

Article

T2T *Colletotrichum lini* Genomes with Hifiasm: ONT R9 and R10 Read Processing and Assembly Guidelines for Fungi

Elizaveta A. Ivankina ^{1,*} , Ekaterina M. Dvorianinova ¹ , Alexander A. Arkhipov ^{1,2} , Antoniy M. Kaplun ^{1,2} , Tatiana A. Rozhmina ³, Ludmila P. Kudryavtseva ³, Nikolai M. Barsukov ¹ , Olesya D. Moskalenko ^{1,2}, Fedor D. Kostromskoy ^{1,4} , Kirill A. Klimov ^{1,2}, Andrei A. Artamonov ^{1,5}, Elena V. Borkhert ¹ , Daiana A. Krupskaya ^{1,5} , Elena N. Pushkova ¹, Nataliya V. Melnikova ¹  and Alexey A. Dmitriev ^{1,*} 

- ¹ Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia; eugenevarenjuk@gmail.com (E.M.D.); arhipov.aleksandr2.0@gmail.com (A.A.A.); atonkaplun@yandex.ru (A.M.K.); keeper@yandex.ru (N.M.B.); lesyamosk.bye-bye2003@yandex.ru (O.D.M.); fkostr@gmail.com (F.D.K.); kiry.klimov.231@gmail.com (K.A.K.); artamonov.contact@gmail.com (A.A.A.); sashai@inbox.ru (E.V.B.); zhernova.d@ya.ru (D.A.K.); pushkova18@gmail.com (E.N.P.); mnv-4529264@yandex.ru (N.V.M.)
- ² Moscow Center for Advanced Studies, 123592 Moscow, Russia
- ³ Federal Research Center for Bast Fiber Crops, 172002 Torzhok, Russia; tatyana_rozhmina@mail.ru (T.A.R.); lpkudryavtseva@icloud.com (L.P.K.)
- ⁴ I.M. Sechenov First Moscow State Medical University, 119991 Moscow, Russia
- ⁵ Faculty of Biology, Lomonosov Moscow State University, 119234 Moscow, Russia
- * Correspondence: sigova1567@gmail.com (E.A.I.); alex_245@mail.ru (A.A.D.)

Abstract

The assembly of telomere-to-telomere (T2T) genomes is essential for understanding genomic architecture, especially in fungal pathogens with complex karyotypes, such as *Colletotrichum lini*, causing flax anthracnose disease. This study provides optimized guidelines for the T2T genome assembly using Oxford Nanopore Technologies (ONT) R9.4.1 and R10.4.1 sequencing data processed with the Hifiasm 0.25.0 assembler (with --ont module). We analyzed ONT sequencing data for four *C. lini* strains and compared basecalling tools (Guppy and Dorado), read filtration strategies (quality thresholds Q10/Q15 and length cut-offs 5 kb/10 kb), and genome coverage levels from 5× to 160×. Our results demonstrated that Dorado-basecalled reads consistently had higher average quality, especially the R10.4.1 data, leading to improved telomere resolution and complete mitochondrial genome assembly. Moderate genome coverage (40–65×) combined with Q15 quality and 5 kb length filtration for R10.4.1 data, or Q10 and 5 kb for R9.4.1 data, produced the most contiguous and complete assemblies. Overfiltration of reads by length and quality or conversely excessive coverage (>90×) reduced assembly quality, causing fragmentation or erroneous chromosome merging. With optimized parameters of ONT R9.4.1 and R10.4.1 sequencing data preprocessing, Hifiasm efficiently generated T2T and near-T2T assemblies of *C. lini* genomes: 53.7–56.1 Mb length, 13–30 contigs, 12–13 chromosomes (including 3–12 T2T chromosomes), complete mitochondrial genome, and >98.5% BUSCO completeness. These findings provide a solid framework for ONT-based fungal genome assembly, facilitating future research on genomic variation and pathogenicity in *Colletotrichum* and related genera.

Keywords: *Colletotrichum lini*; anthracnose; flax pathogen; Nanopore; R9.4.1 and R10.4.1 flow cells; T2T genome assembly; basecalling tools; read processing



Academic Editors: Houxiang Kang, Wei Li and Zhigang Li

Received: 24 October 2025

Revised: 27 November 2025

Accepted: 12 December 2025

Published: 7 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Colletotrichum lini Manns et Bolley is a fungal pathogen that causes anthracnose in flax [1–3]. This disease leads to significant crop losses and is a major concern for flax farmers. Flax is used to produce oil and fiber. Flax fiber is utilized in the textile industry and composite production [4,5]. Linseed oil is rich in lignans and omega-3 fatty acids [6]. It is used as a dietary supplement and in the production of paints and varnishes [7–10]. Therefore, flax plays an important role in agriculture. Anthracnose is one of the most actively spreading diseases affecting flax and leads to substantial economic losses [11,12].

The study of the genomics of fungal pathogens can lead to more effective disease management. Obtaining a complete genome sequence of a phytopathogenic organism is essential for understanding the organization and diversity of its genetic material [13]. Many phytopathogenic species within the genus *Colletotrichum* contain accessory chromosomes, the presence of which may vary among different strains [14–16]. To investigate this phenomenon, telomere-to-telomere (T2T) genome sequences of phytopathogens are required [17]. Such genomes provide opportunities to identify conserved and variable regions of the pathogen genome, as well as to determine the chromosomal composition [18]. Moreover, whole genomes present new possibilities for taxonomy studies [19,20].

The third-generation sequencing on the Oxford Nanopore Technologies (ONT) platform enables the acquisition of long reads. This technology has undergone an upgrade to a new chemistry, which has improved sequencing accuracy and homopolymer resolution [21,22]. The increased accuracy allows for the assembly of high-quality genomes [23].

