PRODUCT HIGHLIGHT

Dovetail™ Hi-C Kit For Genome Assembly

Chromosome level assemblies for your lab

From researchers studying evolutionary biology to genetics, having a contiguous and accurate chromosome-level assembly is the starting point to answering life's biological questions. It can also be its biggest challenge.

The Dovetail™ Hi-C Library Preparation Kit with HiRise™ Scaffolding Software provides labs with direct access to the leading Hi-C genome assembly solution previously only available through Dovetail's service.

FEATURES

- Capture Megabase range information with Dovetail's proximity ligation technology
- Generate chromosome-level assemblies with HiRise™, the leading proximity ligation genome scaffolding software
- Dovetail quality assemblies in your lab; optimized end-to-end from the world's leading service
- Makes running Hi-C in your lab easier than ever with high quality Dovetail™ Hi-C data
- Assemble more genomes to answer more biological questions with a lower cost per sample, on your own schedule
- Compatible with a wide variety of organisms including humans, plants, and animals.

Dovetail™ Hi-C spans Mbps of genomic space needed to generate chromosome level assemblies

Building chromosome-level genome assemblies requires spanning long genomic distances. When building a genome assembly, sequenced DNA fragments are aligned to each other to build progressively longer contigs, with contigs linked together to form scaffolds. Accurate whole chromosome assembly requires sufficiently long scaffolds connected in the correct order and orientation with a technology that can span the distance between the scaffolds. Long-read, linked-read, and mate-pair technologies are limited in their scaffolding abilities by the size of the inserts.

Dovetail™ Hi-C solves the challenge of spanning long genomic distances by utilizing chromatin confirmation capture methods to generate proximity ligated DNA fragments that include regions in close spatial, but distant linear, proximity. Analyzing Dovetail™ Hi-C data with the HiRise™ pipeline enables researchers to accurately assemble genomes up to the chromosome-level.

Dovetail™ Hi-C Kit and HiRise™ software are integrated and optimized for genome assembly

The best approach for generating high-quality assemblies is optimizing the molecular biology technique and scaffolding software together specifically for genome assembly.

With more than a 1000 genomes scaffolded to date, Dovetail has actively optimized the Dovetail™ Hi-C molecular biology protocol and HiRise™ pipeline.

The integrated and optimized Dovetail™ Hi-C solution has been encapsulated into the Dovetail™ Hi-C Kit for you to run in your lab, on your schedule, with your sequencers. For scaffolding, you have the option to either let our highly experienced informatics team scaffold your assembly in-house using our proprietary HiRise™ pipeline or you can scaffold yourself using an open source platform.

Figure 1. Dovetail™ Hi-C proximity ligation workflow

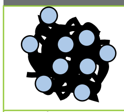
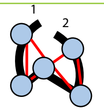
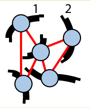
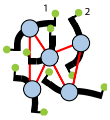
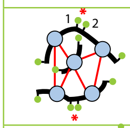
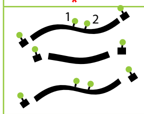
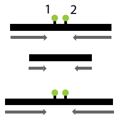
	A	Dovetail™ Hi-C libraries start with endogenous chromatin.
	B	Crosslinking (red lines) the chromatin creates a stabilized nucleosome (blue circles) scaffold.
	C	Restriction endonucleases digest the cross-linked chromatin.
	D	Biotin (green dots) mark digested DNA ends (black lines).
	E	Proximity ligation creates chimeric molecules (ex. 1 and 2).
	F	The crosslinks are reversed.
	G	DNA is purified, and enriched for ligation-containing chimeric molecules. Libraries sequenced as paired-end short reads

Table 1. Examples of N50 improvements for genomes scaffolded using Dovetail™ Hi-C and HiRise™ scaffolding pipeline.

Organism	Genome Size (Gbps)	Dovetail™ Tech	Input N50 (Mbp)	Chicago™ HiRise™ N50 (Mbp)	Dovetail™ Hi-C HiRise™ N50 (Mbp)
Plantain	1.75	Chicago™ + Dovetail™ Hi-C	0.03	1.19	56.71
Coffee	1.19	Dovetail™ Hi-C	1.85	-	82.00
European Common Vole	3.0	ILMN *de novo* + Chicago™ + Dovetail™ Hi-C	0.08	25.67	186.61
African Bullfrog	1.40	Chicago™ + Dovetail™ Hi-C	2.69	15.06	157.52
Rice Weevil	0.50	Chicago™ + Dovetail™ Hi-C	2.29	7.44	53.59
Rattlesnake	1.30	Chicago™ + Dovetail™ Hi-C	0.14	2.47	179.90
Dark-eyed Junco	1.20	ILMN *de novo* + Chicago™ + Dovetail™ Hi-C	0.10	16.74	71.32

In some cases, Dovetail also generated the de novo assembly & performed intermediate range scaffolding using the Chicago™ technique. Meraculous was used for Illumina de novo assemblies.

Dovetail™ Hi-C Kits make running Hi-C in your lab easier than ever

Researchers are increasingly interested in Hi-C data, but significant challenges remain to establishing and maintaining a Hi-C assay in a lab. Available Hi-C protocols are time-intensive and have historically required substantial troubleshooting.

The Dovetail™ Hi-C Kit has been optimized through Dovetail's in-house Dovetail™ Hi-C service. With the Dovetail™ Hi-C Kit you have the option to use your own prefered method of library preparation or you can add the Dovetail™ Library Prep to generate Illumina® sequencing-ready libraries.

