HiRise Scaffolding Report

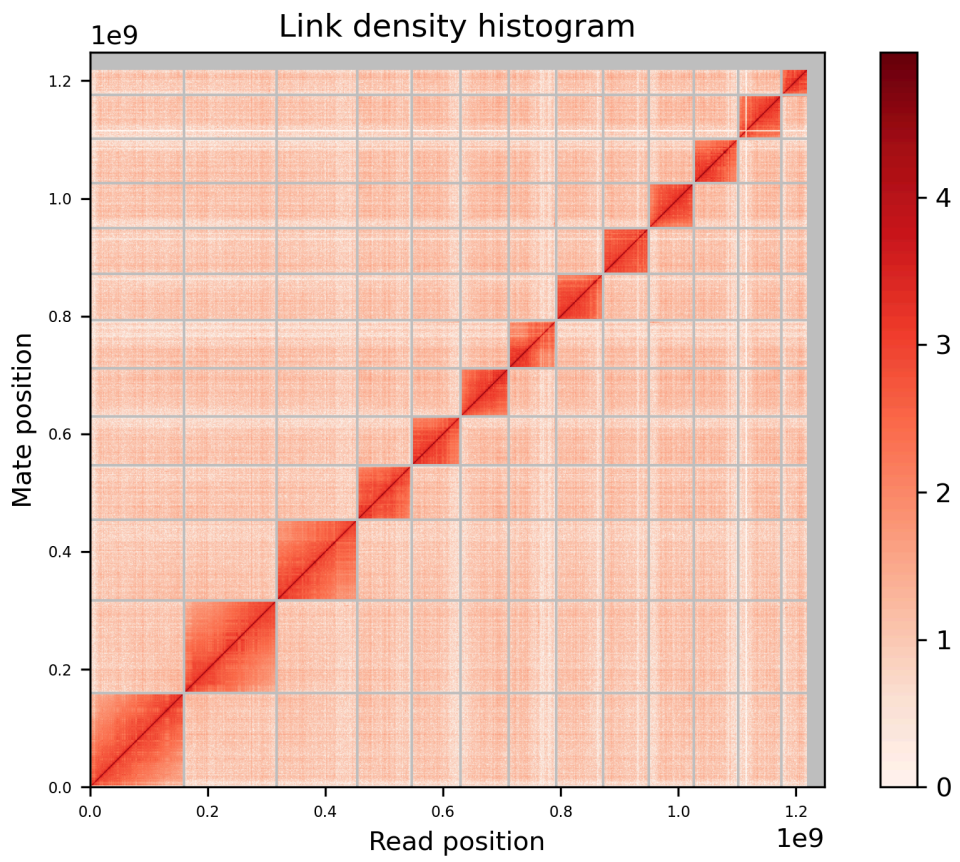
July 28, 2024

Timema cristinae

Patrik Nosil

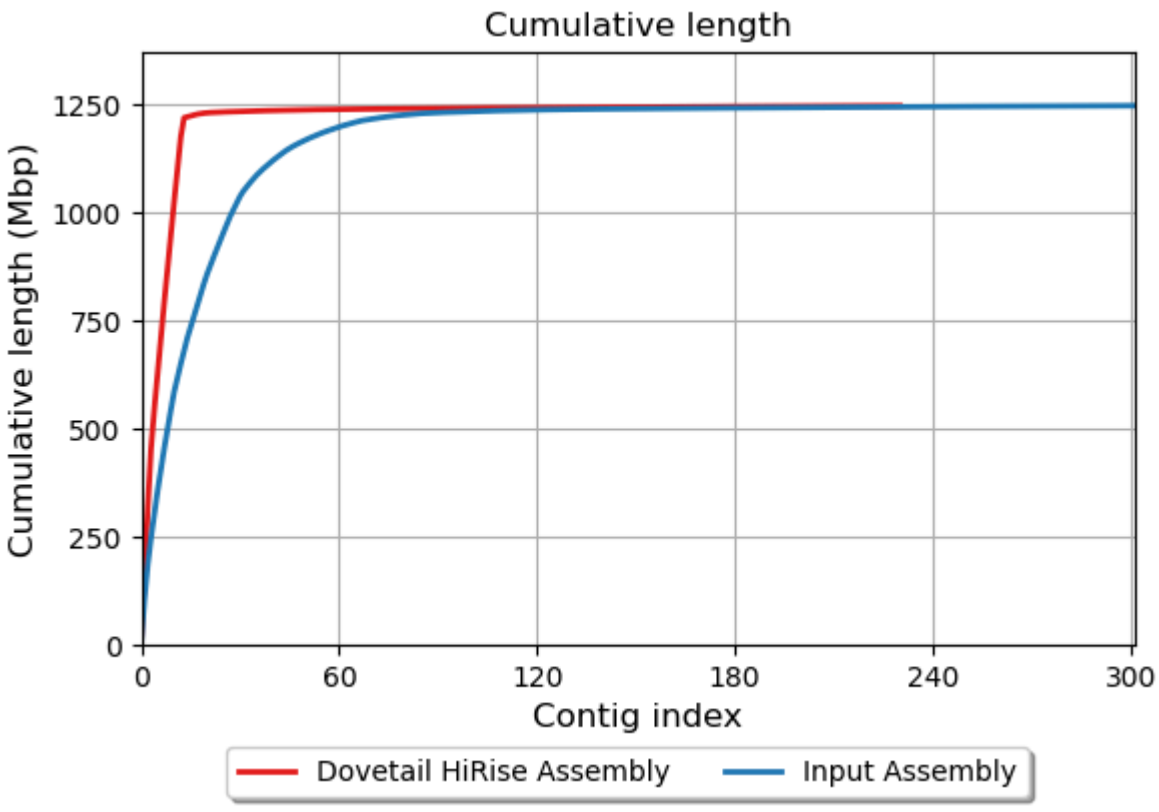
Centre d'Ecologie Fonctionelle & Evolutive

