CDB SEMINAR

Rei Kajitani

School and Graduate School of Bioscience and Biotechnology,
Tokyo Institute of Technology

Friday, April 17, 2015

14:00~15:00   Auditorium C1F

Practical application of Platanus genome assembler

Summary

Although it has been nearly a decade since the emergence of "high-throughput DNA sequencers," various protocols for *de novo* genome assembly exist and standardization have not yet been achieved. Currently, throughput of Illumina sequencers and quality of mate-pair (long-jumping) libraries continue to improve, and single-molecule sequencers are applied for *de novo* assembly, resulting in confusing circumstances to select a strategy.

 In this seminar, I introduce the procedure to construct draft genomes, which utilized Illumina data and Platanus genome assembler (Kajitani et al. 2014). As its novel function, Platanus detects specific graph structures in the scaffolding step and merges highly heterozygous regions that include structural variants. In general, wild-type samples can be highly heterozygous, and it is expected to improve genome assembly for non-model organisms. In the original paper, the performance of Platanus was validated using benchmark data consisting of various species (nematode worms, oyster, bird, snake and fish). It was consequently adopted in the published genome projects: coelacanth (Nikaido et al. 2013), midge (Gusev et al. 2014) and three swallowtail butterflies (Cong et al. 2015; Nishikawa et al. 2015). Practically, the methods usually described in "supplementary materials," such as preprocess of reads and parameter setting in command lines, are influential for results, therefore I also focus on those procedures. With mate-pairs of which insert-sizes >10 kbp, our team has succeeded in achieving mega-order N50 lengths for eukaryotic genomes. Review of whole procedures of *de novo* assembly is helpful not only to improve the quality of draft genomes, but also to understand the progressing technologies related with genome projects.

Host:
Shigehiro Kuraku
Phyloinformatics Unit, CLST
shigehiro.kuraku
@riken.jp
Tel:078-306-3048
(ext:4232)

References
1. Cong, Q., Borek, D., Otwinowski, Z. & Grishin, N. Tiger Swallowtail Genome Reveals Mechanisms for Speciation and Caterpillar Chemical Defense. *Cell Rep* (2015).
2. Gusev, O. *et al.* Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat Commun* **5,** 4784 (2014).
3. Kajitani, R. *et al.* Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24,** 1384–1395 (2014).
4. Nikaido, M. *et al.* Coelacanth genomes reveal signatures for evolutionary transition from water to land. *Genome Res.* **23,**1740–8 (2013).
5. Nishikawa, H. *et al.* A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. *Nat. Genet.* 47, 405–9 (2015).

RIKEN CENTER for DEVELOPMENTAL BIOLOGY (CDB)