Comprehensive Genome Assemblies Made Simple

DNAnexus and Azure for AgBio

AGBIO ASSEMBLY CHALLENGES

De novo genome assembly is a complex task made more challenging by high levels of genetic diversity, repetitive elements, and duplicated genomic regions. Plant species are especially challenging because of their higher ploidy and high repeat content.

Recent advances in long-read sequencing have allowed researchers to improve the accuracy of assemblies through longer contiguous sequences, allowing reads to span repeat elements and unambiguously link neighboring sequences. In addition, layering next-generation mapping into a *de novo* assembly provides a genome-wide view to correctly scaffold the sequence data.

The algorithms used to perform these assemblies can be difficult to tune for each species and require massive computational resources to weave these long-reads into a final, polished assembly. Bringing together multiple technologies (long-read, short-read, next-generation mapping) can improve contigs and scaffolding and provide a comprehensive view of genome assemblies.

THE DNANEXUS DIFFERENCE

DNAnexus provides industry-leading reference-quality genome assembly services. With DNAnexus you have a bioinformatics partner to collaborate on complex assembly, structural variation analysis and physical mapping to achieve complete and accurate views of all types of genomic variation.

- Fast, accurate, & cost-efficient solution
- Expert bioinformatics support
- Secure delivery of results & quality metrics
- Eliminate complex local compute infrastructure, applications deployment, & support
- Leverage multi-omics data & gold standard bioinformatics tools
- Collaborate securely & seamlessly
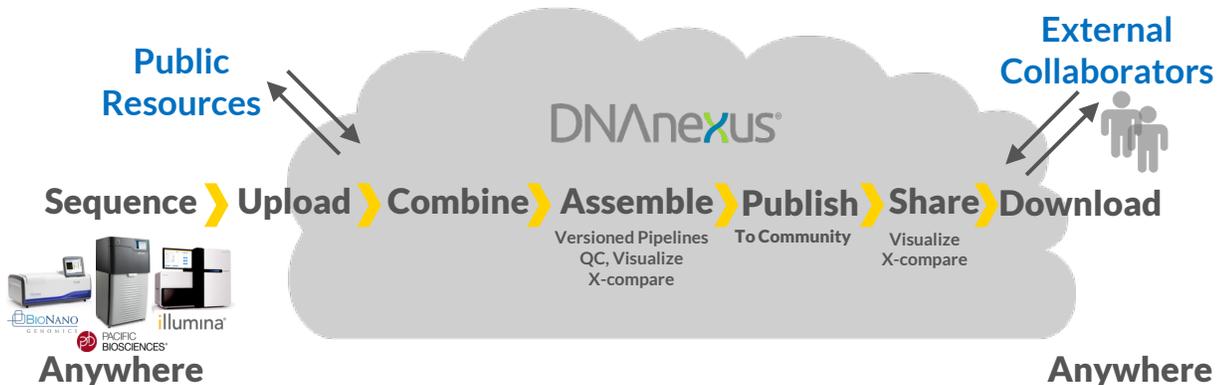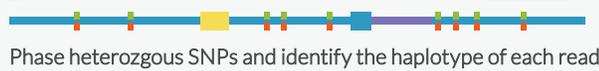
FALCON & FALCON-UNZIP

FALCON is a diploid-aware genome assembler designed for PacBio long-read data. FALCON-Unzip further refines assembled contigs.

