks

**1:00 p.m.** Lunch

Thanks to  
our Partners:





PACBIO®

# Understanding, Curating, and Analyzing your Diploid Genome Assembly

Sarah B. Kingan, Ph.D.  
Senior Scientist, Bioinformatics, PacBio Applications

East Coast UGM, Baltimore, MD  
Tuesday June 27<sup>th</sup> 2017

# AGENDA

- **Understanding Your Diploid Assembly**
  - Assembly Workflow
  - Heterozygosity and Coverage
- **Curating your Assembly**
  - Filtering Contigs
  - Deduplicating Haplotypes
- **Submitting Your Assembly to NCBI**

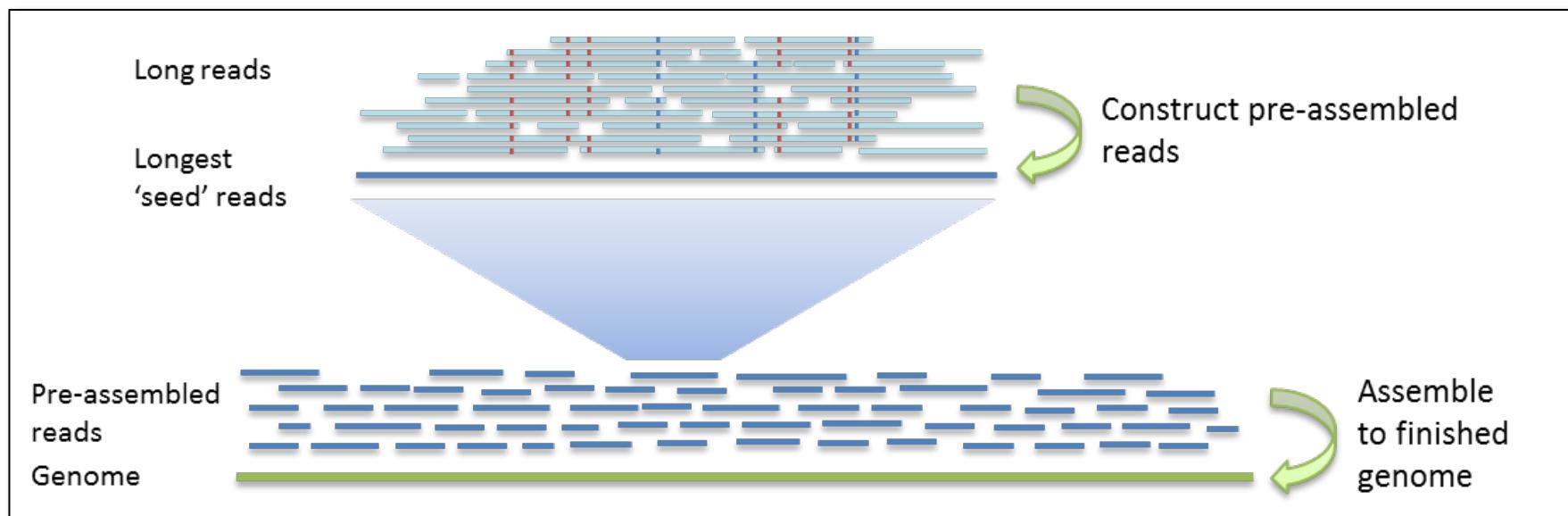
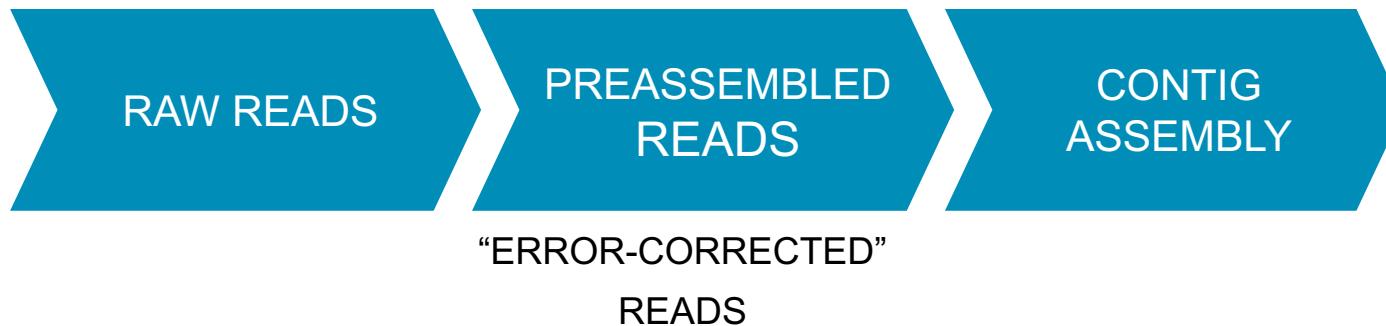


PACBIO®

# Understanding Your Diploid Assembly

Assembly Workflow: From Raw Reads to High Quality Reference

# FALCON / HIERARCHICAL GENOME ASSEMBLY PROCESS (HGAP)



# FALCON AND FALCON-UNZIP



## Phased diploid genome assembly with single-molecule real-time sequencing.

Chin CS, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, Cramer GR, Delledonne M, Luo C, Ecker JR, Cantu D, Rank DR, Schatz MC

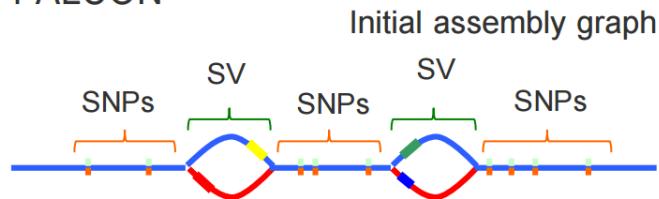
### ABSTRACT

While genome assembly projects have been successful in many haploid and inbred species, the assembly of noninbred or rearranged heterozygous genomes remains a major challenge. To address this challenge, we introduce the open-source FALCON and FALCON-Unzip algorithms (<https://github.com/PacificBiosciences/FALCON/>) to assemble long-read sequencing data into highly accurate, contiguous, and correctly phased diploid genomes. We generate new reference sequences for heterozygous samples including an F1 hybrid of *Arabidopsis thaliana*, the widely cultivated *Vitis vinifera* cv. Cabernet Sauvignon, and the coral fungus *Clavicorona pyxidata*, samples that have challenged short-read assembly approaches. The FALCON-based assemblies are substantially more contiguous and complete than alternate short- or long-read approaches. The phased diploid assembly enabled the study of haplotype structure and heterozygosities between homologous chromosomes, including the identification of widespread heterozygous structural variation within coding sequences.

- FALCON is a **diploid-aware assembler**.
- FALCON-Unzip module performs true **phased assembly** for diploid samples.

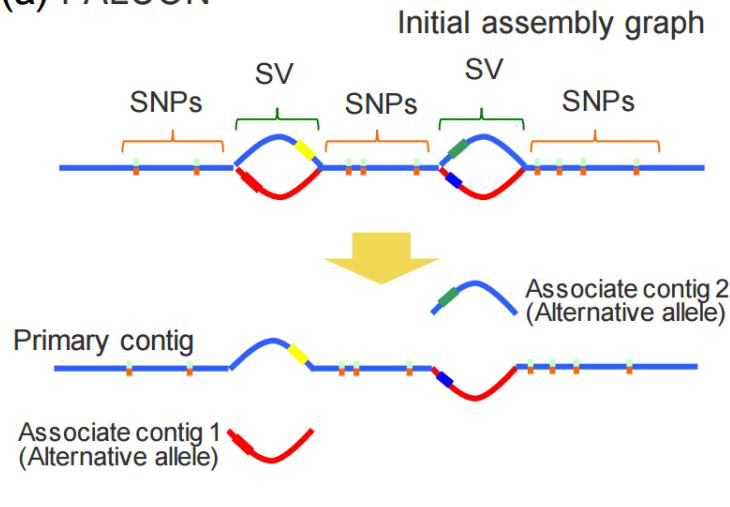
# DIPLOID ASSEMBLY WITH FALCON-UNZIP

(a) FALCON



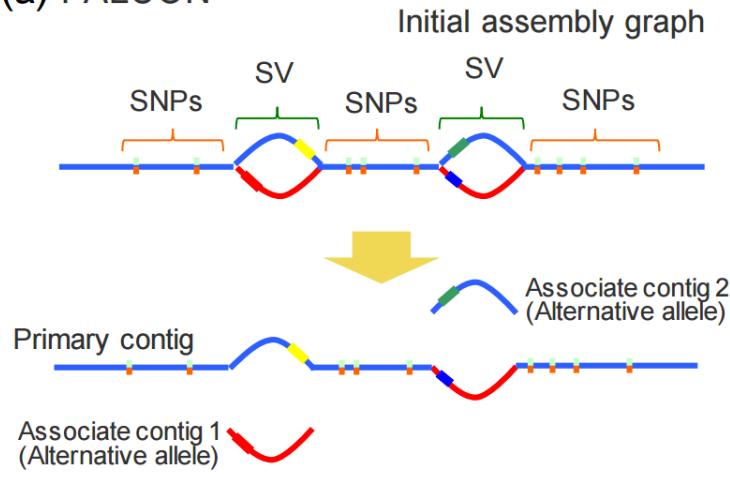
# DIPLOID ASSEMBLY WITH FALCON-UNZIP

(a) FALCON



# DIPLOID ASSEMBLY WITH FALCON-UNZIP

(a) FALCON



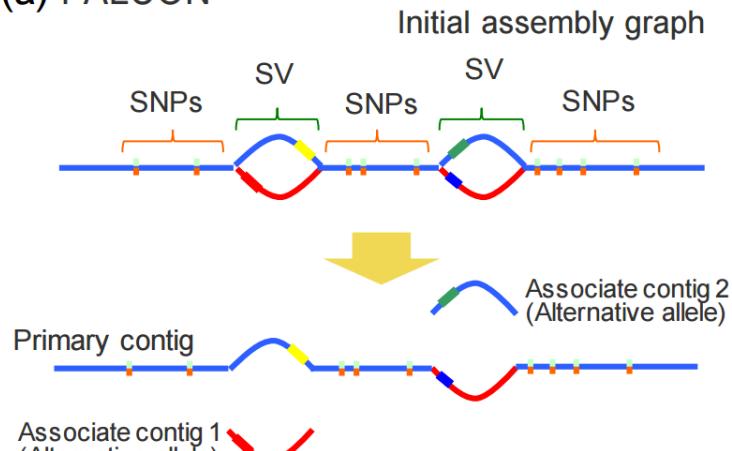
(b)



Phase heterozygous SNPs and identify the haplotype of each read

# DIPLOID ASSEMBLY WITH FALCON-UNZIP

(a) FALCON



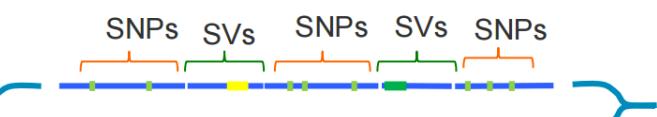
(b)



Phase heterozygous SNPs and identify the haplotype of each read

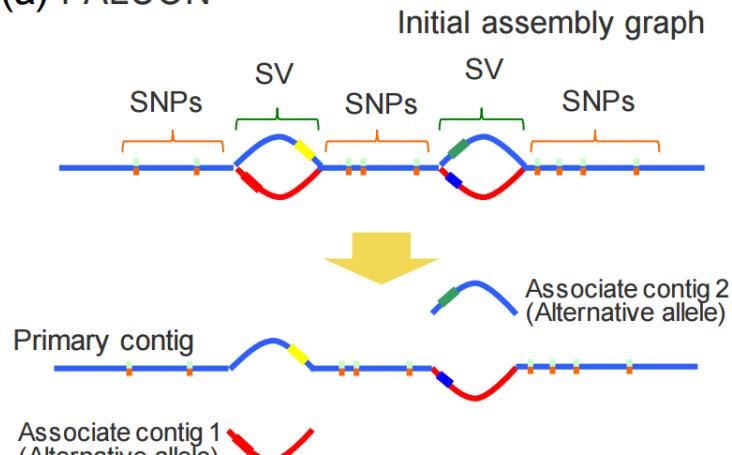
(c) FALCON-Unzip

Haplotype-resolved assembly graph



# DIPLOID ASSEMBLY WITH FALCON-UNZIP

(a) FALCON

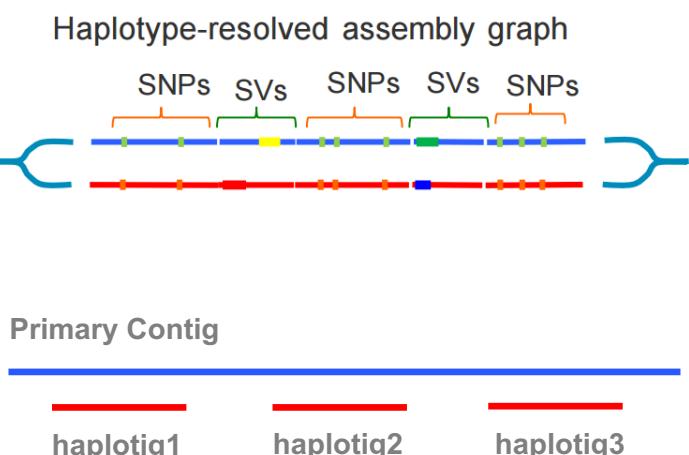


(b)



