

ASSEMBLY OPTIONS FOR YOUR SMRT SEQUENCING DATA

THE LEADER IN LONG-READ SEQUENCING



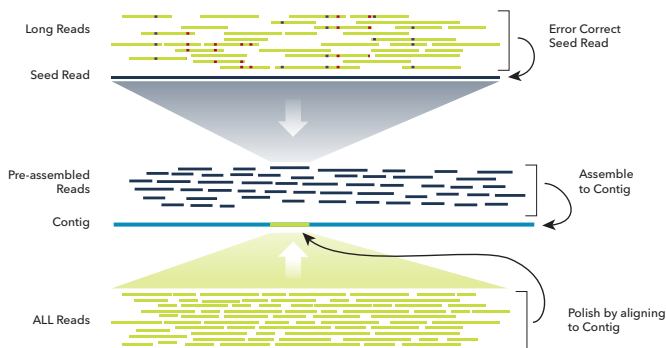
PLANT AND ANIMAL SCIENCES

It is now easier and more affordable than ever to do a genome project with PacBio® Single Molecule, Real-Time (SMRT®) Sequencing—the gold standard for generating contiguous, highly accurate reference genomes. Assembly is no longer a challenge with the wealth of bioinformatics tools developed and optimized for SMRT Sequencing data, enabling you to generate high-quality genome assemblies on **any budget**.

PACBIO GENOME ASSEMBLY

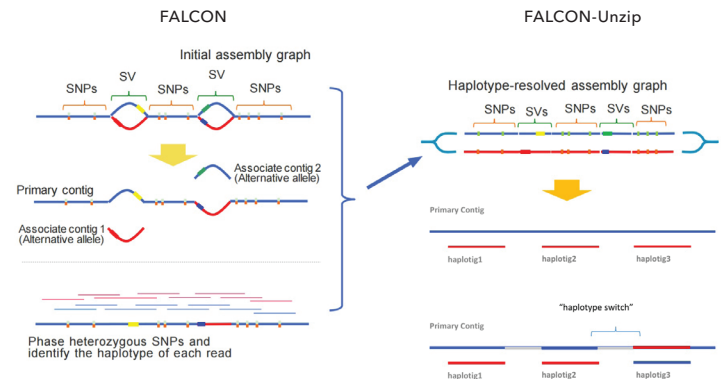
Using the Hierarchical Genome Assembly Process (HGAP) and long-read polishing, SMRT Sequencing data produces genome assemblies with megabase-size contig N50s and >99% accuracy. Explore the options for assembling SMRT Sequencing data below.

SMRT ANALYSIS



- Free for anyone to download and use at pacb.com/software-downloads
- Includes HGAP4, our push-button assembly tool with polishing built into the pipeline
- Use with genomes ≤ 3 Gb

FALCON AND FALCON-Unzip



- Open source for anyone to use at pacb.com/devnet
- FALCON: specialized assembly parameterization for genomes of any size
- FALCON-Unzip: genome phasing of heterozygous or outbred organisms

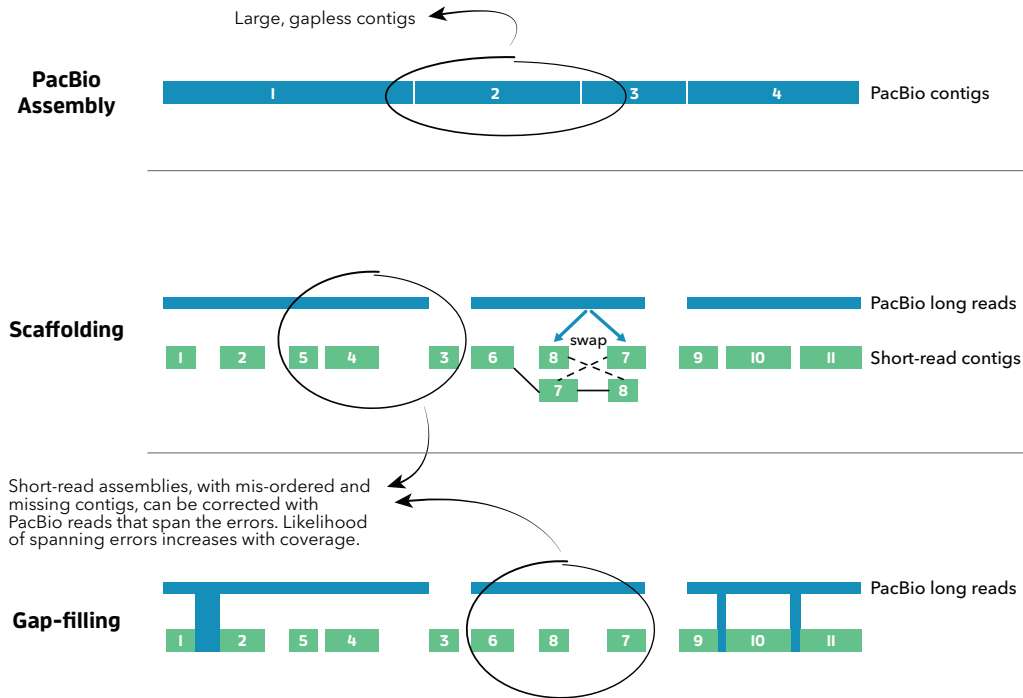
ANALYSIS PARTNERS



With these PacBio analysis partners, you have the option to completely outsource your bioinformatics or compute resource needs. Options range from cloud-based, self-serve analysis to no-hassle, full-service assembly.

HYBRID GENOME ASSEMBLY

The most contiguous and accurate contig assemblies are generated with SMRT Sequencing data alone. Alternatively, for significantly lower cost than a full PacBio assembly, you can also use SMRT Sequencing data to improve highly fragmented draft genomes through scaffolding of contigs or filling gaps left in short-read assemblies.



PUBLICATION SPOTLIGHT

Browse this GigaScience paper¹ where both PacBio and hybrid assemblies were generated for the complex, hexaploid wheat genome.

COVERAGE REQUIREMENTS

Although every genome project is unique in complexity and downstream utility, we have compiled a quick guide below to help you assess the coverage requirements and expected contiguity for genome assembly using SMRT Sequencing data or hybrid assembly approaches.

METHOD	PACBIO COVERAGE	EXPECTED CONTIGUITY	GAPS IN ASSEMBLY
PacBio Assembly	40 to 50-fold per haplotype	Megabases	NO
Scaffolding²	10 to 30-fold	Tens to hundreds of kb	YES
Gap-filling³	5 to 10-fold	Tens to twenties of kb	YES

KEY REFERENCES

- Zimin, A. et al. (2017) The first near-complete assembly of the hexaploidy bread wheat genome, *Triticum aestivum*. GigaScience 6 (11), 1-7.
- Uliano-Silva, M. et al. (2018) A hybrid-hierarchical genome assembly strategy to sequence the invasive golden mussel, *Limnoperna fortunei*. GigaScience 7 (2), 128.
- Xu, S. et al. (2017) Wild tobacco genomes reveal the evolution of nicotine biosynthesis. PNAS 114 (23), 6133-6138.