Contents



Overview

Assembly	Total Length (bp)	N50	L50	N90	L90
Input Assembly	1,248,337,966	31,292,631	12	6,111,372	41
Dovetail HiRise Assembly	1,248,345,166	83,123,753	5	74,320,459	12



Contiguity Metrics

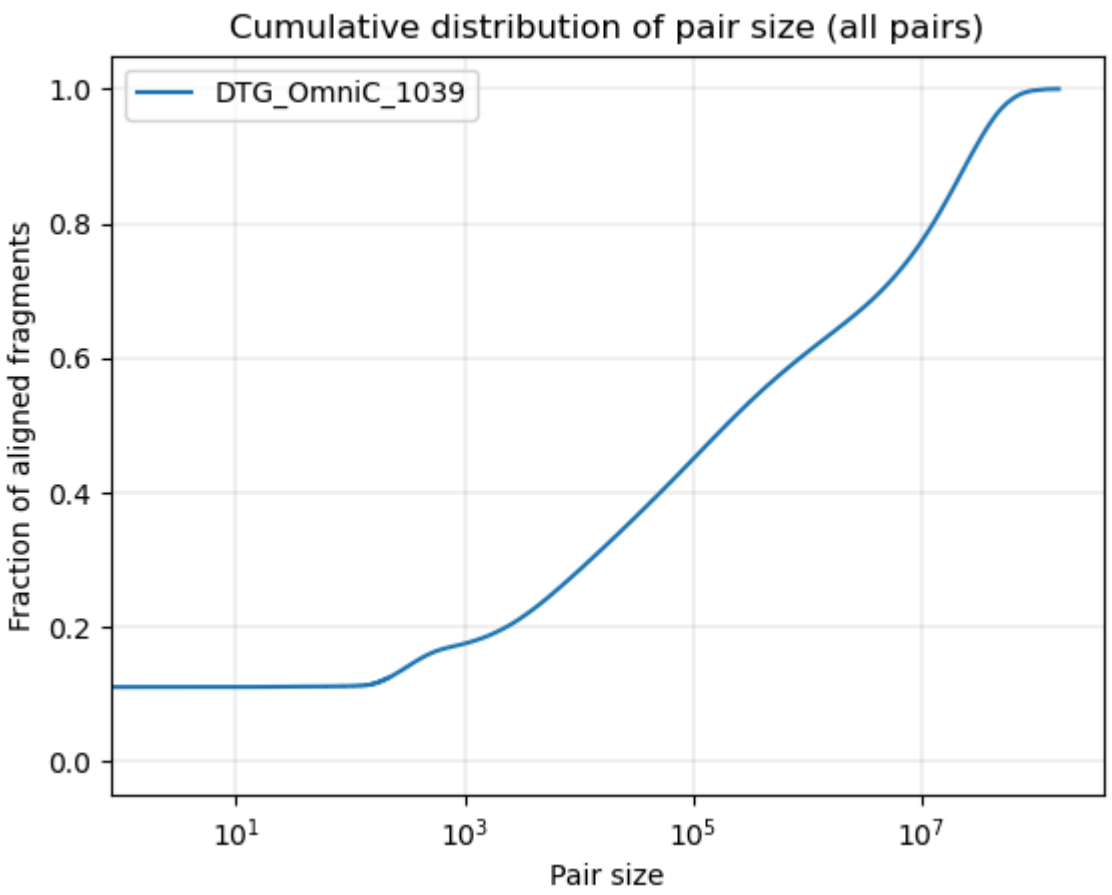
	Input Assembly	Dovetail HiRise Assembly
Largest scaffold	104,484,435	159,663,372
Number of scaffolds	301	230
Number of scaffolds > 1kbp	301	230
Number of gaps	0	68
Number of N's per 100 kbp	0.00	0.58

BUSCO

Assembly	Complete BUSCOs (C)	Complete and single-copy BUSCOs (S)	Complete and duplicated BUSCOs (D)	Fragmented BUSCOs (F)	Missing BUSCOs (M)	Total BUSCO groups searched
Input Assembly	251 (98.43%)	248 (97.25%)	3	3	1	255
Dovetail HiRise Assembly	251 (98.43%)	248 (97.25%)	3	3	1	255

- BUSCO version is: 5.6.1
- The lineage dataset is: eukaryota_odb10 (Creation date: 2024-01-08, number of genomes: 70, number of BUSCOs: 255)

Pair Size Distribution



HiRise Scaffolding Information

Number of joins made by HiRise	68
Number of breaks made to input assembly by HiRise	1
Read-pairs	107,622,264

Materials and Methods

Dovetail Omni-C Library Preparation and Sequencing

For each Dovetail Omni-C library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNAse I, chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter containing ends. After proximity ligation, crosslinks were reversed and the DNA purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce a approximately 30x sequence coverage. Then HiRise used MQ>50 reads for scaffolding (see "read-pair" above for figures).

Scaffolding the Assembly with HiRise

The input de novo assembly and Dovetail OmniC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al, 2016). Dovetail OmniC library sequences were aligned to the draft input assembly using bwa (https://github.com/lh3/bwa). The separations of Dovetail OmniC read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

References

1. Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, Haussler D, Rokhsar DS, Green RE. Genome Research. 2016 Mar;26(3):342-50.