The Hifiasm tool is a widely used genome assembler, which was created for genome assembling from PacBio (Pacific Biosciences) reads. Later, the option of using ONT data as ultra-long reads (--ul) was added. After that, Hifiasm began to be referred to as the T2T assembler [24–28]. The latest version of Hifiasm (0.25.0) includes its own built-in ONT data corrector (--ont) and allows for the creation of continuous assemblies from ONT simplex data. Despite the fact that Hifiasm does not officially support R9.4.1 cells (it is not optimized for them), it still gives a good result using both R9.4.1 and R10.4.1 sequencing data. Previously, this process required the correction of reads using the Dorado correct module (HERRO), but the tool was officially released only for data generated from R10.4.1 flow cells [23,29]. The new version of Hifiasm (0.25.0) creates assemblies without the need for preliminary read correction.

In this research, we sequenced the genome of highly virulent *C. lini* strain #774 on the ONT platform and R10.4.1 flow cell and compared the obtained data with those for three *C. lini* strains' genomes previously sequenced on the R9.4.1 flow cell. We conducted a comprehensive comparative analysis of read processing using two different basecalling software and a variety of read filtration combinations of read quality and minimum length. This study aimed to provide optimized guidelines for ONT R9/R10 read processing and T2T fungal genome assembling.

2. Materials and Methods

2.1. Fungal Material

C. lini strain #774 mycelium was grown in a tube with potato dextrose agar, 39 g/L (Condalab, Madrid, Spain).

2.2. DNA Extraction and Purification

For DNA extraction, we used our previously developed protocol [30,31] with some modifications [32]. We evaluated the quality and quantity of the extracted DNA with spectrophotometry (NanoDrop 2000C; Thermo Fisher Scientific, Waltham, MA, USA), fluorometry (Qubit 4.0; Thermo Fisher Scientific, Waltham, MA, USA), and agarose gel electrophore-

sis (1% agarose) techniques. The obtained DNA with A260/280 ~ 1.9, A260/230 ~ 2.2, and a concentration of ~400 ng/μL was used for the Oxford Nanopore Technologies (ONT) and Illumina library preparation.

2.3. DNA Library Preparation and Sequencing on the Oxford Nanopore Technologies and Illumina Platforms

To prepare the DNA library for sequencing on the ONT platform, the SQK-LSK114 Ligation Sequencing Kit (ONT, Oxford, UK) was used for *C. lini* strain #774. Sequencing was performed on a PromethION instrument with an R10.4.1 flow cell (ONT, Oxford, UK).

The Illumina library was prepared with the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol. Library quality and concentration were assessed using the Qsep1-Plus capillary electrophoresis system (Bi-Optic, New Taipei City, Taiwan) and Qubit 4.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), respectively. Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA, USA) instrument (150 + 150 b).

2.4. Genome Assembly and Quality Analyses

The obtained ONT reads of *C. lini* strain #774 (SRX30895616) and reads of previously sequenced *C. lini* strains #390-1, #757, and #771 (SRX21992304, SRX21992305, SRX21992306) [33] were basecalled using Guppy 6.5.7 (<https://nanoporetech.com/software/other/guppy>, accessed on 12 September 2025) with quality filtration threshold min_qscore = 10 and config files dna_r9.4.1_450bps_sup.cfg (*C. lini* strains #390-1, #757, and #771) and dna_r10.4_e8.1_sup.cfg (*C. lini* strain #774) and using Dorado 0.9.6 (<https://github.com/nanoporetech/dorado>, accessed on 12 September 2025) with config file dna_r9.4.1_e8_sup@v3.6 (*C. lini* strains #390-1, #757, #771) and Dorado 1.0.2 with config file dna_r10.4.1_e8.2_400bps_sup@v5.2.0 (*C. lini* strain #774) with quality filtration threshold min_qscore = 10. To remove adapters, Porechop 0.2.4 was used (<https://github.com/rrwick/Porechop>, accessed on 12 September 2025). Quality and quantity of the obtained reads were analyzed with SeqKit v2.4.0 [34]. If needed, reads were corrected using the HERRO algorithm [35] with Dorado 0.9.6 (correct script). The obtained Illumina reads of *C. lini* strain #774 were processed using Cutadapt 2.8 (-a AGATCGGAAGAG -A AGATCGGAAGAG) [36] and Trimmomatic 0.39 (PE, SLIDINGWINDOW:3:28, MINLEN:50) [37]. Illumina reads of the previously sequenced *C. lini* strains #390-1, #757, and #771 (SRX21992307, SRX21992308, SRX21992309) were trimmed according to the same scheme.

Assemblies were produced by Hifiasm 0.25.0-r726 with the --ont module for uncorrected reads and without any special modules for corrected reads [38]. The assemblies were polished with Pilon 1.24 [39] with Illumina reads. The prior alignments before polishing were produced with BWA 0.7.18-r1243 [40].

To analyze the quality of the obtained assemblies, completeness and contiguity statistics were calculated using BUSCO 5.8.0 (glomerellales_odb10) and QUAST 5.0.2 [41,42]. The following reference genome was used for QUAST reference-based statistics: *Colletotrichum higginsianum* IMI 349063 (NCBI Genome, GCA_001672515.1). Tidk 0.2.31 was used for the identification and visualization of telomeric repeats (<https://github.com/tolkit/telomeric-identifier>, accessed on 12 September 2025). The genome assemblies were aligned to each other using LAST 1471 (<https://gitlab.com/mcfrith/last>, accessed on 12 September 2025).