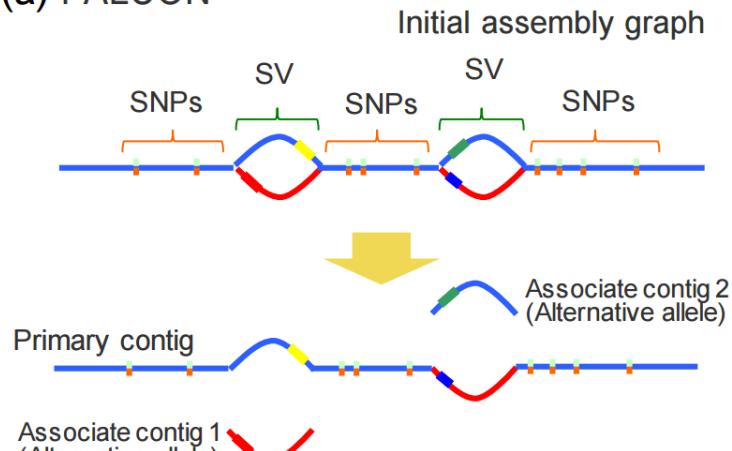
Phase heterozygous SNPs and identify the haplotype of each read

(c) FALCON-Unzip



# DIPLOID ASSEMBLY WITH FALCON-UNZIP

(a) FALCON

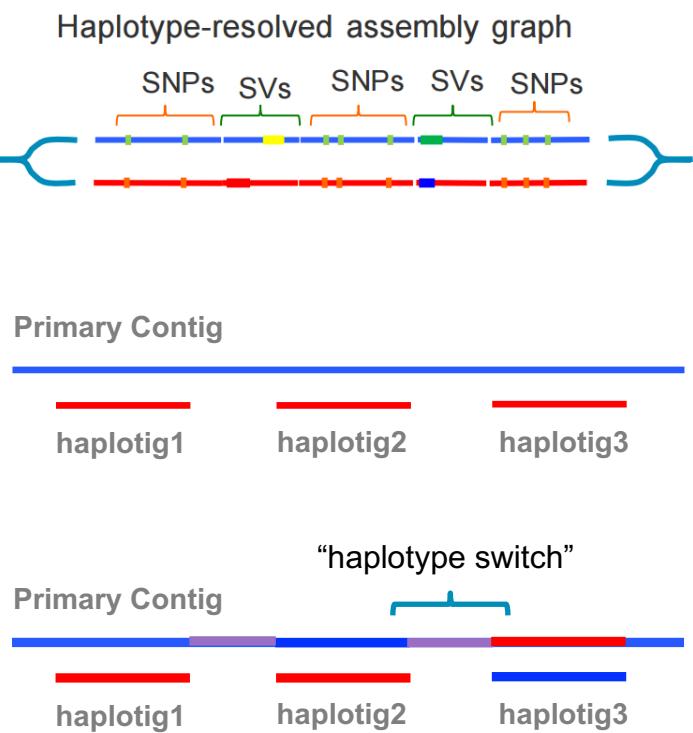


(b)



Phase heterozygous SNPs and identify the haplotype of each read

(c) FALCON-Unzip



# EXAMPLE ASSEMBLY OF WATER BUFFALO

	FALCON	FALCON-Unzip
<b>Primary Length</b>	2.66 Gb	2.65 Gb
<b>Primary N50</b>	18.7 Mb	18.8 Mb
<b>Secondary Length</b>	0.218 Gb	1.53 Gb
<b>Proportion Phased</b>	8.2 %	58 %

PRIMARY CONTIG

SECONDARY CONTIGS

7-fold increase in  
haplotype phasing  
with Unzip module



Djambalawa, Wiki Commons

Acknowledgements:

Tim Smith, USDA-ARS

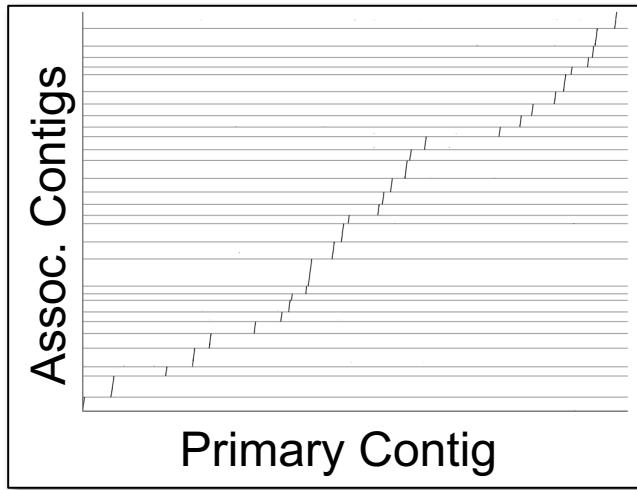
John Williams, University of Adelaide

Paola Ajmone-Marsan, Università Cattolica  
del S. Cuore

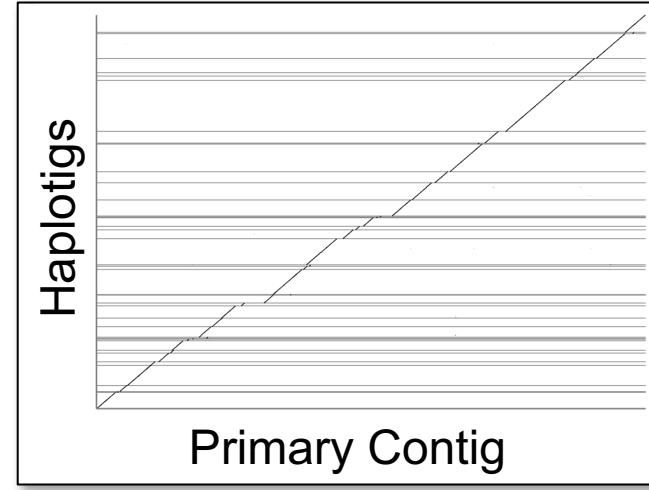
David Hume, Mick Watson, Roslin Institute

# INCREASED HAPLOTIG CONTINUITY WITH FALCON-UNZIP

FALCON



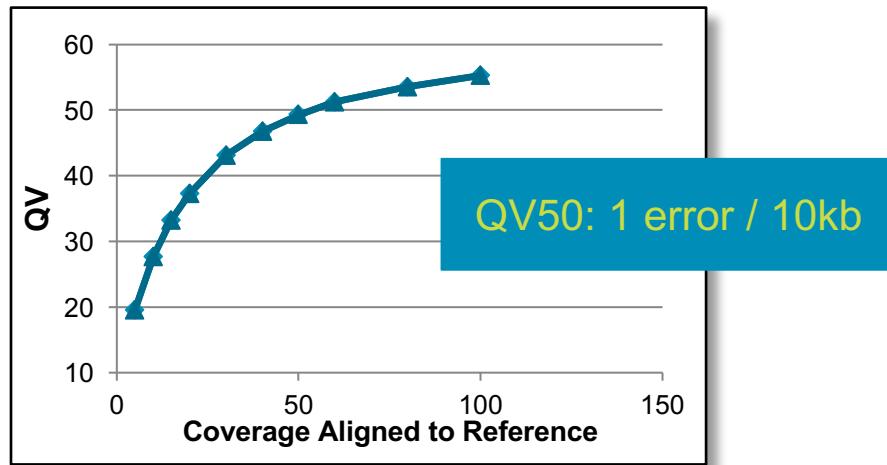
FALCON-UNZIP



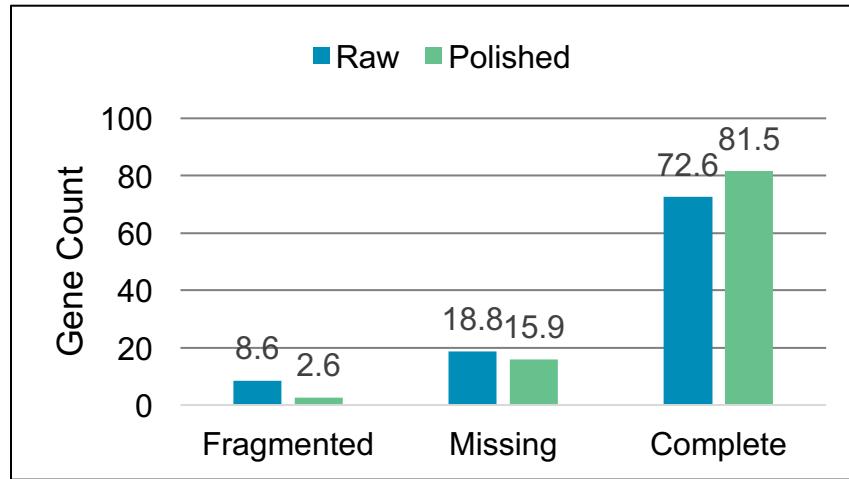
000078F	FALCON	FALCON-Unzip
Primary Contig Length	12.9 Mb	12.9 Mb
Number Secondary Contigs	30	34
Total Secondary Length	1.21 Mb	10.6 Mb
Secondary Contig N50	42.5 kb	470 kb
Proportion Phased	9.3 %	82%

# POLISHING WITH ARROW: INCREASED REFERENCE QUALITY

## CONSENSUS BASE ACCURACY



## GENOME COMPLETENESS WITH BUSCO



## Consensus Base Accuracy

- Sequel 2.0 Chemistry
- Bacterial Genomes

## Genome Completeness

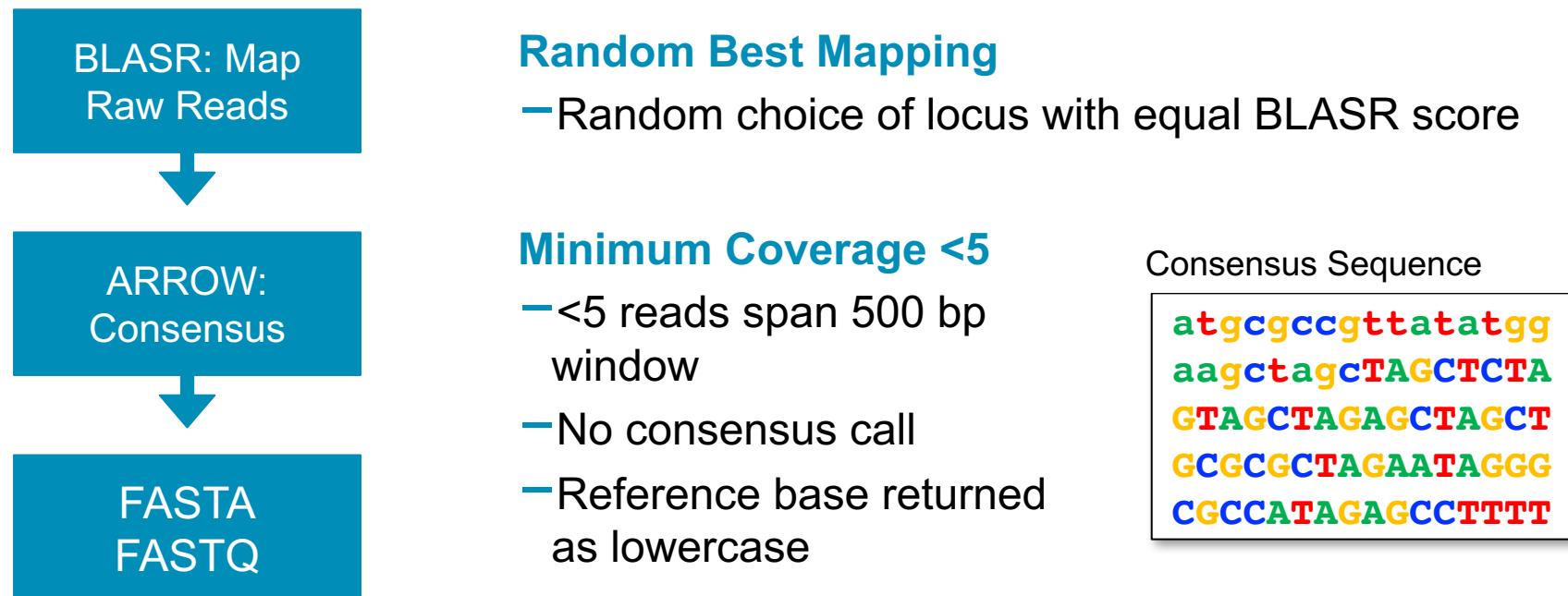
- Avian Genome
- 50-fold Raw Coverage
- BUSCO2 analysis with eukaryota geneset