- Improves assembly continuity; 30- to >100-fold more contiguous than short-read assemblers
- Generates phased diploid assemblies, enabling the study of haplotype structure & the identification of heterozygous structural variation
- Genome phasing of the human genome HG002 on DNAnexus revealed 90% of haplotype assignment is consistent with informative SNPs
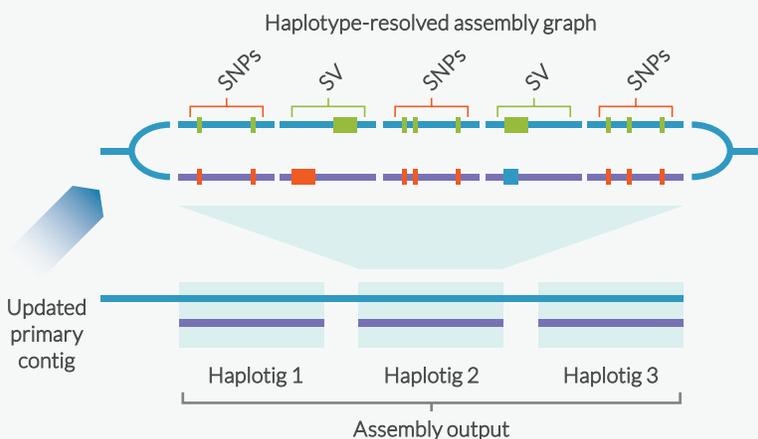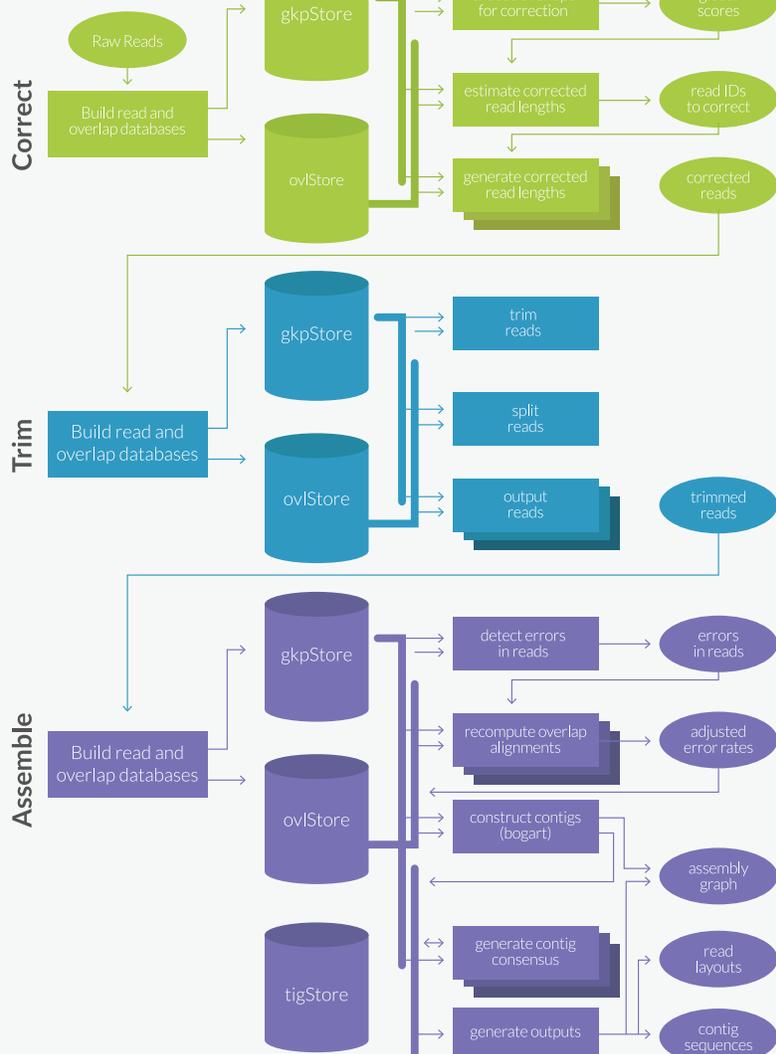- Eliminates complicated installation; ready to run on DNAnexus
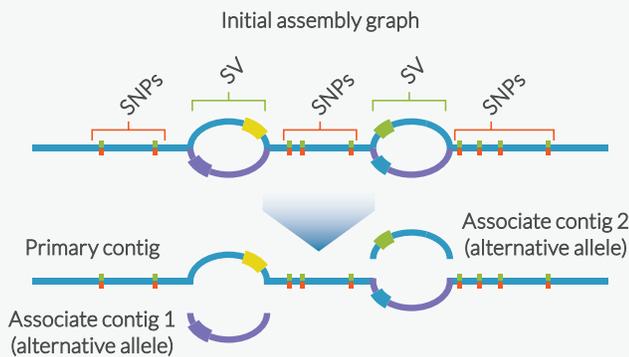
CANU (COMING SOON)

Derived from the Celera assembler, Canu is a genome assembler that is specifically designed for noisy single-molecule sequences.

- Halves depth-of-coverage requirements
- Improves assembly continuity while simultaneously reducing runtime by an order of magnitude on large genomes
- Uses two data instruments – PacBio and Oxford Nanopore for improved assembly
- Eliminates complicated installation; ready to run on DNAnexus

A. FALCON

Initial assembly graph

SNPs, SV, SNPs, SV, SNPs

Primary contig

Associate contig 2 (alternative allele)

Associate contig 1 (alternative allele)

B.

Phase heterozgous SNPs and identify the haplotype of each read

C. FALCON-Unzip

Haplotype-resolved assembly graph

SNPs, SV, SNPs, SV, SNPs

Updated primary contig

Haplotig 1   Haplotig 2   Haplotig 3

Assembly output

CANU

Correct
Raw Reads
Build read and overlap databases
gkpStore
ovlStore
choose overlaps for correction → global scores
estimate corrected read lengths → read IDs to correct
generate corrected read lengths → corrected reads

Trim
Build read and overlap databases
gkpStore
ovlStore
trim reads
split reads
output reads → trimmed reads

Assemble
Build read and overlap databases
gkpStore
ovlStore
tigStore
detect errors in reads → errors in reads
recompute overlap alignments → adjusted error rates
construct contigs (bogart) → assembly graph
generate contig consensus → read layouts
generate outputs → contig sequences

READY TO TAKE YOUR ASSEMBLIES TO DNANEXUS?

DNAnexus is a market leader in *de novo* genome assembly. Our key experience includes 3000 Rice Genomes Project, Vertebrate Genomes Project, along with many complex individual assemblies including the tobacco genome (4.5 billion base pairs, tetraploid & highly repetitive).

Let us be your bioinformatics partner. Contact us at info@dnanexus.com.