3. Results

3.1. Basecalling of ONT Reads

We chose four *C. lini* strains for this study: two highly virulent strains (#390-1, #774), one moderately virulent strain (#757), and one lowly virulent strain (#771). The genomes of *C. lini* strains #390-1, #757, and #771 were previously sequenced by us on the ONT platform with the R9.4.1 flow cell [33]. The highly virulent strain #774 genome was sequenced on the ONT platform with the R10.4.1 flow cell in the present study. To assess the impact of the basecalling software on the quality of reads and assemblies, we performed basecalling using two tools, Dorado and Guppy, with an average read quality filtration threshold of Q10 (min_qscore = 10). The characteristics of the obtained data after the Porechop adapter trimming are in Table 1.

Table 1. Statistics of the basecalled ONT genome sequencing data after adapter trimming for *C. lini* strains #390-1, #757, and #771 (R9.4.1 data) and #774 (R10.4.1 data).

Strain	Basecaller	Data Volume, Gb	Genome Coverage ¹ , ×	Max Read Length, kb	N50, kb	Percentage of Data with Q > 20, %	Average Read Q	GC, %
390-1	Dorado 0.9.6	5.7	105	104.8	11.5	60.1	13.3	53.92
	Guppy 6.5.7	5.4	100	377.4	11.6	59.2	13.1	54.09
757	Dorado 0.9.6	8.7	160	147.5	5.8	59.9	13.3	53.77
	Guppy 6.5.7	8.2	150	115.7	5.9	59.1	13.0	53.95
771	Dorado 0.9.6	6.3	115	116.2	1.2	60.1	13.4	53.35
	Guppy 6.5.7	6.0	110	219.7	1.3	59.6	13.2	53.56
774	Dorado 1.0.2	4.6	85	152.2	12.8	91.8	19.2	53.18
	Guppy 6.5.7	4.6	85	150.5	12.7	82.1	17.1	53.29

¹ The estimated genome length is 55 Mb.

According to these statistics and other research, Dorado and Guppy basecallers have different quality assessment approaches [43]. For the R9.4.1-sequenced data, the volume of the obtained reads for each strain varied by 0.3–0.5 Gb, which is 5–10× additional genome coverage. Maximum read length was also different. For the strain #390-1, the maximum read length varied from 104.8 kb (Dorado) to 377.4 kb (Guppy). We suppose that it was because of the quality assessment differences, so the longest read did not pass the quality threshold during Dorado basecalling. For all strains' data, the Dorado-basecalled reads had a higher average read quality. However, the difference in the average read quality for the R9.4.1-sequenced data was not significant between the basecallers (variation of 0.2–0.3), while for the strain #774 (R10.4.1-sequenced data), the read quality value was 19.2 for Dorado and 17.1 for Guppy. The latest Dorado versions were created only for R10.4.1 sequencing data, so we considered their basecalling to be more accurate.

3.2. Processing of ONT Reads

For the analysis of read N50 and genome coverage effects, we filtered reads by average quality (Q > 10 or Q > 15) and by length (length > 5 kb or length > 10 kb). In addition, we corrected reads obtained from the R10.4.1 flow cell (*C. lini* strain #774) using the built-in Dorado correct module. The statistics of the received reads were analyzed with SeqKit (Figure 1).

Strain	Read type*	Data Volume, Gb	Coverage**	N50, kb	GC(%)	Percent from Raw Data Volume	Data Volume, Gb	Coverage**	N50, kb	GC(%)	Percent from Raw Data Volume
390-1	Dorado 0.9.6						Guppy 6.5.7				
	raw	5.7	105	11.5	53.92	100%	5.4	100	11.6	54.09	100%
	Q10, 5kb	4.1	75	16.0	54.06	72%	3.9	70	16.0	54.22	73%
	Q10, 10kb	3.1	60	19.1	54.08	55%	3.0	55	19.1	54.25	55%
	Q15, 5kb	1.0	20	15.6	53.90	18%	0.7	10	15.5	53.92	13%
	Q15, 10kb	0.8	15	19.0	53.96	13%	0.5	10	18.9	54.00	9%
757	Dorado 0.9.6						Guppy 6.5.7				
	raw	8.7	160	5.8	53.77	100%	8.2	150	5.9	53.95	100%
	Q10, 5kb	4.9	90	9.8	53.93	56%	4.6	85	9.8	54.12	56%
	Q10, 10kb	2.4	45	14.3	53.99	27%	2.3	40	14.3	54.18	27%
	Q15, 5kb	1.2	20	9.6	53.71	14%	0.8	15	9.5	53.70	10%
	Q15, 10kb	0.6	10	14.2	53.85	7%	0.4	10	14.2	53.84	5%
771	Dorado 0.9.6						Guppy 6.5.7				
	raw	6.3	115	1.2	53.35	100%	6.0	110	1.3	53.56	100%
	Q10, 5kb	2.3	40	16.7	53.81	37%	2.2	40	16.7	54.00	37%
	Q10, 10kb	1.8	30	20.2	53.91	28%	1.7	30	20.2	54.09	28%
	Q15, 5kb	0.6	10	16.4	53.53	9%	0.4	10	16.2	53.45	6%
	Q15, 10kb	0.4	10	20.2	53.69	7%	0.3	5	20.1	53.63	5%
774	Dorado 1.0.2						Guppy 6.5.7				
	raw	4.6	85	12.8	53.18	100%	4.6	85	12.7	53.29	100%
	Q10, 5kb	3.6	65	17.0	53.37	78%	3.6	65	16.9	53.50	78%
	Q10, 10kb	2.7	50	21.7	53.53	58%	2.7	50	21.6	53.66	58%
	Q15, 5kb	3.3	60	17.1	53.40	72%	3.1	55	16.8	53.44	67%
	Q15, 10kb	2.5	45	21.7	53.56	54%	2.3	40	21.3	53.61	50%
	Q10 corr, 10kb	2.1	40	20.5	54.27	46%	2.0	40	23.0	54.38	44%
	Q15 corr, 10kb	2.0	40	20.5	54.25	44%	2.0	35	22.0	54.32	43%

Figure 1. Statistics of the raw and filtered by quality ($Q > 10$ or $Q > 15$) and minimum read length (>5 kb or >10 kb) ONT data for *C. lini* strains #390-1, #757, and #771 (R9.4.1 data) and #774 (R10.4.1 data). * Read type: raw—basecalled and adapter trimmed with Porechop, Q10—with quality threshold $Q = 10$, Q15—with quality threshold $Q = 15$, 5 kb—with minimum read length of 5 kb, 10 kb—with minimum read length of 10 kb, corr—corrected with Dorado correct module. ** The estimated genome length is 55 Mb. The quality of the N50 value is indicated by the green (best)–white (worst) color scale. The percent of data volume from raw reads is indicated with gray–blue filling lines.

