



Shilpa Garg, Ph.D.

Assistant Professor,
Department of Biology
University of Copenhagen

Research Snapshot

Research Area	Genome assembly
Species/ Sample Type	Human genomic DNA
Arima Product	Arima Genome Assembly HiC kit
Application/ Workflow	Genome Assembly

Chromosome-Scale, Haplotype-Resolved Assembly of Human Genomes

Humans are diploid organisms, typically containing two copies of every chromosome. However, most genome assemblies are represented as haploid and miss heterozygous sequences. To accurately represent diploid genomes, genome assemblies are “phased” to achieve haplotype-level resolution. No one has been able to achieve chromosome-long phasing without using the parent genomes, which are not always available. To overcome this, diploid assembly (DipAsm) combines long-read assembly and Hi-C-based phasing for single individuals to generate a chromosome-scale phased assembly within one day.

Garg, S., et al. (2021) [Chromosome-scale, haplotype-resolved assembly of human genomes](#). *Nature Biotechnology* **39**, 309-312.

Research Question

Can we accurately produce phased assembly for a single individual?

How Did Arima Genomics Make a Difference?

“The Arima team is easy to work with and very collaborative. The technical support really made the difference in our project.”

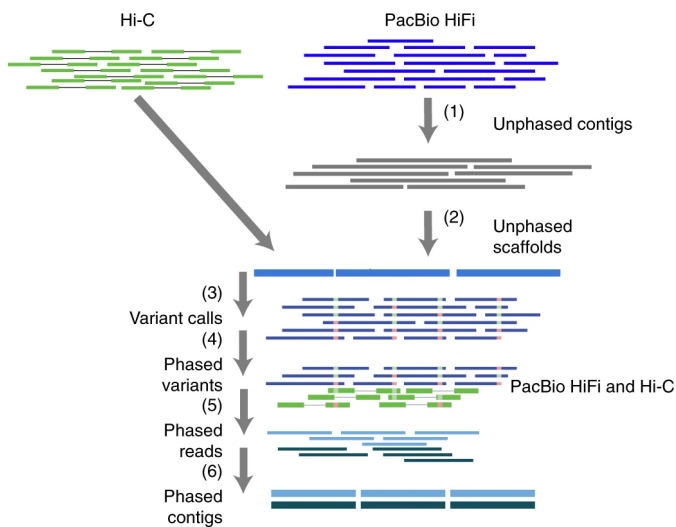


Technology

With the Arima Genome Assembly Kit you can take your genome assembly from draft to reference quality with our efficient 6-hour workflow and data that is compatible with the latest assembly and scaffolding pipelines.

Experiment Overview

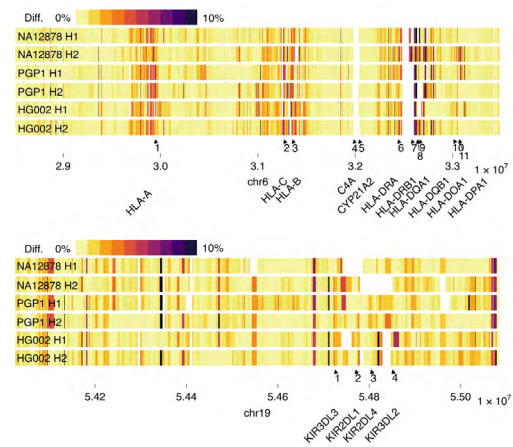
- Convert genomic DNA from four genomes into a SMRTbell™ (PacBio) library
- Sequence on PacBio Sequel System to yield long high-fidelity (HiFi) reads
- Hi-C sequencing with the Arima High Coverage Hi-C kit
- Phased sequence assembly
 - » Scaffold an unphased Peregrine¹ assembly with 3D-DNA² or HiRise³, Call small variants with DeepVariant⁴
 - » Phase variants with WhatsHap⁵ and HapCUT2⁶
 - » Partition the reads and assemble each partition independently with Peregrine again
- Evaluate variant calling accuracy



Results and Future Directions

DipAsm accurately produces chromosome-long phased assembly using data from PacBio HiFi and Arima High-Coverage Hi-C. In contrast to trio binning, this method does not use pedigree data and can phase *de novo* mutations. The Hi-C portion of the method renders it easier to use and more widely adoptable than alternative methods like Strand-seq.

Using this approach, the team was able to fully phase the assemblies and reconstruct most of the HLA and KIR regions with two contigs for each haplotype. Based on the patterns of local sequence divergence they determined that the two haplotypes in each individual are distinct from one another. They concluded that their assembly approach can be used to phase even these highly polymorphic regions.



The *de novo* method is a milestone since generating an assembly without a reference sequence furthers the goals of the Human Genome Reference Project, which aims for unbiased characterization of human genome diversity. Sequencing polymorphic regions at high resolution will also advance personalized, precision medicine.

1. Chin C-S, Khalak A. Human Genome Assembly in 100 Minutes. *bioRxiv*. 2019:705616.
2. Dudchenko O, Batra SS, Omer AD, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356(6333):92-95.
3. Putnam NH, O'Connell BL, Stites JC, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research*. 2016;26(3):342-350.
4. Poplin R, Chang P-C, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nature biotechnology*. 2018;36(10):983-987.
5. Martin M, Patterson M, Garg S, et al. WhatsHap: fast and accurate read-based phasing. *bioRxiv*. 2016:085050.
6. Edge P, Bafna V, Bansal V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Research*. 2017;27(5):801-812.



Learn More

[Where Will Genomes Take Us Next: How Chromosome-Scale Assemblies Are Unlocking New Biology](#)