70% reduction in Fragmented Genes  
15% reduction in Missing Genes  
12% increase in Complete Genes

**Acknowledgement:**  
Erich Jarvis, Rockefeller University

# POLISHING WITH ARROW: WORKFLOW

METHOD	ASSEMBLY	POLISHING
HGAP4 - SMRT Link	✓	✓
FALCON	✓	resequencing pipeline from pbsmrtpipe/SMRT Link
FALCON-Unzip	✓	✓ (phased) plus optional resequencing



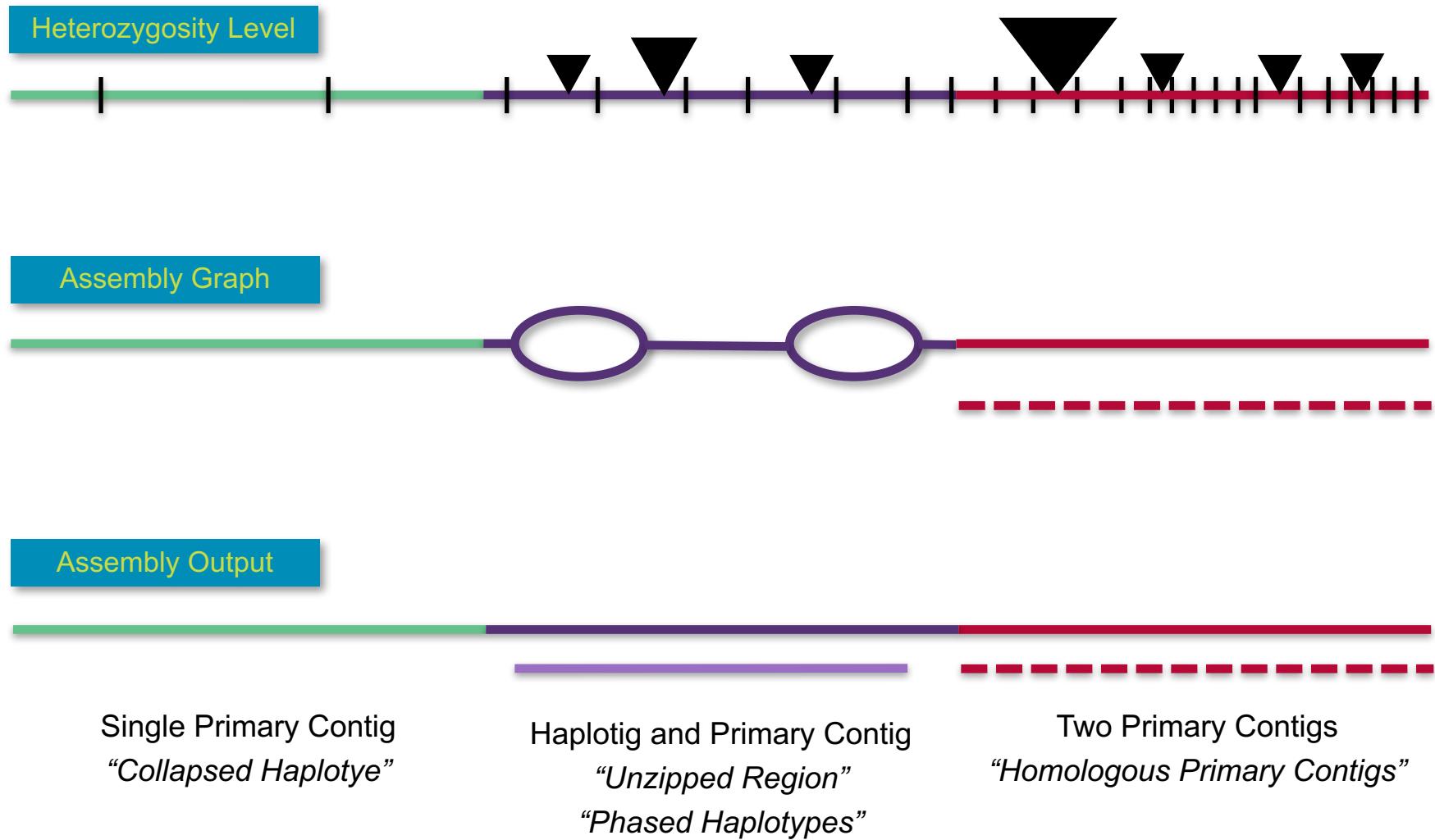


PACBIO®

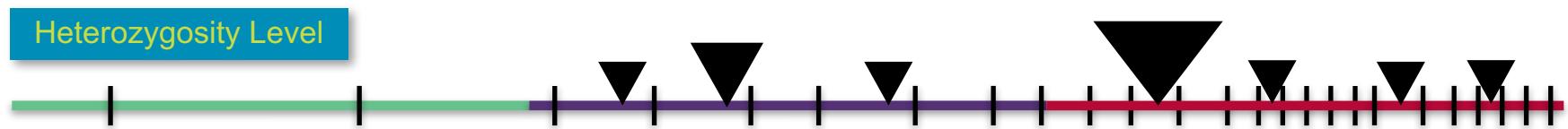
# Understanding Your Diploid Assembly

Heterozygosity, Assembly Structure, and Coverage

# IMPACT OF HETEROZYGOSITY ON ASSEMBLY PROCESS

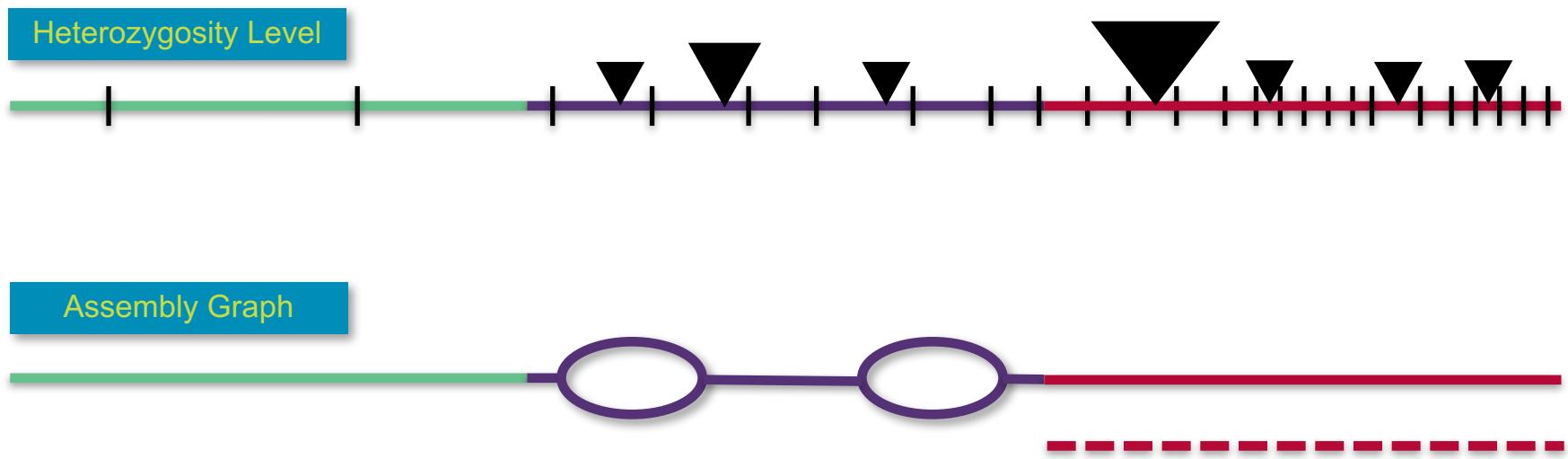


# IMPACT OF HETEROZYGOSITY ON ASSEMBLY PROCESS

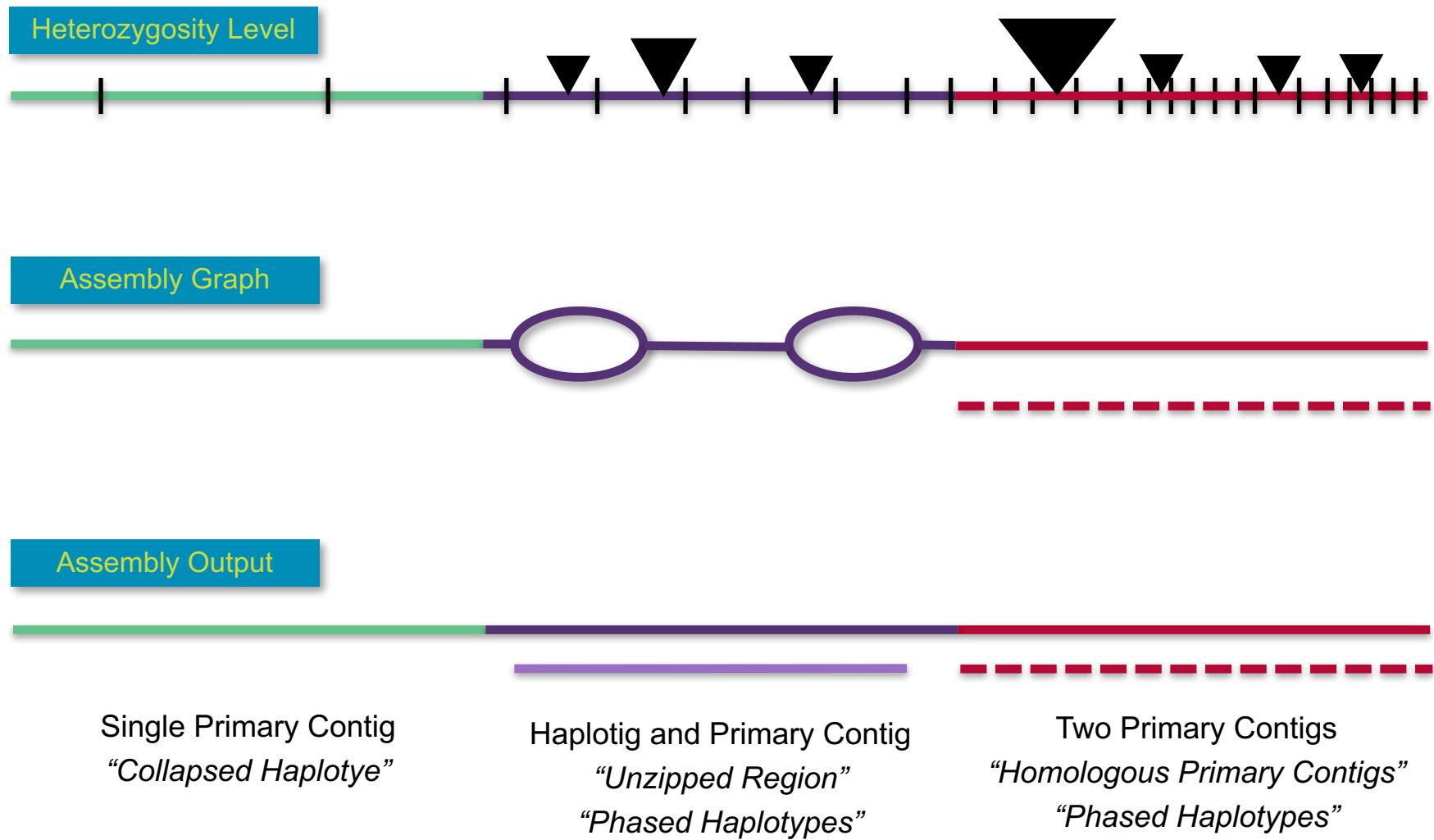


Modified from Chin et al. 2016

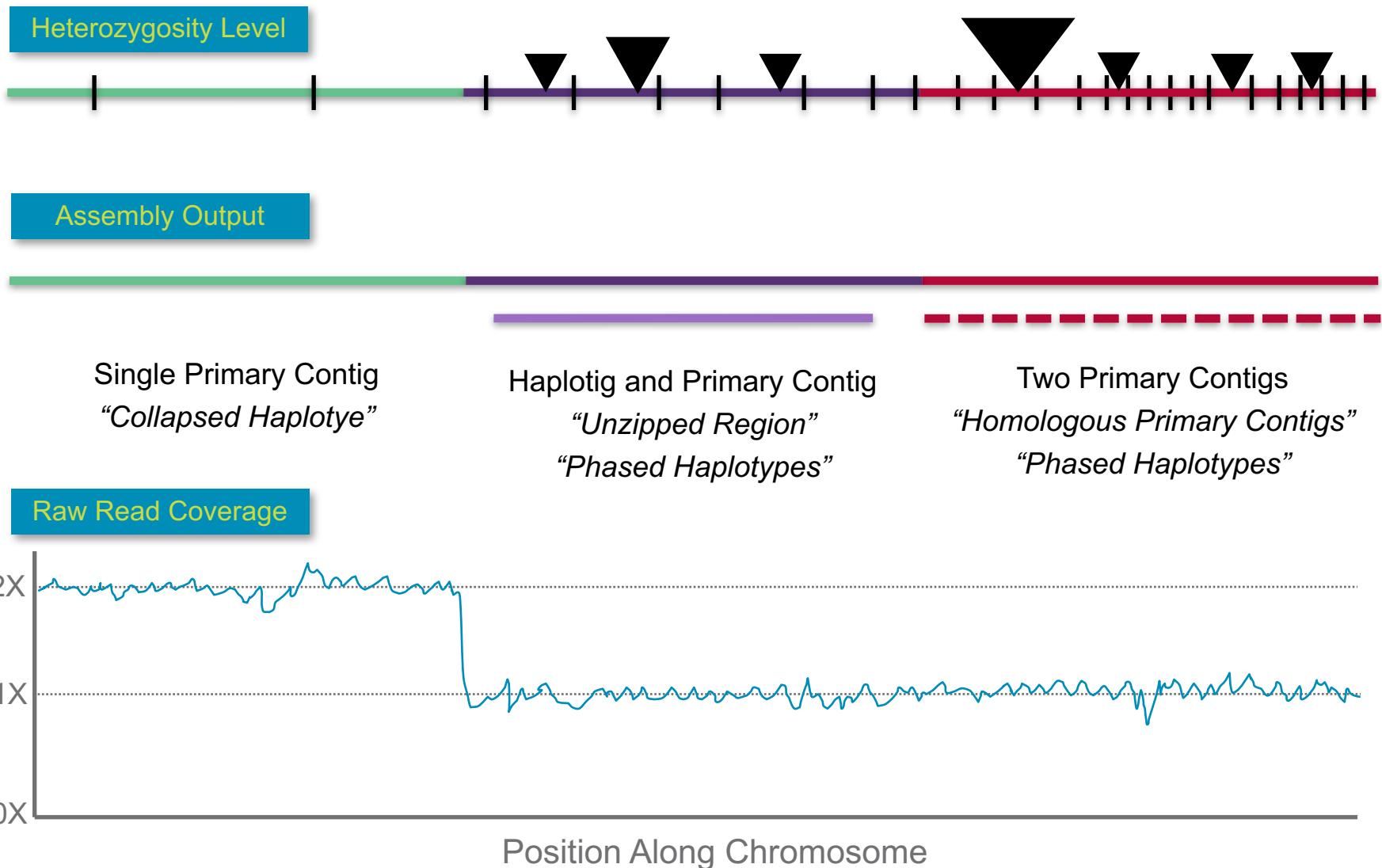