The data volume was between 0.3 and 8.7 Gb among all strains. The *C. lini* strain #774 genome coverage with R10.4.1 reads varied from $85\times$ to $35\times$. The other strains' genome coverage with R9.4.1 reads varied from $160\times$ to $5\times$. This difference was connected to the higher average read quality of R10.4.1 sequencing.

3.3. Impact of Read Processing on Assembly Quality

To assemble *C. lini* genomes, we used the Hifiasm tool. Using the generated combinations of ONT reads, we produced genome assemblies. The uncorrected reads were used with the `--ont` module, and the assemblies from the corrected *C. lini* strain #774 reads were obtained without any special modules. Hifiasm was developed to produce T2T genome assemblies, so one of the most significant assembly characteristics was the number of T2T-assembled chromosomes. Assembly quality was analyzed with QUAST and BUSCO. The presence of assembled telomeric repeats was verified with Tidk (Figure 2).

Flow-cell	Strain	Basecaller	Read Type	Coverage	Total Length, Mb	T2T Contigs	T Contigs	Number of Contigs	Largest Contig Length, Mb	N50, Mb	L50	GC, %	BUSCO	
													C, %	D, %
R9.4.1	390-1	Dorado 0.9.6	raw	105	54.7	2	8	76	6.6	3.6	6	53.94	98.8	0.2
			Q10, 5kb	75	54.5	3	7	19	6.6	3.5	6	53.97	98.8	0.2
			Q10, 10kb	60	54.6	3	8	23	6.6	3.4	6	53.99	98.8	0.2
			Q15, 5kb	20	54.4	2	8	20	6.6	4.1	5	53.98	98.8	0.2
			Q15, 10kb	15	53.8	0	7	34	3.7	2.4	9	54.02	98.0	0.2
		Guppy 6.5.7	raw	100	54.9	2	6	81	8.1	6.5	4	53.94	98.8	0.2
			Q10, 5kb	70	54.6	3	8	23	9.9	3.5	5	53.97	98.8	0.2
			Q10, 10kb	55	54.6	2	12	21	6.6	3.5	6	53.97	98.8	0.2
			Q15, 5kb	10	53.7	0	6	61	3.5	1.8	12	54.01	97.7	0.2
			Q15, 10kb	10	52.5	0	0	169	1.8	0.5	32	54.05	96.1	0.6
	757	Dorado 0.9.6	raw	160	54.6	0	6	184	9.7	4.1	5	53.94	98.8	0.4
			Q10, 5kb	90	53.7	0	6	27	6.1	3.6	6	54.03	98.8	0.4
			Q10, 10kb	45	53.4	0	6	37	4.8	2.3	8	54.04	98.4	0.3
			Q15, 5kb	20	53.6	0	6	26	7.7	3.1	6	54.02	98.7	0.2
			Q15, 10kb	10	52.7	0	0	161	2.1	0.6	30	54.07	97.6	0.4
		Guppy 6.5.7	raw	150	54.3	1	1	171	7.4	4.3	5	53.96	98.8	0.5
			Q10, 5kb	85	53.6	0	6	23	7.9	3.7	5	54.01	98.8	0.2
			Q10, 10kb	40	53.4	0	9	34	4.9	3.4	7	54.01	98.8	0.2
			Q15, 5kb	15	53.3	0	3	41	3.9	2.8	9	54.00	98.8	0.2
			Q15, 10kb	10	47.5	0	0	447	0.9	0.1	105	54.19	89.8	0.6
	771	Dorado 0.9.6	raw	115	54.7	2	8	176	7.2	4.7	5	53.86	98.6	0.3
			Q10, 5kb	40	54.0	0	11	30	6.7	4.5	5	53.94	98.5	0.1
			Q10, 10kb	30	52.3	0	2	113	2.0	0.8	23	53.93	95.2	0.2
			Q15, 5kb	10	52.0	0	0	156	1.6	0.6	29	53.95	96.0	0.3
			Q15, 10kb	10	47.3	0	0	309	0.9	0.2	67	54.04	87.6	0.7
		Guppy 6.5.7	raw	110	54.6	0	8	159	16.0	4.5	4	53.84	98.6	0.3
			Q10, 5kb	40	53.8	0	11	35	6.2	3.1	7	53.96	98.2	0.2
			Q10, 10kb	30	48.7	0	5	86	2.4	1.1	17	54.11	90.2	0.3
			Q15, 5kb	10	44.7	0	0	453	0.7	0.1	110	54.02	82.8	1.0
			Q15, 10kb	5	33.3	0	0	383	0.4	0.1	112	54.12	63.3	0.5
R10.4.1	774	Dorado 1.0.2	raw	85	56.2	6	4	99	8.6	6.5	4	53.40	98.8	0.2
			Q10, 5kb	65	55.0	6	4	20	6.8	5.8	5	54.00	98.7	0.2
			Q10, 10kb	50	54.9	7	4	20	6.8	5.8	5	53.99	98.8	4.5
			Q15, 5kb	60	55.0	6	5	22	6.8	5.8	5	53.95	98.8	0.2
			Q15, 10kb	45	55.1	7	3	25	6.8	5.8	5	53.98	98.8	4.5
		Guppy 6.5.7	raw	85	55.1	7	3	73	6.8	5.8	5	53.89	98.8	0.2
			Q10, 5kb	65	54.7	5	5	27	6.8	5.8	5	53.99	98.8	0.2
			Q10, 10kb	50	56.1	6	5	13	6.8	6.0	5	54.04	98.8	0.2
			Q15, 5kb	55	54.3	5	6	14	6.8	5.8	5	54.01	98.8	0.2
			Q15, 10kb	40	56.1	5	6	15	6.8	6.0	5	54.03	98.8	0.2
		Dorado 1.0.2	Q10 corr, 10kb	40	54.1	4	6	14	6.8	5.8	5	54.02	98.7	0.2
			Q15 corr, 10kb	40	54.3	5	4	15	6.8	5.8	5	54.03	98.8	0.5
		Guppy 6.5.7	Q10 corr, 10kb	40	53.9	2	4	17	6.8	5.8	5	54.04	98.7	0.3
			Q15 corr, 10kb	40	54.0	1	7	14	6.8	5.8	5	54.03	98.8	0.3

