

## DOCTORAL THESIS

### Resolving Complex Genomic Structures for Studying Hybrid Incompatibilities between *Caenorhabditis* Nematodes

DING, Qiutao

*Date of Award:*  
2021

[Link to publication](#)

#### General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

## 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

Declaration.....	i
Abstract.....	ii
Acknowledgements.....	iv
Table of Contents.....	v
List of Figures.....	vii
List of Tables.....	ix
List of Abbreviations.....	x
Chapter 1 Literature Review.....	1
1.1 Long-read sequencing in genome finishing.....	1
1.2 Status of nematode genome assembly.....	5
1.3 <i>Caenorhabditis</i> species as a model for HI study.....	7
1.3.1 Intraspecific HI in <i>C. elegans</i> .....	9
1.3.2 Interspecies HI in <i>Elegans</i> supergroup.....	13
1.4 Objectives.....	16
Chapter 2 Long read sequencing reveals a dynamic <i>C. elegans</i> 5S rDNA cluster.....	17
2.1 Abstract.....	17
2.2 Introduction.....	18
2.3 Materials and Methods.....	23
2.3.1 Library preparation and ONT sequencing.....	23
2.3.2 Sequence acquisition and alignment.....	24
2.3.3 Identification of variation in 5S rDNA units.....	26
2.3.4 Reconstruction of rDNA clusters.....	28
2.3.5 Draft genome assembly and quality assessment.....	29
2.3.6 Estimation of rDNA copy number.....	30
2.3.7 Validation of genomic localization and structure of assembled rDNA clusters.....	31
2.3.8 Molecular biology, transgenesis, and imaging.....	32
2.3.9 Data access.....	33
2.4 Results.....	34
2.4.1 Genomic architecture of the 5S rDNA cluster.....	34
2.4.2 Variations in genomic organization of the 5S rDNA cluster.....	39
2.4.3 rDNA in eccDNA and copy number variations in chromosomal rDNA cluster.....	46
2.4.4 45S rDNA cluster in genome.....	50
2.4.5 Genomic environment of the rDNA cluster is partially compatible with RNA Pol II transcription.....	54

2.5 Discussion.....	56
2.6 Supplementary information .....	58
Chapter 3 Physical mapping of <i>C. nigoni</i> fosmid library using nanopore sequencing	76
3.1 Abstract.....	76
3.2 Introduction.....	76
3.3 Materials and Methods.....	78
3.3.1 Construction of <i>C. nigoni</i> fosmid sequencing library.....	78
3.3.2 Clustering and mapping of fosmid reads .....	81
3.4 Results.....	83
3.4.1 ONT sequencing of <i>C. nigoni</i> fosmid library .....	83
3.4.2 Physical clone mapping of the <i>C. nigoni</i> fosmid library .....	84
3.5 Discussion.....	87
Chapter 4 An improved transgenesis method in <i>Caenorhabditis briggsae</i> for hybrid incompatibility loci interaction screening.....	88
4.1 Abstract.....	88
4.2 Introduction.....	89
4.3 Materials and Methods.....	92
4.3.1 Nematode strains.....	92
4.3.2 Introgression .....	93
4.3.3 Molecular Biology and transgenesis.....	94
4.3.4 Quantification of P0 insertion frequency.....	97
4.3.5 Selection marker for transformation .....	97
4.3.6 Development of negative selection marker .....	98
4.4 Results.....	99
4.4.1 Heat shock treatment significantly increased the efficiency of transgene insertion .....	99
4.4.2 A collection of single copy insertions animals expressing RFP ...	101
4.4.3 A dual-color labeling system for interacting HI loci screening ....	102
4.4.4 The lethal PEEL-1 is nonlethal in <i>C. briggsae</i> .....	103
4.4.5 Limited success of using <i>sup-35</i> as a negative selection marker for extrachromosomal array .....	108
4.5 Discussion.....	109
4.5.1 Potential cause of low efficiency transgenesis.....	110
4.5.2 Divergent response to PEEL-1 and SUP-35 in <i>C. briggsae</i> .....	110
4.6 Supplementary information .....	112
Reference .....	115
CURRICULUM VITAE .....	125