# IMPACT OF HETEROZYGOSITY ON ASSEMBLY PROCESS



# IMPACT OF HETEROZYGOSITY ON ASSEMBLY PROCESS



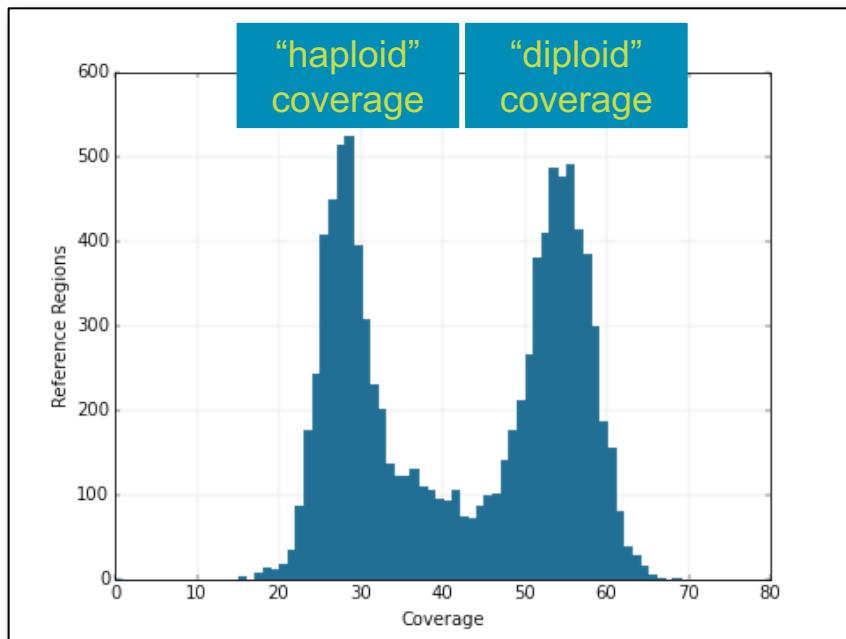
# RAW READ COVERAGE AND ASSEMBLY STRUCTURE



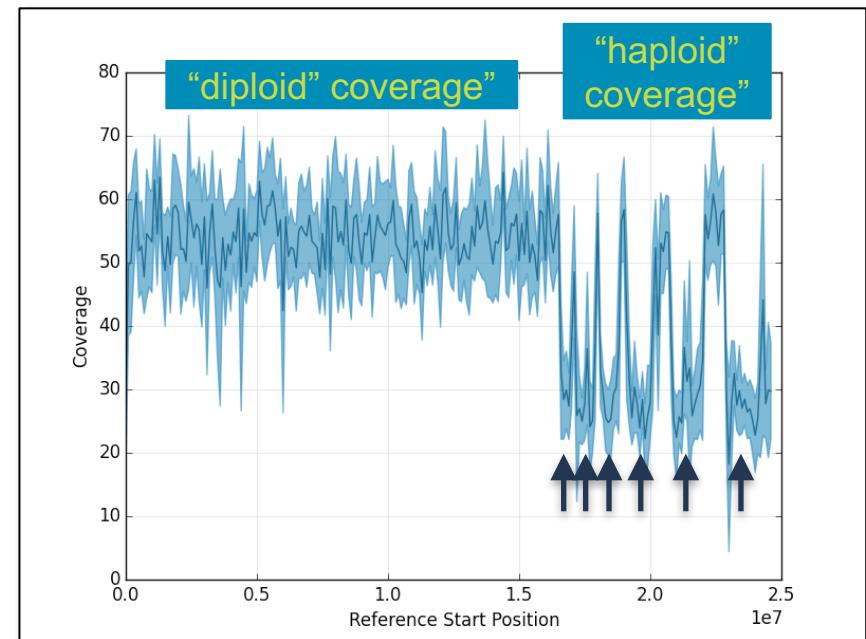
# SMRT LINK COVERAGE REPORTS

Graphical Outputs from Resequencing Pipeline / HGAP4

COVERAGE HISTOGRAM: GENOME



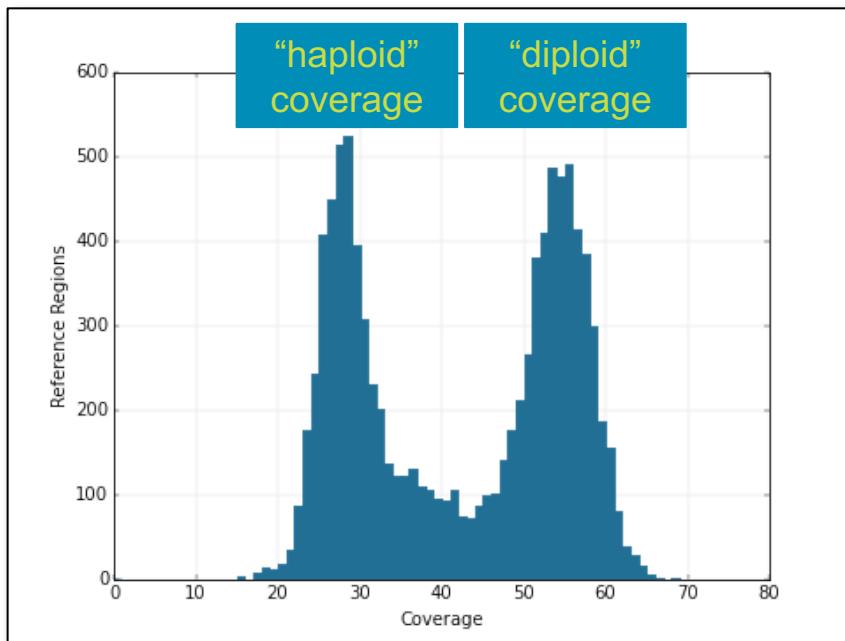
COVERAGE PLOT: CONTIG



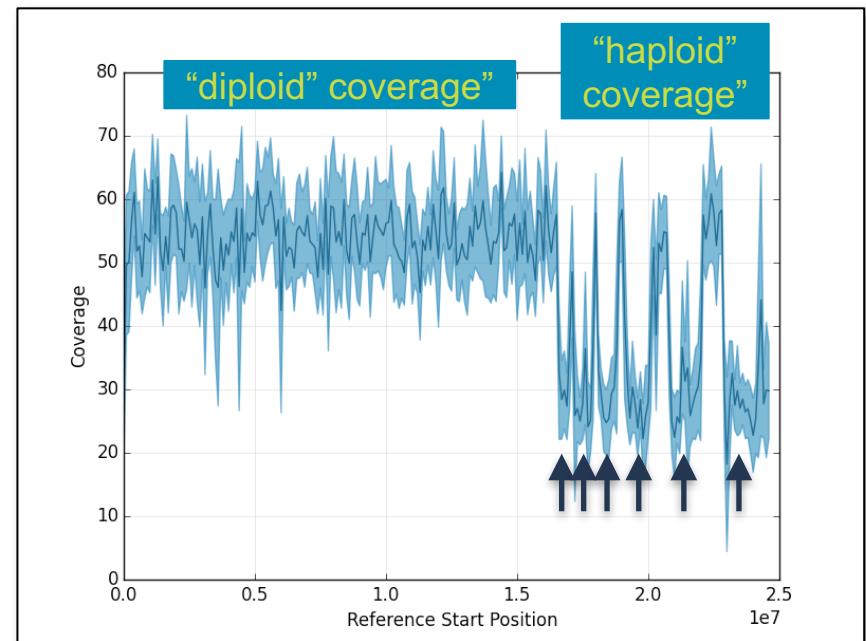
# SMRT LINK COVERAGE REPORTS

Graphical Outputs from Resequencing Pipeline / HGAP4

COVERAGE HISTOGRAM: GENOME



COVERAGE PLOT: CONTIG



PRIMARY CONTIG

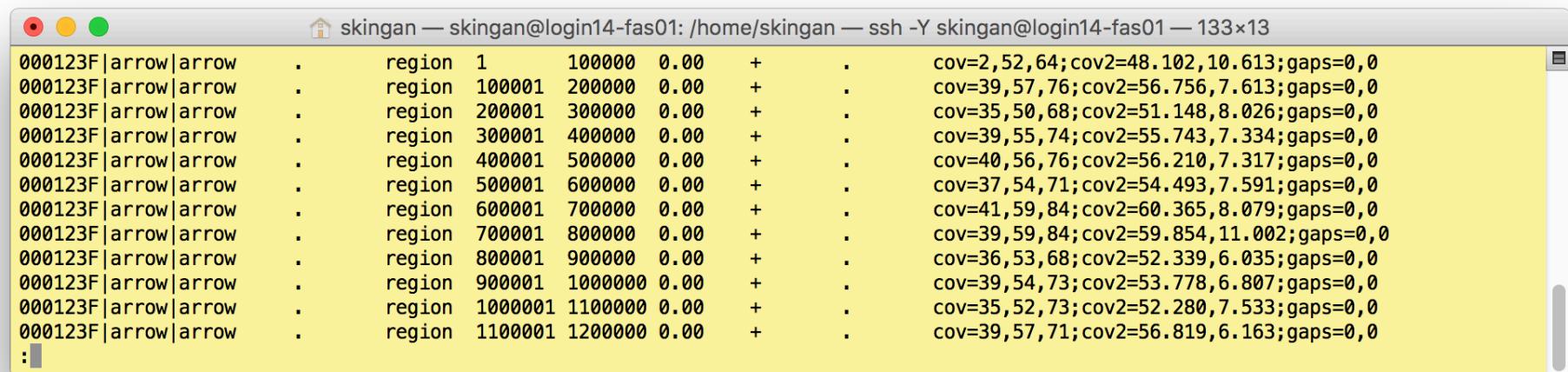
SECONDARY CONTIGS

# SMRT LINK COVERAGE SUMMARY FILES

`alignment_summary.gff`: coarse coverage across all contigs

## —SMRT Link job directory

- `myJob/tasks/pbreports.tasks.summarize_coverage-0/alignment_summary.gff`
- File Format Specs: <https://github.com/ben-lerch/SAT>