Figure 2. Statistics of genome assemblies produced by Hifiasm for *C. lini* strains #390-1, #757, #771, and #774 from ONT raw data and data with different filtration by quality ($Q > 10$ or $Q > 15$) and minimum read length (>5 kb or >10 kb). T2T contigs—number of telomere-to-telomere assembled contigs. T contigs—number of contigs with telomeric repeats on one end. BUSCO: C—complete, D—duplicated. The quality of the values is indicated by the green (best)–white–red (worst) color scale. The assemblies with a total length under 50 Mb (incomplete assembly) are indicated in red in the “Total Length” and “Read Type” columns. The assemblies with the largest contig length over 6.8 Mb (erroneously merged contigs) are indicated in red in the “Largest Contig Length”, “N50”, “L50”, and “Read Type” columns.

The estimated *C. lini* genome length is ~55 Mb. The *C. lini* genome contains ten core chromosomes (chromosome length > 0.9 Mb) and from two to four accessory chromosomes

(chromosome length < 0.9 Mb) [23,32]. We checked the total length of all assemblies and marked those with atypical lengths (less than 50 Mb) in red in the “Total Length” and “Read Type” columns in Figure 2. The largest contig length was also checked, and the assemblies with the largest contig over 6.8 Mb (estimated length of the largest *C. lini* chromosome) were marked in red in the corresponding column and in the “Read type” column as the assemblies with merged chromosomes. We also indicated such assemblies in red in the “N50” and “L50” columns since the erroneously merged contigs affect these metrics.

There were five assemblies with a total length of less than 50 Mb; three of them were produced from Guppy-basecalled reads of strain #771, one from Dorado-basecalled reads of strain #771, and one from Guppy-basecalled reads of strain #757. All these assemblies were produced from R9.4.1 sequencing data. The majority of the assemblies with erroneously merged contigs (eight of nine) were also produced from R9.4.1 sequencing data.

The length of the genome assemblies obtained from raw reads was 54.3–56.2 Mb, and the largest contig length differed from 6.6 to 16.0 Mb, while the largest *C. lini* chromosome (Chromosome 1) is ~6.8 Mb long [23,32]. Only two raw read-based assemblies had no mis-assembled chromosomes—*C. lini* strain #390-1 genome assembly from Dorado-basecalled raw reads and *C. lini* strain #774 genome assembly from Guppy-basecalled raw reads. The raw read-based assemblies had from 76 to 184 contigs, despite the non-zero number of T and T2T contigs. This indicated that there were a significant number of small contigs in these assemblies. So, the Hifiasm assembler with unfiltered reads had a tendency to merge together two or more chromosomes while increasing the number of small contigs.

The genome assemblies of *C. lini* strain #774 from R10.4.1 sequencing data had no variety in the largest contig length (except for the assembly from raw Dorado-basecalled reads). For this strain, the N50 and BUSCO completeness values also had insignificant differences between the read types. The performance gap between R10.4.1 and R9.4.1 technologies was most apparent in the telomere resolution capability. Assemblies generated using R10.4.1 data had a substantially higher number of T2T contigs (5–7) compared to those based on R9.4.1 data (0–3).

GC content remained consistent across all assemblies (53.93–54.07%), indicating that neither sequencing technology nor read filtering introduced significant systematic bias in base composition. This consistency provides confidence in the biological validity of the assembled sequences.

The majority of the analyzed assemblies had more than 98.6% BUSCO completeness, which is comparable with that of the reference assemblies of the genus *Colletotrichum* from NCBI (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=5455&reference_only=true, accessed on 12 September 2025). There was no correlation between the basecalling tool and the assembly completeness or percentage of missing genes. However, for three strains, the assemblies from Dorado-basecalled data had more assembled telomeres than the assemblies from Guppy-basecalled data.

For further analyses and for choosing the best (the most complete and contiguous) genome assembly for each strain, we took only assemblies from reads, which were not marked with red in Figure 2. In order to obtain the fairest and most unambiguous comparison, we performed the normalization of some characteristics for such assemblies using the following formula:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}, \quad (1)$$

where X is the value to be normalized, X_{min} is the minimum value of this characteristic among all analyzed assemblies, and X_{max} is the maximum value of this characteristic among all analyzed assemblies.

For normalization, we chose the number of contigs with both assembled telomeres (T2T contigs), the number of contigs with assembled telomeric repeats only at one end (T contigs) (with a 0.5 ratio), the number of contigs, the largest contig length, and N50, L50, and BUSCO (complete (C) and duplicated (D)) metrics. The total length and GC content are more qualitative than quantitative since there is no direct correlation between the assembly quality and the value of these metrics. So, we kept them non-normalized. The best assembly for every strain was chosen by the sum of normalized metrics (Figure 3).

