DOCTORAL THESIS

Resolving Complex Genomic Structures for Studying Hybrid Incompatibilities between Caenorhabditis Nematodes

DING, Qiutao

Date of Award:
2021

Abstract

Systematic characterization of hybrid incompatibility (HI) between related species remains the key to understanding speciation. The genetic basis of HI has been intensively studied in *Drosophila* species, but remains largely unknown in other species, including nematodes. The recent discovery of a *C. briggsae* sister species, *C. nigoni*, has opened up the possibility of dissecting the genetic basis of HI in nematode species. However, the existing nematode genomes are far from satisfaction due to numerous errors or missing sequences. This thesis aims to improve these genomes and to establish methodologies to facilitate the study of speciation genetics using the two species.

Unlike the relatively short reads produced by next-generation sequencing (NGS) platforms, the reads generated with third-generation sequencing (TGS) platforms, including nanopore sequencing from Oxford Nanopore Technologies (ONT) and single molecule real-time (SMRT) sequencing from Pacific Biosciences (PacBio), are usually long, which are up to 2 million bps. These so-called long reads are expected to facilitate numerous applications, including *de novo* genome assembly with improved continuity and genome finishing by closing gaps usually left by repetitive sequences. Even in the "finished" *C. elegans* genome, there are still many gaps caused by repetitive sequences, for example, those by ribosomal DNA (rDNA) clusters.

In this thesis, I first used TGS (interchangeable with long-read sequencing) methods to close the gaps left by the tandemly repetitive rDNAs in both *C. elegans* and *C. briggsae*. I found that the rDNA units have undergone a wide range of variations between and intra-strains, including insertions and deletions (INDELs) and single nucleotide polymorphisms (SNPs), which allow unambiguous placement of individual units within the rDNA cluster. Intriguingly, the arrangement and composition of individual rDNA units can serve as a reliable marker for inference of strain origin not only between strains but also between populations within a single strain. I also demonstrated that the genomic environment of the rDNA cluster is compatible with RNA polymerase II transcription. Next, I developed a pipeline for physical mapping of genomic fosmids using ONT sequencing, which was used to complement the HI loci resulted from *C. nigoni* genome. Lastly, I improved a *C. briggsae* transgenesis method with a single copy of insertion, with which I successfully generated 54 transgenic lines expressing dominant markers. These markers were intensively used to identify HI loci between the two closely related species. Overall, this thesis provides an invaluable resource for the study of speciation genetics and illustrates the power of ONT DNA sequencing in genome finishing

Table of Contents

List of Figures

List of Tables