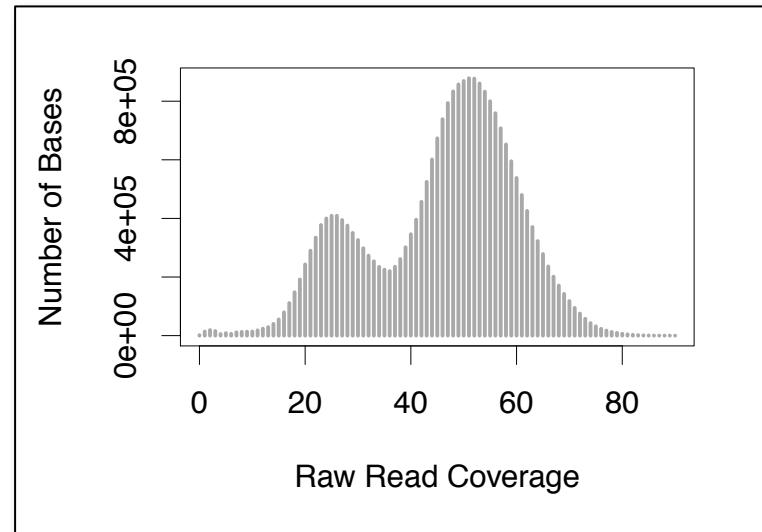
```
skingan — skingan@login14-fas01: /home/skingan — ssh -Y skingan@login14-fas01 — 133x13
000123F|arrow|arrow . region 1 100000 0.00 + . cov=2,52,64;cov2=48.102,10.613;gaps=0,0
000123F|arrow|arrow . region 100001 200000 0.00 + . cov=39,57,76;cov2=56.756,7.613;gaps=0,0
000123F|arrow|arrow . region 200001 300000 0.00 + . cov=35,50,68;cov2=51.148,8.026;gaps=0,0
000123F|arrow|arrow . region 300001 400000 0.00 + . cov=39,55,74;cov2=55.743,7.334;gaps=0,0
000123F|arrow|arrow . region 400001 500000 0.00 + . cov=40,56,76;cov2=56.210,7.317;gaps=0,0
000123F|arrow|arrow . region 500001 600000 0.00 + . cov=37,54,71;cov2=54.493,7.591;gaps=0,0
000123F|arrow|arrow . region 600001 700000 0.00 + . cov=41,59,84;cov2=60.365,8.079;gaps=0,0
000123F|arrow|arrow . region 700001 800000 0.00 + . cov=39,59,84;cov2=59.854,11.002;gaps=0,0
000123F|arrow|arrow . region 800001 900000 0.00 + . cov=36,53,68;cov2=52.339,6.035;gaps=0,0
000123F|arrow|arrow . region 900001 1000000 0.00 + . cov=39,54,73;cov2=53.778,6.807;gaps=0,0
000123F|arrow|arrow . region 1000001 1100000 0.00 + . cov=35,52,73;cov2=52.280,7.533;gaps=0,0
000123F|arrow|arrow : region 1100001 1200000 0.00 + . cov=39,57,71;cov2=56.819,6.163;gaps=0,0
```

cov	cov2	gaps
min	mean	number continuous gaps
median	s.d.	number gap bases
max		

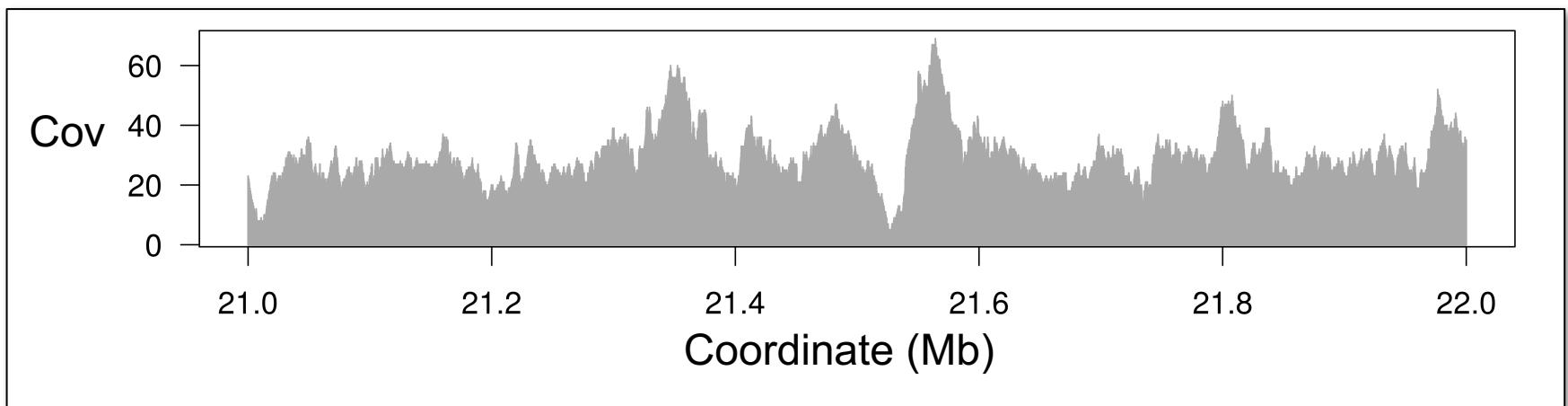
# TOOLS FOR CUSTOM COVERAGE ANALYSIS

- Merge BAM files (N=24)
  - pbmerge
  - samtools merge
- Coverage calculation
  - samtools depth
  - bedtools genomecov
- Visualization
  - R – text file/dataframe

**CONTIG COVERAGE HISTOGRAM**  
bedtools genomecov, R



**CONTIG COVERAGE WINDOW** bedtools genomecov/samtools depth, R



# ALIGNMENTS AND VISUALIZATION

## Recommended Tools for Haplotype Alignment and Analysis

### — Subset Reference Sequence

- samtools faidx

### — Alignment

- MUMmer (v4, multi-threaded support)
- NUCmer, delta-filter, show-coords, show-snps, etc

### — Visualization

- mummerplot
- assemblytics
- gepard

### — FALCON Assembly Tools

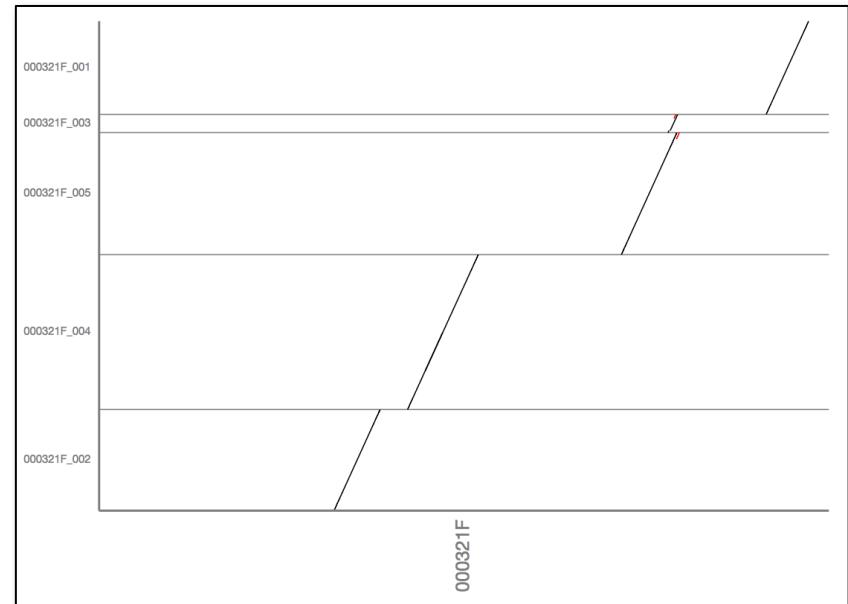
- <https://github.com/PacificBiosciences/apps-scripts/>

- FALCONAssemblyTools repo

### HAPLOTIGS TO PRIMARY CONTIG DOTPLOT

`alignHaplotigs2Primary.sh`

`Assemblytics`





PACBIO®

# Assembly Finishing

Filtering, Circularizing, Haplotype Deduplication

# GUIDELINE FOR CONTIG FILTERING

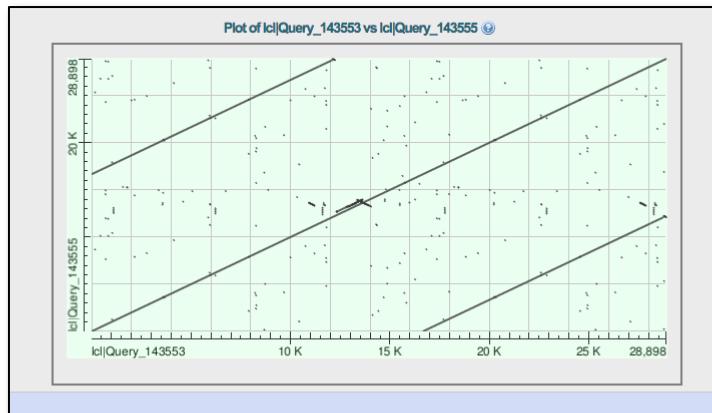
<https://github.com/PacificBiosciences/apps-scripts/tree/master/FALCONAssemblyTools>

## Circularize organelle

- Identify
  - high coverage
  - “circular ctg” FALCON annotation
  - blast hit to organelle
- Circularize and polish
  - minimus2, circulator

## SELF-ALIGNMENT OF MITO CONTIG

### BLAST



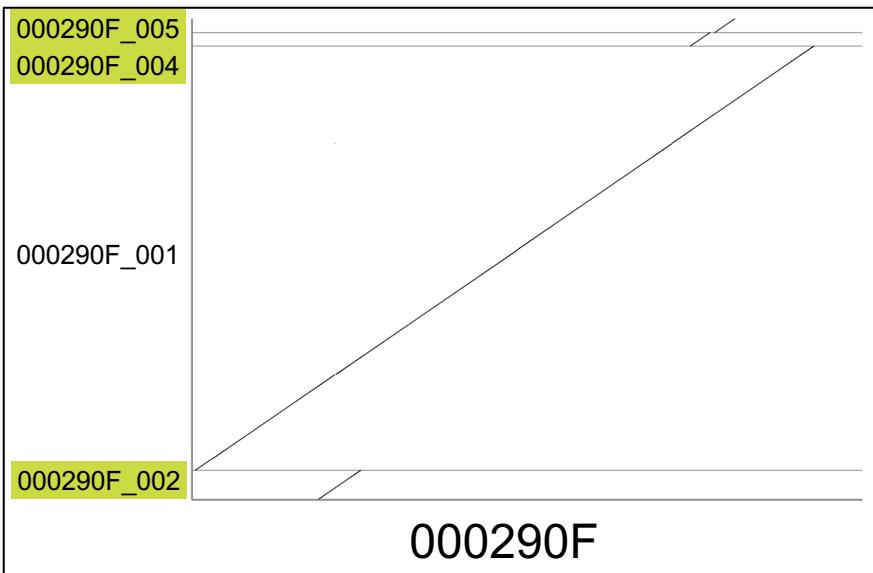
## Remove low quality contigs

- Filter out contigs with >50% unpolished bases (lowercase)