Flow-cell	Strain	Basecaller	Read Type	Non normalized			Normalized										
				Coverage	Total Length, Mb	GC, %	T2T Contigs	T Contigs	Number of Contigs	Largest Contig Length, Mb	N50	L50	BUSCO		Total		
													C, %	D, %			
R9.4.1	390-1	Dorado 0.9.6	raw	105	54.7	53.94	0.29	0.67	0.86	0.97	0.59	0.99	1.00	0.98	6.00		
			Q10, 5kb	75	54.5	53.97	0.43	0.58	0.99	0.97	0.58	0.99	1.00	0.98	6.22		
			Q10, 10kb	60	54.6	53.99	0.43	0.67	0.98	0.97	0.57	0.99	1.00	0.98	6.24		
			Q15, 5kb	20	54.4	53.98	0.29	0.67	0.98	0.97	0.68	1.00	1.00	0.98	6.23		
			Q15, 10kb	15	53.8	54.02	0.00	0.58	0.95	0.52	0.40	0.96	0.98	0.98	5.08		
		Guppy 6.5.7	Q10, 10kb	55	54.6	53.97	0.29	1.00	0.98	0.96	0.58	0.99	1.00	0.98	6.28		
			Q15, 5kb	10	53.7	54.01	0.00	0.50	0.89	0.49	0.29	0.93	0.97	0.98	4.80		
			Q15, 10kb	10	52.5	54.05	0.00	0.00	0.65	0.22	0.07	0.75	0.92	0.89	3.49		
			757	Dorado 0.9.6	Q10, 5kb	90	53.7	54.03	0.00	0.50	0.97	0.89	0.59	0.99	1.00	0.93	5.63
					Q10, 10kb	45	53.4	54.04	0.00	0.50	0.95	0.69	0.37	0.97	0.99	0.95	5.18
	Q15, 10kb	10			52.7	54.07	0.00	0.00	0.66	0.27	0.08	0.77	0.97	0.93	3.68		
	Guppy 6.5.7	Q10, 10kb		40	53.4	54.01	0.00	0.75	0.95	0.71	0.57	0.98	1.00	0.98	5.56		
		Q15, 5kb	15	53.3	54.00	0.00	0.25	0.94	0.54	0.47	0.96	1.00	0.98	5.01			
	771	Dorado 0.9.6	Q10, 5kb	40	54.0	53.94	0.00	0.92	0.96	0.98	0.74	1.00	0.99	1.00	6.13		
			Q10, 10kb	30	52.3	53.93	0.00	0.17	0.77	0.26	0.11	0.83	0.90	0.98	3.93		
			Q15, 5kb	10	52.0	53.95	0.00	0.00	0.68	0.19	0.08	0.78	0.92	0.95	3.60		
		Guppy 6.5.7	Q10, 5kb	40	53.8	53.96	0.00	0.92	0.95	0.90	0.50	0.98	0.98	0.98	5.75		
	R10.4.1	774	Dorado 1.0.2	Q10, 5kb	65	55.0	54.00	0.86	0.33	0.98	1.00	0.97	1.00	1.00	0.98	6.95	
				Q10, 10kb	50	54.9	53.99	1.00	0.33	0.98	1.00	0.97	1.00	1.00	0.00	6.12	
Q15, 5kb				60	55.0	53.95	0.86	0.42	0.98	1.00	0.97	1.00	1.00	0.98	6.99		
Q15, 10kb				45	55.1	53.98	1.00	0.25	0.97	0.99	0.97	1.00	1.00	0.00	6.06		
Guppy 6.5.7			raw	85	55.1	53.89	1.00	0.25	0.86	1.00	0.97	1.00	1.00	0.98	6.94		
			Q10, 5kb	65	54.7	53.99	0.71	0.42	0.97	1.00	0.97	1.00	1.00	0.98	6.84		
			Q10, 10kb	50	56.1	54.04	0.86	0.42	1.00	0.99	1.00	1.00	1.00	0.98	7.03		
			Q15, 5kb	55	54.3	54.01	0.71	0.50	1.00	1.00	0.97	1.00	1.00	0.98	6.91		
			Q15, 10kb	40	56.1	54.03	0.71	0.50	1.00	0.99	1.00	1.00	1.00	0.98	6.93		
Dorado 1.0.2			Q10 corr, 10kb	40	54.1	54.02	0.57	0.50	1.00	0.99	0.97	1.00	1.00	0.98	6.75		
			Q15 corr, 10kb	40	54.3	54.03	0.71	0.33	1.00	0.99	0.97	1.00	1.00	0.91	6.74		
Guppy 6.5.7			Q10 corr, 10kb	40	53.9	54.04	0.29	0.33	0.99	0.99	0.97	1.00	1.00	0.95	6.36		
			Q15 corr, 10kb	40	54.0	54.03	0.14	0.58	1.00	0.99	0.97	1.00	1.00	0.95	6.35		

Figure 3. Normalized statistics for the genome assemblies of *C. lini* strains #390-1, #757, #771, and #774 produced by Hifiasm from ONT raw data and data with different filtration by quality ($Q > 10$ or $Q > 15$) and minimum read length (>5 kb or >10 kb). BUSCO: C—complete, D—duplicated. The quality of the values is indicated by the green (best)–white (worst) color scale. The “Total” is the sum of all normalized characteristics. Numbers in bold are the maximum “Total” values for each strain. The corresponding values in the “Read Type” and “Coverage” columns are colored green.