## List of Figures

Figures	Titles	Page no.
Fig. 1.1	Paternal effect by zygotic lethality from sperm-derived PEEL-1	10
Fig. 1.2	A maternal-effect HI gene in N2	12
Fig. 1.3	Asymmetric HI phenotypes in F1 and backcrossing progeny between <i>C. briggsae</i> and <i>C. nigoni</i>	14
Fig. 2.1	Structure of the <i>C. elegans</i> N2 5S rDNA cluster	37
Fig. 2.2	Structural variations within the 5S rDNA cluster between <i>C. elegans</i> N2 strain and other N2-derived strains	43
Fig. 2.3	Structural variations within the 5S rDNA clusters between <i>C. elegans</i> N2 and CB4856 strains	45
Fig. 2.4	Compositions of rDNA cluster in eccDNA and copy number variations in the rDNA cluster	49
Fig. 2.5	Comparison of 45S rDNA units and cluster between strains and species	53
Supplementary Fig. 2.1	Coverage changes in rDNA clusters	58
Supplementary Fig. 2.2	The average read length in the rDNA clusters was smaller than those in other genomic loci	59
Supplementary Fig. 2.3	Presence of <i>cel</i> -5S unit 3 in the 330 <i>C. elegans</i> wild isolates	60
Supplementary Fig. 2.4	Evaluation of the ONT reads assembled <i>C. briggsae</i> AF16 genome	61
Supplementary Fig. 2.5	Characterization of the 5S rDNA units in <i>C. briggsae</i> AF16	62
Supplementary Fig. 2.6	Association of rDNA clusters with chromosomes is revealed with Hi-C data	63
Supplementary Fig. 2.7	Extrachromosomal rDNAs were detected in endogenous circular DNA sequencing data	64
Supplementary Fig. 2.8	Large INDELS in the 45S rDNA cluster of <i>C. elegans</i> and <i>C. briggsae</i>	65
Supplementary Fig. 2.9	Terminal duplication model of chrIL, chrIR, and chrIVL ends in <i>C. elegans</i> CB4856 ancestor	66

Supplementary Fig. 2.10	Altered expression patterns of transgenes in rDNA clusters	67
Supplementary Fig. 2.11	A tiling path of ONT reads for the <i>C. elegans</i> N2 5S rDNA cluster consensus is reconstructed for illustration	68
Fig. 3.1	Workflow for fosmid clone localization sequencing	80
Fig. 3.2	Fosmid clustering	82
Fig. 3.3	Distribution of the length of genomic inserts in the fosmid collection	85
Fig. 4.1	A modified <i>miniMos</i> -based transgene insertion method in <i>C. briggsae</i>	96
Fig. 4.2	Improved <i>MosI</i> -dependent insertion frequency	100
Fig. 4.3	Genomic map of fluorescence marker insertions over <i>C. briggsae</i> genome (CB4)	102
Fig. 4.4	Demonstration of dual-color labeling system in identified HI interacting introgressions for phenotype rescue	106
Fig. 4.5	Comparison of negative selection efficiency of PEEL-1 and SUP-35	107
Supplementary Fig. 4.1	Overview of the plasmid construct	112
Supplementary Fig. 4.2	Identification of <i>hsp-16.41</i> promoter equivalent in <i>C. briggsae</i>	113

## List of Tables

<b>Tables</b>	<b>Titles</b>	<b>Page no.</b>
Table 2.1	Statistics of ONT reads used for characterization of 5S rDNA cluster	36
Table 2.2	Variants of 5S rDNA unit sequences in <i>C. elegans</i> N2	39
Supplementary Table 2.1	Identified 5S rDNA variants in <i>C. elegans</i> CB4856	69
Supplementary Table 2.2	Estimated copy numbers of 5S rDNA unit and 45S rDNA unit using the TGS data	69
Supplementary Table 2.3	Proportions of <i>cel-5S</i> unit 3 in 330 <i>C. elegans</i> isolates	70
Supplementary Table 2.4	Proportions of 5S rDNA variants in genomic DNA and ERC	73
Supplementary Table 2.5	Proportions of 45S rDNA variants in <i>C. elegans</i> (N2) genomic DNA and ERC	73
Supplementary Table 2.6	Proportions of variants in 45S rDNA in <i>C. elegans</i> (CB4856)	74
Supplementary Table 2.7	Correlation coefficient of the estimated copy numbers of rRNA genes	74
Supplementary Table 2.8	PCR primers used for target amplification and <i>miniMos</i> vector construction	74
Supplementary Table 2.9	Estimated poly(A) length of all rRNA reads using DRS data	75
Table 3.1	Sample table of fosmid-based physical mapping based on genome coordinates	86
Supplementary Table 4.1	Transgene insertion site	114
Supplementary Table 4.2	PCR primers for vector amplification	114