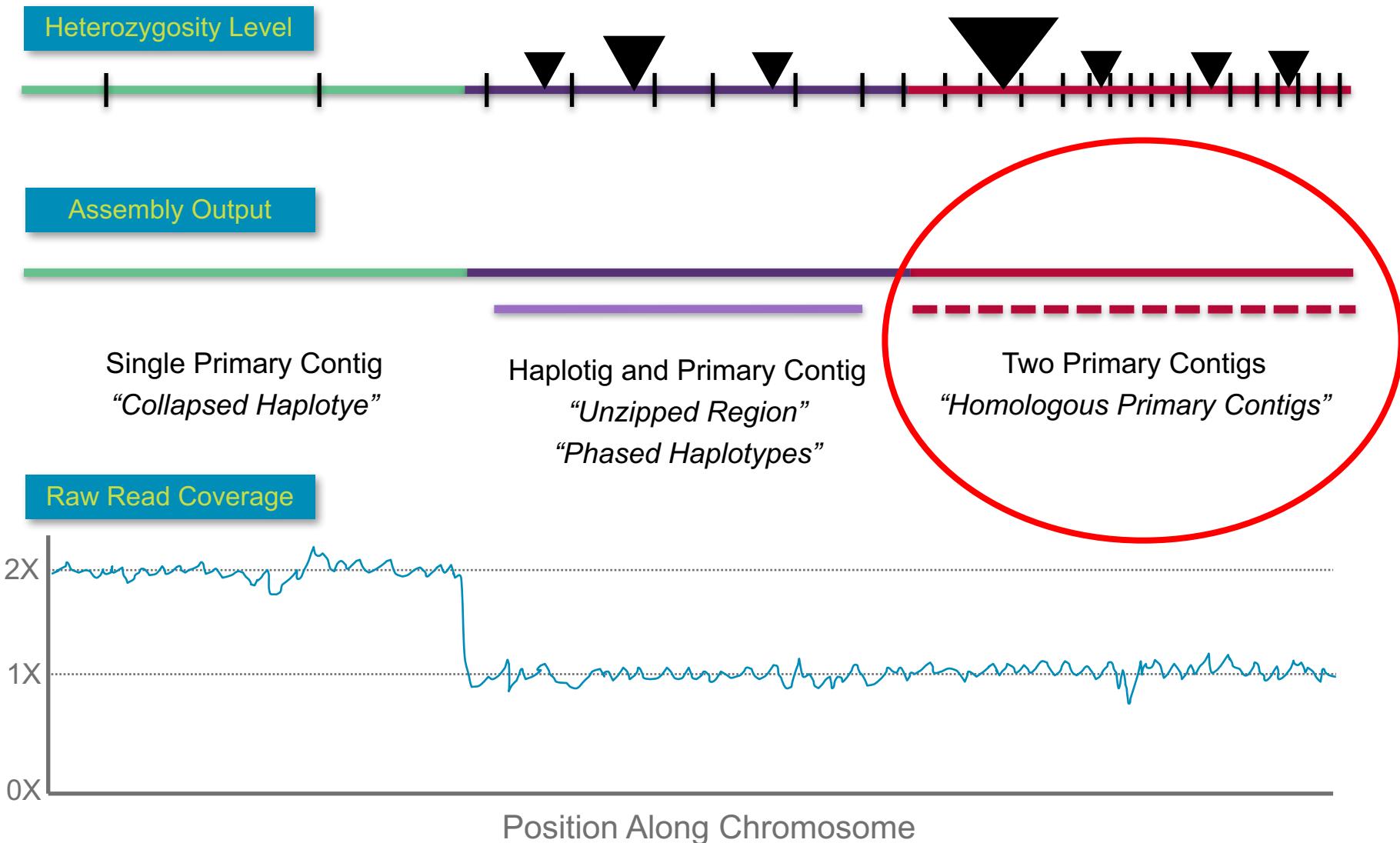
## Remove nested haplotigs

- Short haplotigs that align within longer haplotigs

## HAPLOTIGS ALIGNED TO PRIMARY



# DEDUPLICATING PRIMARY CONTIGS



## DOES MY ASSEMBLY HAVE HOMOLOGOUS PRIMARY CONTIGS?

**Primary assembly length is longer than haploid genome size**

- Inbred individual: diploid assembly: assembly length = 1N
- F1 hybrid: haploid assembly: assembly length = 2N

**Haploid coverage on primary contigs in regions without haplotigs**

**BUSCO analysis on primary contigs indicates widespread duplicated genes**

# METHODS TO IDENTIFY HOMOLOGOUS PRIMARY CONTIGS

## BUSCO/Gene Annotation

- Pros: simple, works for highly divergent haplotypes
- Cons: unannotated contigs excluded
- Usage Case: high contiguity assembly, highly divergent haplotypes

## All-By-All Alignments

- Pros: simple
- Cons: high compute time/manual curation
- Usage Case: small genome (<1 Gb)

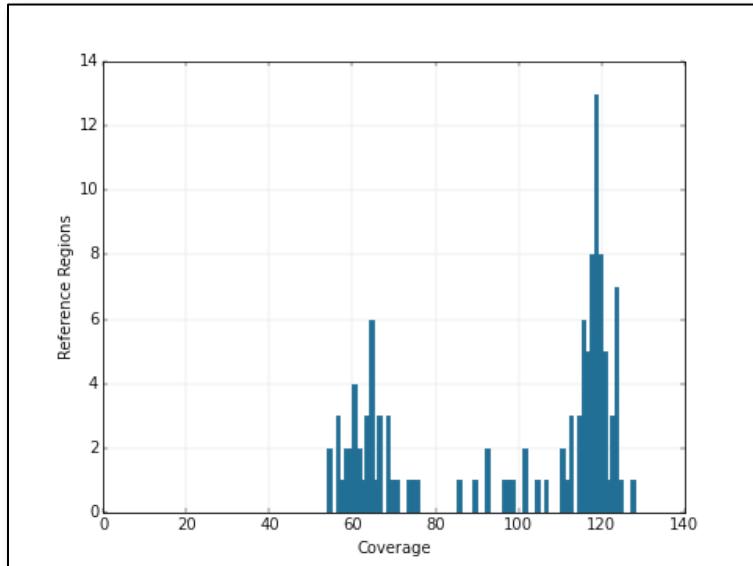
## Purge Haplotypes Pipeline (Mike Roach)

- Pros: uses coverage and pairwise identity
- Cons: some manual curation
- Usage Case: many

## EXAMPLE: AEDES MOSQUITO FALCON-UNZIP ASSEMBLY

- Expected Genome Size: ~1.3 Gb
- Primary Contig Length: 1.69 Gb

### BIMODAL COVERAGE HISTOGRAM



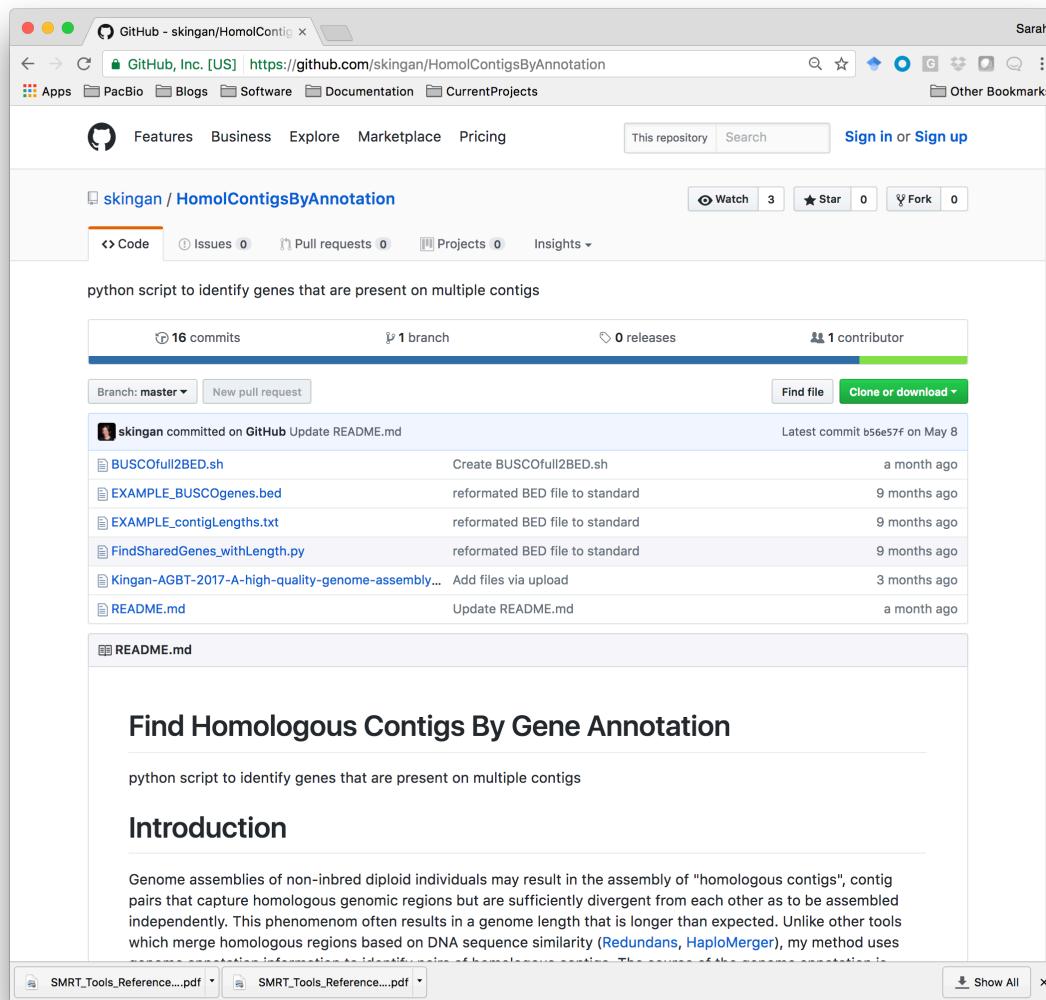
### BUSCO ANALYSIS: ARTHROPOD GENESET (N = 2675)

ASSEMBLY	Aedes PACBIO
COMPLETE	98%
MISSING	2%
FRAGMENTED	10%
DUPLICATED	32%

**Acknowledgement:**  
**Aedes Genome Working Group**  
**Leslie Vosshall, Ben Matthews,**  
**Rockefeller University**

# BUSCO METHOD

[github.com/skingan/HomolContigsByAnnotation](https://github.com/skingan/HomolContigsByAnnotation)



python script to identify genes that are present on multiple contigs

16 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Find file Clone or download

skingan committed on GitHub Update README.md Create BUSCOfull2BED.sh reformatted BED file to standard reformatted BED file to standard reformatted BED file to standard Add files via upload Update README.md

Latest commit b56e57f on May 8 a month ago 9 months ago 9 months ago 9 months ago 3 months ago a month ago

README.md

**Find Homologous Contigs By Gene Annotation**

python script to identify genes that are present on multiple contigs

**Introduction**

Genome assemblies of non-inbred diploid individuals may result in the assembly of "homologous contigs", contig pairs that capture homologous genomic regions but are sufficiently divergent from each other as to be assembled independently. This phenomenon often results in a genome length that is longer than expected. Unlike other tools which merge homologous regions based on DNA sequence similarity ([Redundans](#), [Haplomerger](#)), my method uses

SMRT\_Tools\_Reference....pdf SMRT\_Tools\_Reference....pdf Show All

*Convert BUSCO  
Output to BED*



**Input: BED-formatted Gene Annotation**



**Output: list of contig pairs that share genes**



*User Curation of Contig Pairs*



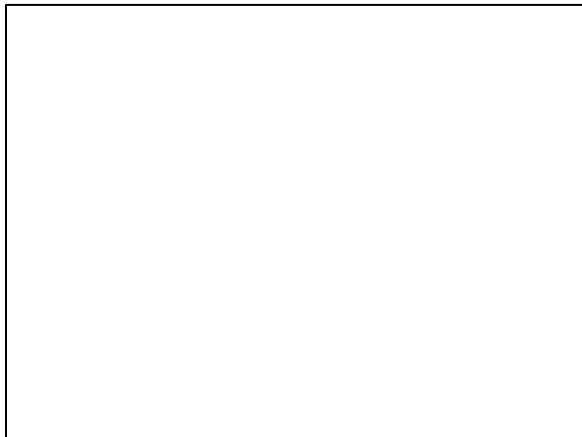
## ALL-BY-ALL ALIGNMENTS

[https://github.com/gconcepcion/chain\\_filter](https://github.com/gconcepcion/chain_filter)

### Automated shell script for contig alignments

- Each contig aligned to all shorter contigs using NUCmer
  - uses multi-threaded version of MUMmer4
- High quality alignments filtered and output in BED format
- Multi-sequence alignment can be rerun and visualized