The protocol is extremely robust. In a validation study 100% of samples generated high-quality Hi-C sequencing libraries. The protocol for generating proximity-ligated DNA has also been optimized to simplify and shorten the work flow 1 day, greatly reducing the obstacles associated with running the Hi-C assay in your lab.

HiRise™ assembly software scaffolds up to whole chromosomes with high confidence

Dovetail™ Hi-C generates proximity ligation data that links regions of a chromosome together ranging from hundreds of bps to Mbps apart. The HiRise™ software is built on a statistical model that takes into account the unique read pair separation distribution of this data.

Initially, HiRise™ aligns Dovetail™ Hi-C data to scaffolds of the existing draft assembly and generates a custom statistical model based on the read-separation distribution of the Dovetail™ Hi-C data.

The software then scans through the draft assembly for errors, then breaks and rejoins the scaffold when an error is identified. HiRise™ iterates over many candidate arrangements to produce a better scoring assembly, then refines the likelihood model for additional rounds of scaffolding improvement. The pipeline repeats this process until it produces the most confident arrangement of scaffolds supported by the data.

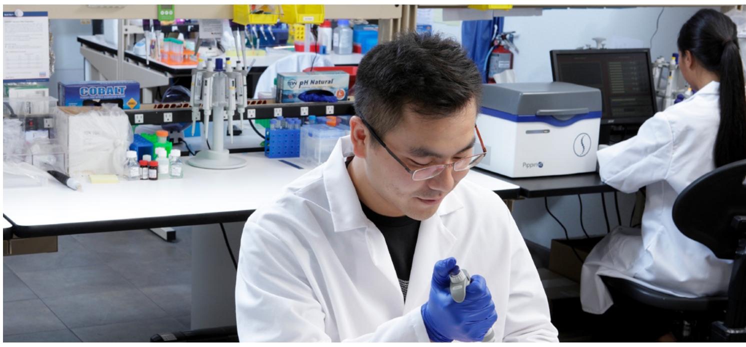


Table 2. Dovetail™ Hi-C Kit Specifications

Number of reactions	8
Sample types	Mammalian tissue, blood and cell culture, plants, and animals
Libraries per genome	1 library per 3 Gbp genome recommended for most genomes*. See protocol for specific recommendations
Required equipment	Standard genomics lab equipment
Reagents not included	General purpose reagents, formaldehyde, DNA purification beads
Optional Add-ons	1. Dovetail™ Library Module for Illumina® 2. Dovetail™ Primer Set for Illumina® 3. Dovetail™ Hi-C Plant and Animal Samples Filter Set

**As determined by QC sequencing metrics.*

Table 3. HiRise™ Software Specifications

Input assembly metrics	Minimum of N50 >1Mbp and N90 >20kb recommended
Input assembly format	FASTA
Input assembly technologies	Virtually all genome assembly technologies and assemblers that produce FASTA files
Dovetail™ Hi-C data format	FASTQ files
Dovetail™ Hi-C sequencing format	2X 150 paired-end reads
Dovetail™ Hi-C sequencing coverage	100M read-pairs per gigabase of haploid genome

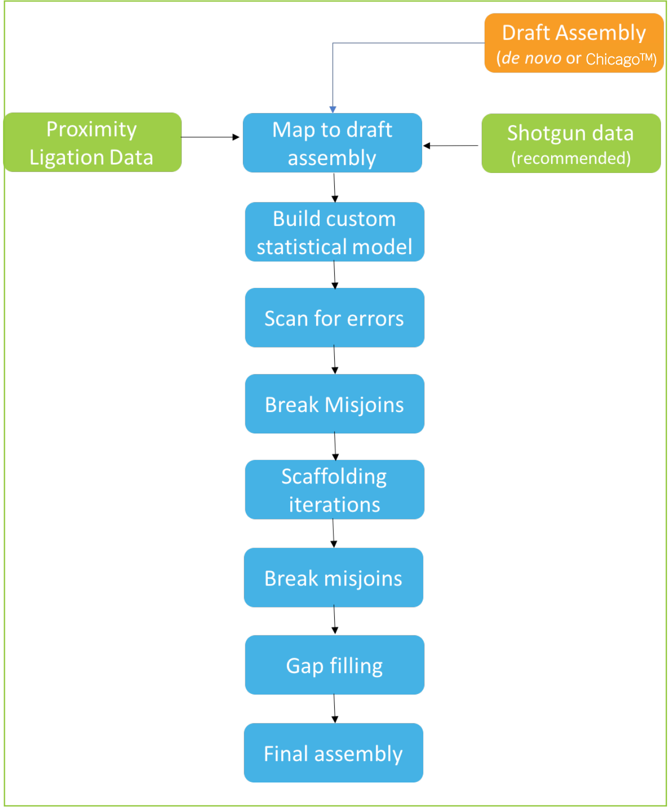
Assemble more genomes and answer more interesting biological questions

For researchers involved in genome assembly, there is a clear recognition of the value of generating assemblies for the organism of interest and not just using the assembly of a model organism. This allows for asking and answering biological questions that cannot be answered satisfactorily using another organism as a reference.

Some of the hurdles to generating more assemblies are the cost per assembly and not being able to control the timeline.

Dovetail™ Hi-C with HiRise™ for genome assembly makes it easier to generate chromosome-level assemblies for all organisms of interest at a lower cost per sample, flexible pooling options from 1 to 8 samples at a time, and the ability to run on your own schedule.

Figure 2: Process workflow of HiRise™ scaffolding pipeline from inputting a draft assembly and Dovetail™ Hi-C sequencing data to output of the final scaffolded assembly



ORDERING INFORMATION

Contact Orders@dovetail-genomics.com or (831) 713-44656