Building High Quality, Chromosome-Scale, De Novo Genome Assemblies by Scaffolding Next-Generation Sequencing Assemblies with Bionano Genome Maps

Jian Wang, Andy Wing Chun Pang, Ernest T Lam, Warren Andrews, Thomas Anantharaman, Alex Hastie, Henry B. Sadowski, Michael Saghbini, Luna Zhang, Zhanyang Zhu, Mike Austin, Han Cao, and Mark Borodkin

Bionano Genomics, San Diego, California, USA

Abstract

With the exception of a few model organisms, many biologically and economically important plants and animals still lack a reference-quality genome assembly that is crucial to the understanding of their biology. Their genomes are often complex and highly repetitive, making generation of high-quality assemblies almost impossible with next-generation sequencing (NGS) alone and without access to long-range structural information. Bionano genome mapping provides a solution to reconstruct the full genomic architecture of large and complex genomes.

Here, we present a novel direct enzymatic labeling approach which maintains the integrity of the DNA and allows us to create very contiguous Bionano maps which can then be used to scaffold NGS sequence assemblies to produce highly contiguous and structurally accurate hybrid assemblies that can span most repeat regions. This direct labeling method is compatible with a vast array of organisms.

We validated our approach with the human NA12878 genome. Starting with NGS assemblies with N50 ranging from 0.18 – 0.9 Mbp, we produced hybrid assemblies with N50 from 70 to 80 Mbp. Chromosome-arm length scaffolds were assembled in 20 out of 23 chromosomes, and alignments show that they were consistent with the hg19 reference. The hybrid assemblies incorporated 80-90% of total NGS sequences with over 99% scaffolding accuracy. We will also show equally impressive scaffolds for a variety of plants and animals. For a low cost and only several days from sample-to-scaffold, this new method promises to set a new standard for genome finishing.

Methods

1. Extraction of long DNA molecules
2. Label DNA at specific sequence motifs
3. Saphyr Chip linearizes DNA in NanoChannel arrays
4. Saphyr automates imaging of single molecules in NanoChannel arrays
5. Molecules and labels detected in images by instrument software
6. Bionano Access software assembles optical maps

(1) Long molecules of DNA are labeled with Bionano reagents by (2) incorporation of fluorophores at a specific sequence motif throughout the genome. (3) The labeled genomic DNA is then linearized in the Saphyr Chip using NanoChannel arrays (4) Single molecules are imaged by Saphyr and then digitized. (5) Molecules are uniquely identifiable by distinct distribution of sequence motif labels (6) and then assembled by pairwise alignment into de novo genome maps.

DLS (Direct Label and Stain) Labeling Chemistry

- DLE-1 is the first of a new class of Bionano DLS enzymes, no more to come
- Single enzymatic reaction: no mixing, no repair step, no NLS
- No fragile sites: optical maps are >100x higher (human)
- Works on the same principle of tagging recognition motifs, highly specific
- Long molecules: No Fragile Sites – Increased Label Density

Unprecedented Maps

Full chromosome arm of human Chr3 was assembled

Bionano map of 147 Mbp was assembled in Maize B73

Hybrid-scaffolding NGS sequence assembly with Bionano maps:
1. Sequence contigs are converted into hybrid maps and aligned to Bionano maps
2. Assembly errors in sequence assembly were detected and corrected
3. NGS contigs are ordered and oriented into ultra-long super-scaffolds

Key features of Bionano hybrid scaffolds:
- Configuration: scaffold N50 over 20 Mbp, up to 700x improvement over input NGS
- Contaminants: >95% of NGS contigs incorporated in hybrid scaffold
- Cost-effective: final scaffold can be generated in <1 week with <1000 dollars in cost
- Compatible with many species: DLE-1 maps successfully generated for >15 different species

Hybrid scaffold of human NA12878 genome

- NGS corrected by total length incorporated in final scaffold
- ~100% accuracy

Conclusions

Bionano Genomics' mapping solution provide an accurate and direct view of the global architecture of genome sequences. Integrating Bionano mapping data with NGS sequence data present both a global, top-down view along with single-nucleotide level details of the genome. The scaffolds generated with this data have set a new standard for genome assembly that can be accomplished in less than one week for less than 1000 dollars.

Reference

1) Cao, H., et al., Rapid detection of structural variation in a human genome using NanoChannel-based genome mapping technology. Gigascience (2014); 3(1):34