#### SINGLE ALIGNMENT



#### MULTIPLE ALIGNMENT



# PURGE HAPLOTIGS

MIKE ROACH, AUSTRALIAN WINE RESEARCH INSTITUTE

[https://bitbucket.org/mroachawri/purge\\_haplots/](https://bitbucket.org/mroachawri/purge_haplots/)

## Semi-automated pipeline to remove haplotigs from primary contigs

- Input: BAM of mapped PacBio reads to primary contigs
- Output: curated haploid representation of assembly
  - Record of association between excluded and retained primary contigs

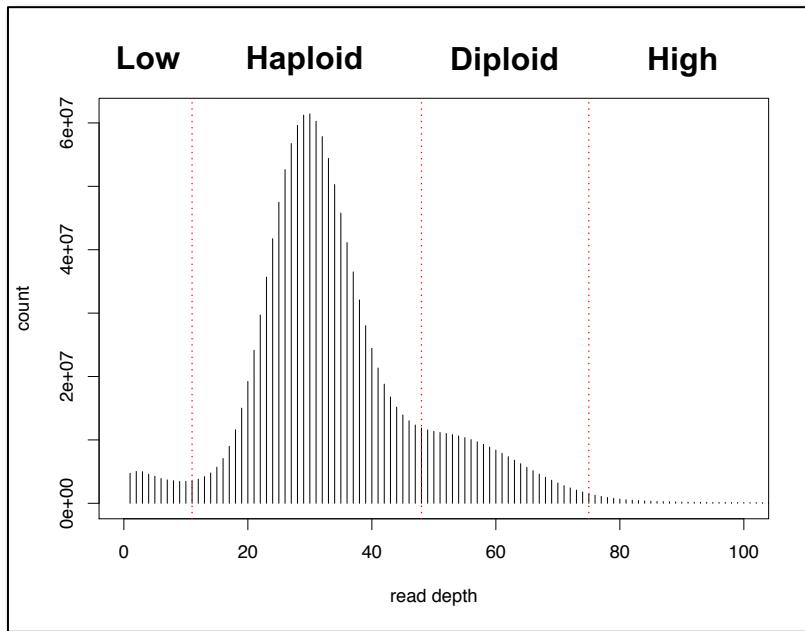
## Pipeline

1. coverage histogram and user-defined coverage cut offs
2. contig-specific coverage analysis to identify candidates
3. alignments and iterative purging of candidates

# PURGE HAPLOTIGS: EXAMPLE FROM BARBERRY

Acknowledgement:  
Iago Hale, UNH

## 1. PRIMARY CONTIG COVERAGE HISTOGRAM



## 2. INDIVIDUAL CONTIG COVERAGE ASSESSMENT

- <80% contig length has diploid coverage
- 4470 / 4672 contigs flagged as “suspect”



## 3. ITERATIVE REASSIGNMENT

- All-by-all BLAST to assign two best hits to “suspect” contigs
- NUCmer alignment and summary stats
- Categorization as “repeat” or “haplotig”

## OUTPUT: CURATED ASSEMBLY

- Revised haploid genome
- Log of reassignment

```
000000F,PRIMARY <- 001282F,REPEAT  
                  <- 003081F,HAPLOTIG
```



# METHODS TO IDENTIFY HOMOLOGOUS PRIMARY CONTIGS

## BUSCO/Gene Annotation

— [github.com/skingan/HomolContigsByAnnotation](https://github.com/skingan/HomolContigsByAnnotation)

## All-By-All Alignments

— [https://github.com/gconcepcion/chain\\_filter](https://github.com/gconcepcion/chain_filter)

## Purge Haplotypes Pipeline (Mike Roach)

— [https://bitbucket.org/mroachawri/purge\\_haplotigs/](https://bitbucket.org/mroachawri/purge_haplotigs/)