The performed analysis revealed an optimal coverage for achieving high-quality *C. lini* genome assemblies. The moderate coverage ($40\text{--}65\times$) consistently produced the best results. Excessively high coverage ($>90\times$) did not necessarily improved the assembly quality. Coverage over $90\times$ often led to the merging of chromosomes. We concluded that, when a certain threshold of coverage is surpassed, read quality and read length become more significant factors for successful assembly.

For R10.4.1 data specifically, coverage between 40× and 65× combined with Q10 quality filtration and 5–10 kb minimum read length filtration produced the highest-quality assemblies (total scores 6.84–7.03). The comparison of corrected and uncorrected R10.4.1 reads revealed that, in most cases, overall assembly quality was higher for assemblies obtained from uncorrected reads (total scores 6.35–6.75 for corrected data versus 6.06–7.03 for uncorrected data).

Filtration of R9.4.1 data for quality and read length demonstrated a complex relationship with assembly contiguity and completeness (Figure 4). The Q10 quality filtration and 5 kb or 10 kb minimum read length filtration generally improved assemblies' contiguity by reducing the number of contigs (19–27 contigs for Q10-filtered assemblies versus 76–184 contigs for raw read-assemblies). However, excessive filtration with Q15 and a minimum read length of 10 kb was detrimental. The most significant assembly quality deterioration was observed when Q15 and 10 kb minimum read length filtration resulted in significant assembly fragmentation (169–447 contigs) (Figure 4b) and reduction in the N50 value (0.1–0.6 Mb), so T2T contigs completely disappeared (Figure 4c). Moreover, all assemblies from R9.4.1 data with the “Q15, 10 kb” read type had significantly lower BUSCO completeness and higher missing BUSCO values in comparison with other assemblies (Figure 4d). We believe that this was related to insufficient genome coverage with sequencing data. This pattern suggests that, while removing low-quality reads can benefit assembly, overly aggressive filtration eliminates reads essential for spanning repetitive regions and maintaining assembly continuity.

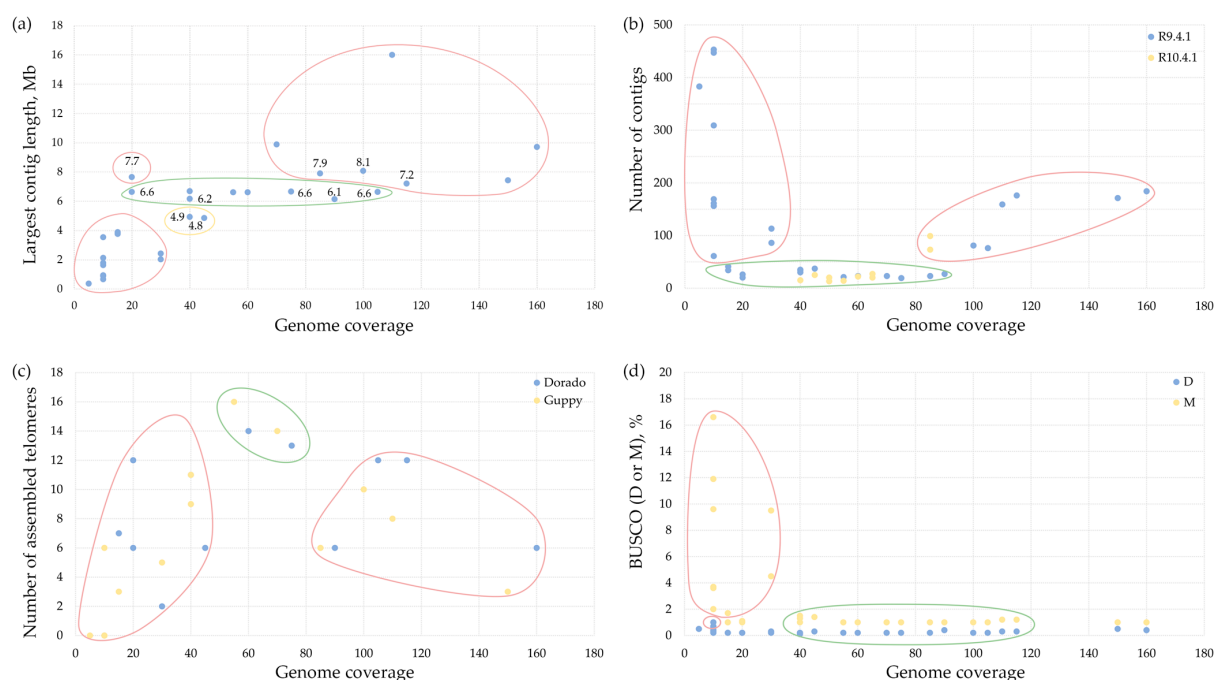


Figure 4. Relationship of genome coverage and (a) the largest contig length, (b) the number of contigs, (c) the number of assembled telomeres, (d) BUSCO (D—duplicated and M—missing) for the Hifiasm-produced genome assemblies from R9.4.1 data for *C. lini* strains #390-1, #757, and #771. The quality of the values is indicated by the green (best)–yellow–red (worst) color scale (see the color of shapes around the dots).

3.4. Comparison of the Obtained Genome Assemblies

We made whole-genome alignments of the best obtained genome assemblies of four *C. lini* strains to the previously assembled complete genome of *C. lini* strain #655-1 (Figure 5) [23]. For R9.4.1-sequenced genomes (strains #390-1, #757, and #771), there were some fragmented chromosomes. *C. lini* strain #390-1 chromosomes 1, 3, 8, 9, 10, 11, and

12 were completely assembled (seven out of thirteen). *C. lini* strain #757 had four (out of twelve) completely assembled chromosomes (9, 10, 11, and 12) and one nearly complete chromosome (7), which might be a strain diversity. *C. lini* strain #771 had three (out of thirteen) completely assembled chromosomes (1, 3, 13) and some nearly complete chromosomes (5, 11, and 12). *C. lini* strain #774, sequenced using the R10.4.1 flow cell, had all twelve complete chromosomes in both assemblies. However, the assembly from Guppy-basecalled reads had repeats on chromosome 8, which were not present in other assemblies of this strain's genome. Therefore, we concluded that these repeats were assembly errors.