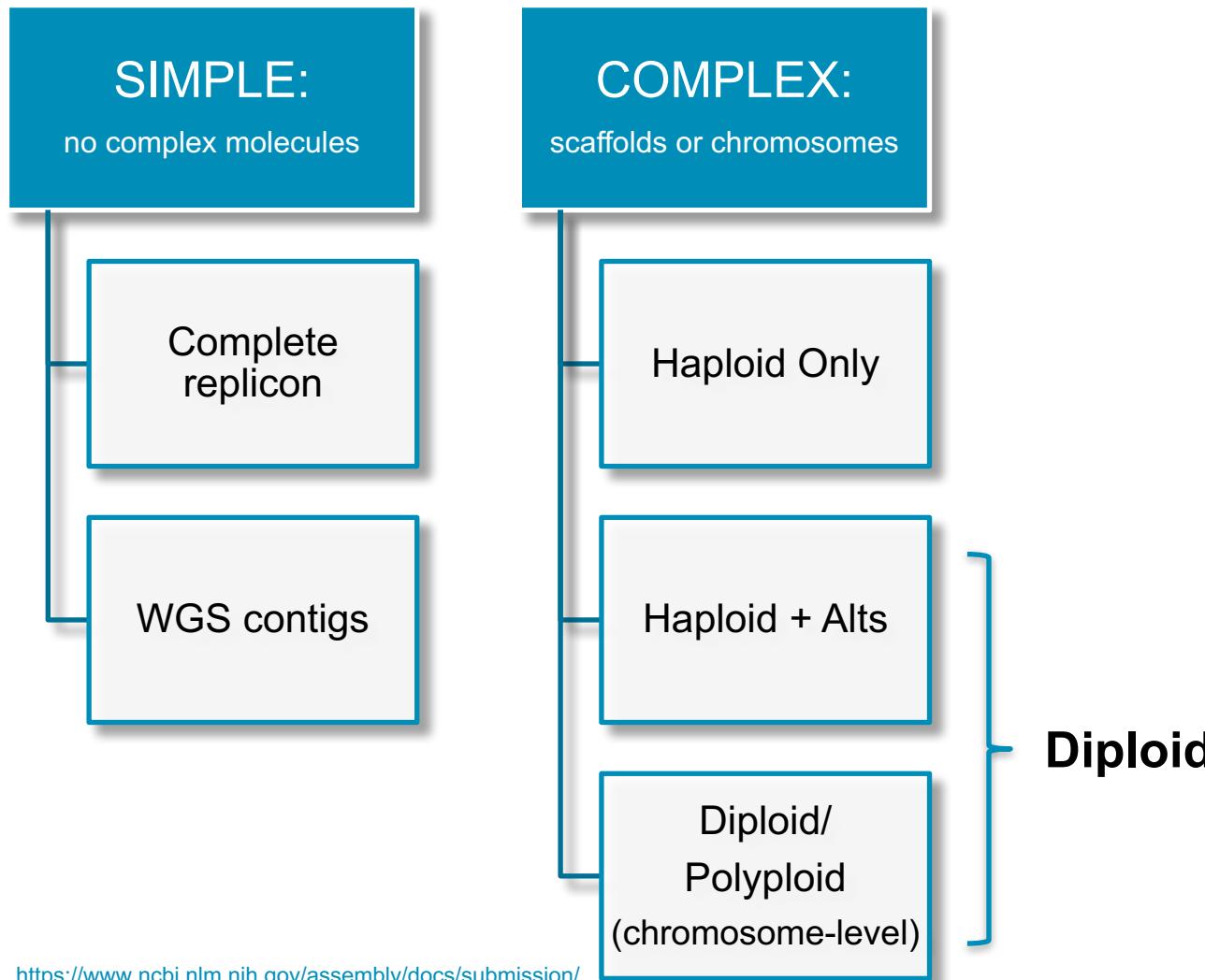


PACBIO®

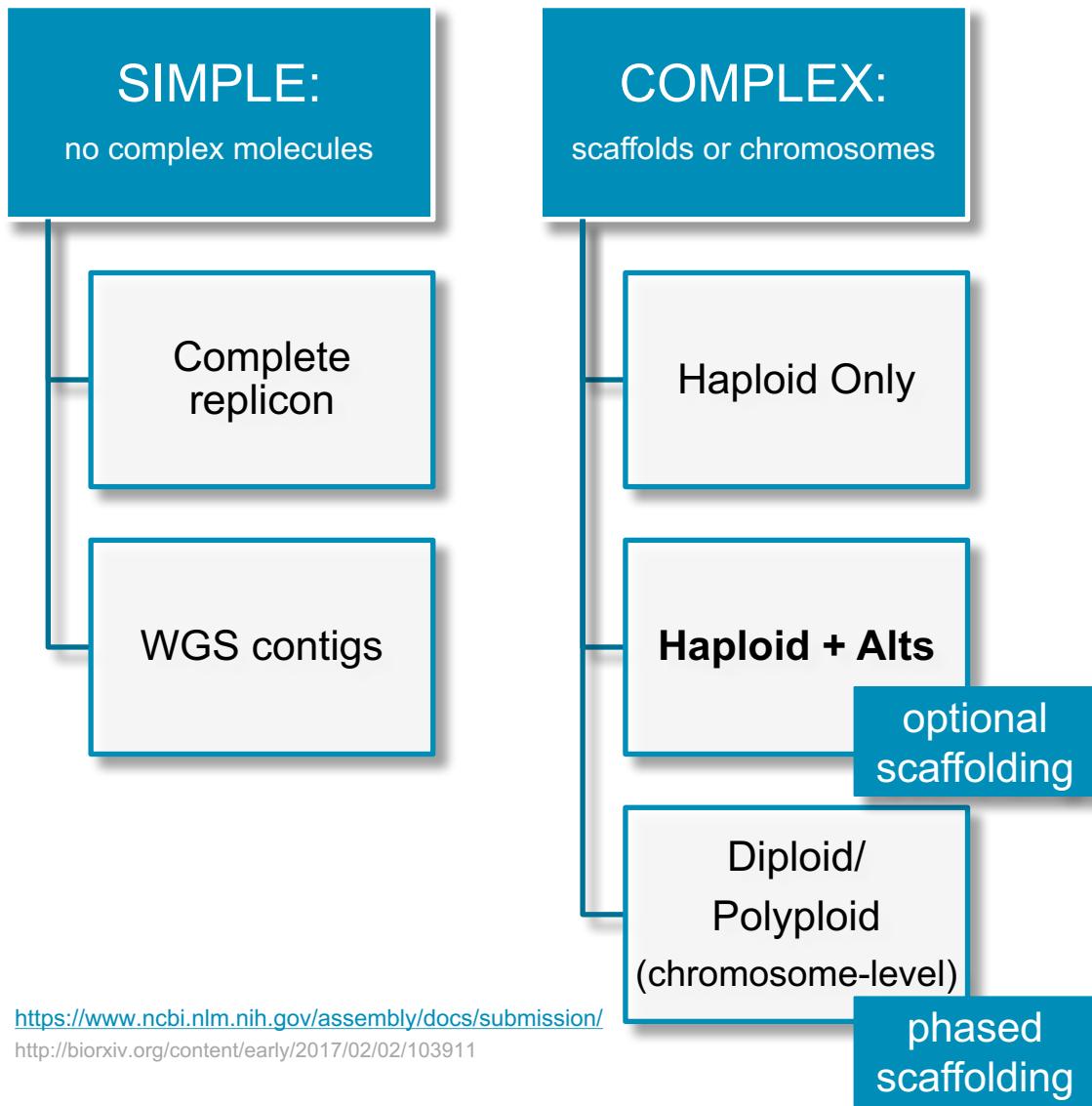
# Sharing Your Assembly

Diploid Assembly Submission to NCBI

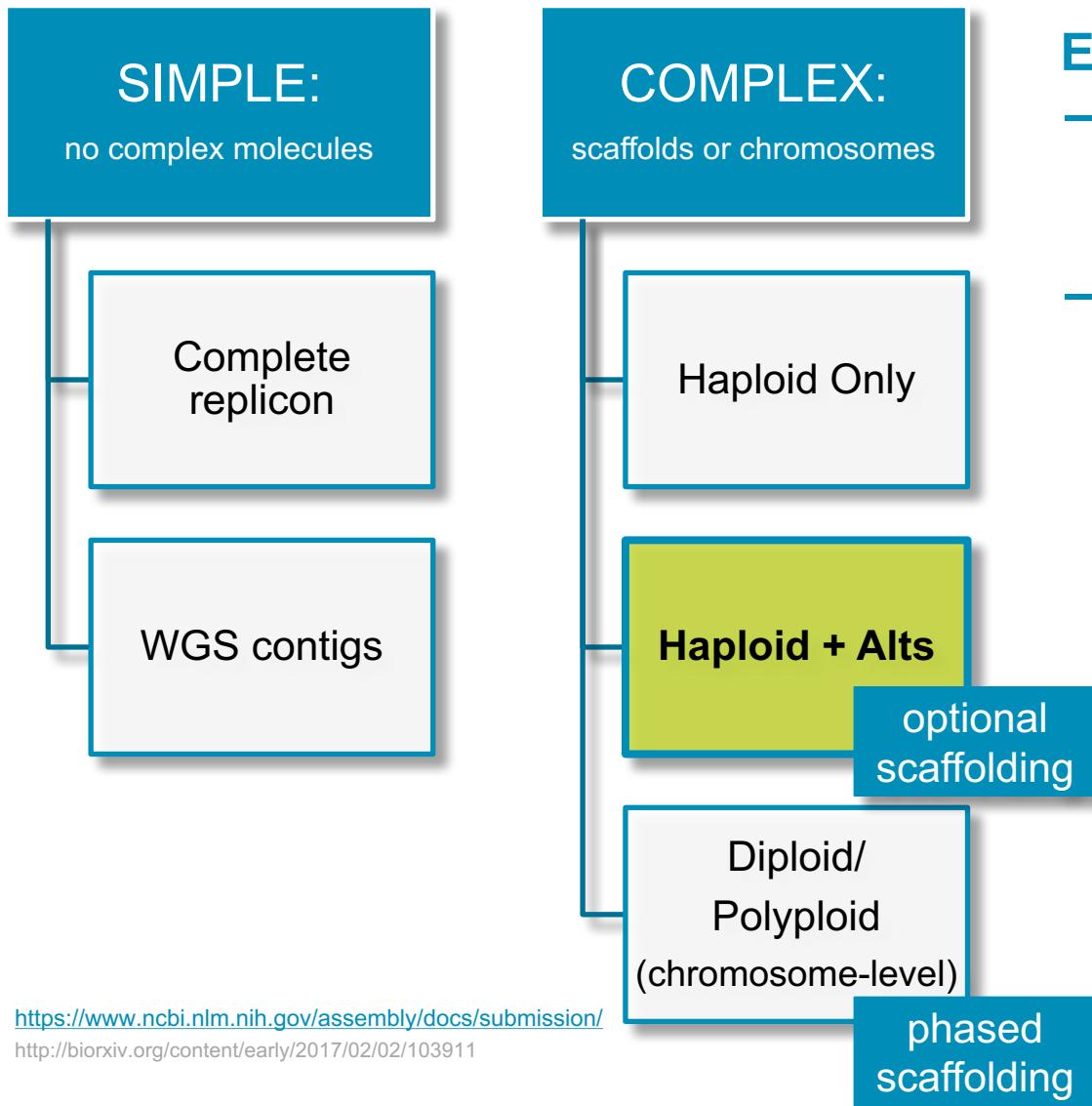
# ASSEMBLY SUBMISSION TYPES



# ASSEMBLY SUBMISSION TYPES



# ASSEMBLY SUBMISSION TYPES



## Examples of Haploid + Alts

- GCA\_001753755.2
  - *Arabidopsis thaliana* F1 from Chin et al. 2016
- GCA\_002008985.2
  - zebra finch Korlach et al. 2017

# ASSEMBLY UNITS: PRIMARY AND ALT

Zebra finch: GCA\_002008985.2

The screenshot shows a web browser window with the title "Tgut\_diploid\_1.0 - Assembly". The URL in the address bar is [https://www.ncbi.nlm.nih.gov/assembly/GCA\\_002008985.2#/def\\_asm\\_haplotigs](https://www.ncbi.nlm.nih.gov/assembly/GCA_002008985.2#/def_asm_haplotigs). The browser interface includes a menu bar with "Sarah" and a toolbar with "Other Bookmarks". The main content area has tabs for "Assembly Definition" and "Assembly Statistics". Under "Assembly Definition", there is a section titled "Global assembly definition" with a note: "Click on the table row to see sequence details in the table to the right". Below this is a table with columns: "Assembly Unit Name", "Sequence name", "Chromosome assignment", "GenBank ID", and "n/a". The "Assembly Unit Name" column shows "Primary Assembly" followed by "haplotigs", which is highlighted with a blue border. The "Sequence name" column lists various sequence identifiers, and the "GenBank ID" column lists corresponding identifiers like MUGN01001535.1.

Assembly Unit Name	Sequence name	Chromosome assignment	GenBank ID	n/a
Primary Assembly	000002F_097	na	MUGN01001535.1	n/a
haplotigs	000003F_001	na	MUGN01002051.1	n/a
	000035F_034	na	MUGN01002022.1	n/a
	000048F_017	na	MUGN01002877.1	n/a
	000054F_006	na	MUGN01001978.1	n/a
	000056F_015	na	MUGN01001387.1	n/a
	000070F_006	na	MUGN01001544.1	n/a
	000076F_003	na	MUGN01001493.1	n/a
	000101F_020	na	MUGN01002686.1	n/a
	000116F_002	na	MUGN01001811.1	n/a
	000162F_004	na	MUGN01002109.1	n/a

- FALCON-Unzip:
  - Primary Contigs and Haplotigs
- FALCON:
  - Primary Contigs and Associated Contigs

# REGIONS

Zebra finch: GCA\_002008985.2

Tgut\_diploid\_1.0 - Assembly - Sarah

Secure https://www.ncbi.nlm.nih.gov/assembly/GCA\_002008985.2/#/def\_region-region256

Apps PacBio Blogs Software Documentation CurrentProjects Other Bookmarks

Assembly Definition Assembly Statistics

Global assembly definition

Download the full sequence report

Region: region256 000051F:2729484-2810743

Sequence name	Chromosome assignment	GenBank ID	RefSeq ID	Sequence role
000051F_069	MUGN01001394.1	n/a	n/a	alt-scaffold

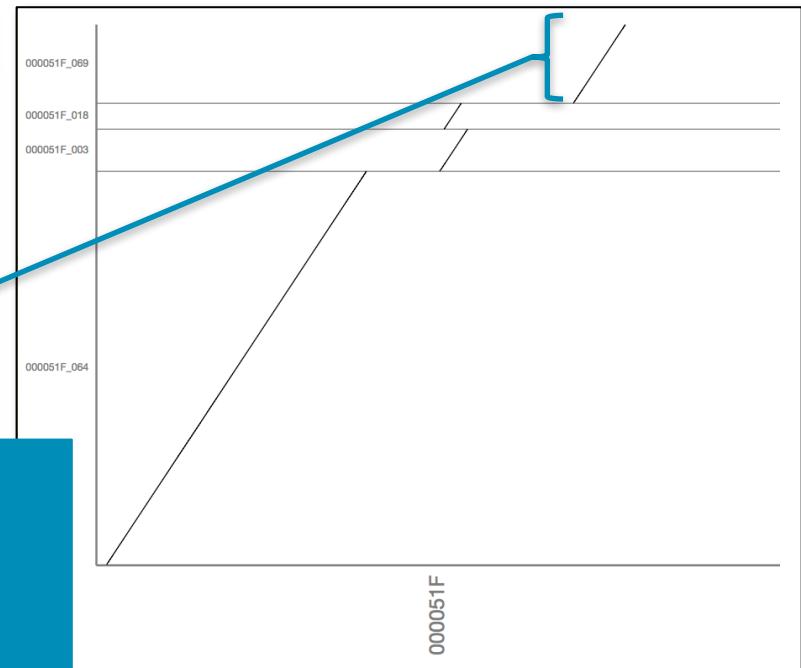
Click on the table row to see sequence details in the table to the right

Assembly Unit Name  
Primary Assembly  
haplotigs

Regions

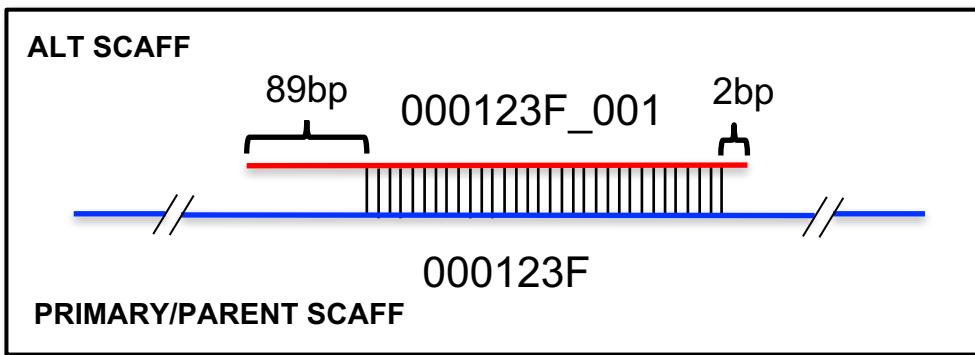
Name	Location
region252	000051F:1442298-1540929
region253	000051F:1568019-2025689
region254	000051F:2035102-2422023
region255	000051F:2530758-2572306
<b>region256</b>	<b>000051F:2729484-2810743</b>
region257	000051F:2834537-2900690
region258	000051F:3067152-3115221
region259	000051F:3138398-3198125
region260	000051F:3329914-3433901
region261	000051F:3506509-3541465

"regions" defined by single  
haplotigs or overlapping  
haplotigs



# ALTERNATE LOCUS PLACEMENT FILE

- Required for haploid + alts submission
- Details placement of alt sequences relative to primary assembly



HEADER	EXAMPLE
<code>alt_asm_name</code>	haplotigs
<code>prim_asm_name</code>	Primary Assembly
<code>alt_scaf_name</code>	000123F_001
<code>parent_type</code>	SCAFFOLD
<code>parent_name</code>	000123F
<code>ori</code>	+
<code>alt_scaf_start</code>	90
<code>alt_scaf_stop</code>	41595
<code>parent_start</code>	8663681
<code>parent_stop</code>	8708427
<code>alt_start_tail</code>	89
<code>alt_stop_tail</code>	2

# PLACEMENT FILE TOOLS

[https://github.com/skingan/NCBI\\_DiploidAssembly](https://github.com/skingan/NCBI_DiploidAssembly)

## generate\_placement.py

- Written by Jason Chin
- Runs NUCmer and generates placement file

## nucmer2ncbiPlacement.py

- Written by Sarah Kingan
- generates placement file from directory of filtered NUCmer alignments
- Contains suggested MUMmer commands

The screenshot shows a GitHub repository page for `skingan / NCBI_DiploidAssembly`. The repository has 15 commits, 1 branch, 0 releases, and 1 contributor. The commits are listed as follows:

Commit	Description	Date
skingan committed on GitHub	Update README.md	Latest commit f1dfba on May 17
AthalPlacementFile.txt	Create AthalPlacementFile.txt	5 months ago
README.md	Update README.md	a month ago
generate_placement.py	Create generate_placement.py	5 months ago
nucmer2ncbiPlacement.py	minor clean up	4 months ago

The page also features a section titled "Diploid Assembly Submission to NCBI" which states: "NCBI now accepts diploid genome submissions! This repository contains a collection of scripts to aid you in generating the placement file required by NCBI. Details can be found [here](#). Refer to the *Arabidopsis thaliana* assembly from Chin et al. 2016 as you prepare your FALCON-Unzip assembly for submission to NCBI."

### Placement File

The placement file for an unscaffled assembly has the following fields:

1. alt\_asm\_name: name of the assembly-unit that includes the alternate scaffold.
2. prim\_asm\_name: name of the assembly-unit on which the alternate scaffold is being placed. Expected to be 'Primary Assembly' in most cases.
3. alt\_scaf\_name: name of the alternate scaffold being placed
4. parent\_type: type of object on which the alternate scaffold is being placed, either CHROMOSOME or SCAFFOLD

## RESOURCES

### FALCON

- <http://pb-falcon.readthedocs.io/>
- <https://github.com/PacificBiosciences/FALCON-integrate>
- <https://github.com/PacificBiosciences/apps-scripts/tree/master/FALCONAssemblyTools>

### SMRT Analysis

- <http://www.pacb.com/support/software-downloads/>
- [http://programs.pacificbiosciences.com/l/1652/2017-02-01/3rzxn6/184345/SMRT\\_Tools\\_Reference\\_Guide\\_\\_v4.0.0\\_.pdf](http://programs.pacificbiosciences.com/l/1652/2017-02-01/3rzxn6/184345/SMRT_Tools_Reference_Guide__v4.0.0_.pdf)



PACBIO®

[www.pacb.com](http://www.pacb.com)

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2017 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSEngine are trademarks of GenDx. FEMTO Pulse and Fragment Analyzer are trademarks of Advanced Analytical Technologies.

All other trademarks are the sole property of their respective owners.