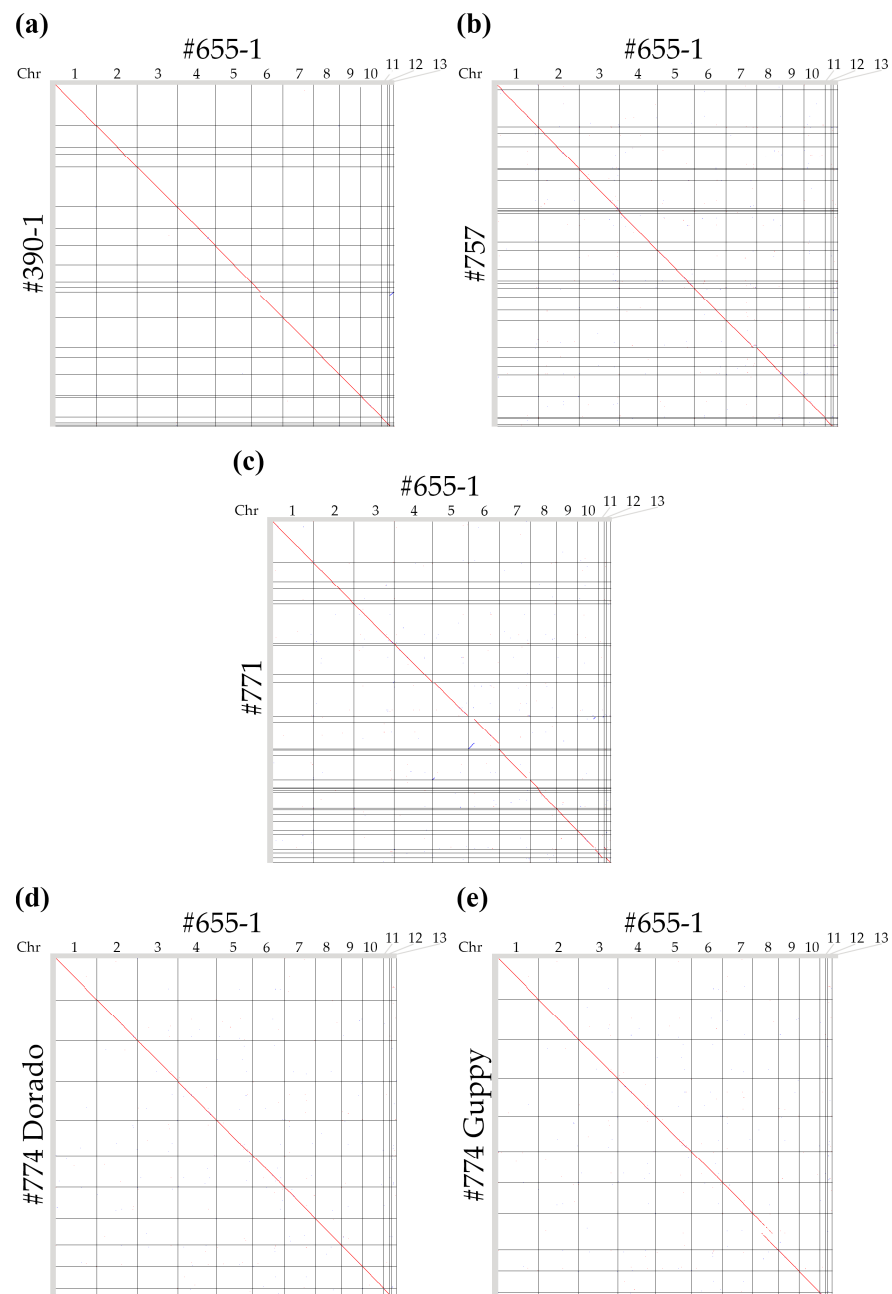


Figure 5. The results of alignment of genome assemblies for (a) *C. lini* strain #390-1 (Dorado-basecalled, “Q10, 10 kb” read type) and *C. lini* strain #655-1; (b) *C. lini* strain #757 (Dorado-basecalled, “Q10, 5 kb” read type) and *C. lini* strain #655-1; (c) *C. lini* strain #771 (Dorado-basecalled, “Q10, 5 kb” read type) and *C. lini* strain #655-1; (d) *C. lini* strain #774 (Dorado-basecalled, “Q15, 5 kb” read type) and *C. lini* strain #655-1; (e) *C. lini* strain #774 (Guppy-basecalled, “Q10, 10 kb” read type) and *C. lini* strain #655-1. Red lines indicate the forward orientation of the aligned sequences, and blue lines indicate the reverse orientation of the aligned sequences.

The quality of the data from R10.4.1 flow cell allowed us to obtain a T2T-level assembly of *C. lini* genome with all complete chromosomes without gaps. Data from R9.4.1 flow cell, which had a lower quality, were less suitable for this purpose, but Hifiasm also produced assemblies of *C. lini* genomes that were close to the T2T level. The comparison of the obtained assemblies with the assemblies of *C. lini* genomes from our previous study [33] revealed that they were very similar. In our previous research, we generated *C. lini* genome assemblies using Canu and then polished them in four iterations with ONT and Illumina reads using various tools. This resulted in the assemblies of 26–32 contigs. These assemblies did not contain a single copy of the mitochondrial genome, and it was necessary to manually search for it and separate it. Fragmented contigs were also present. Considering the time and CPU resources used to assemble the *C. lini* genome, we found that Hifiasm was more efficient than Canu in terms of the time–quality ratio.

According to our previous research [23], the best Illumina polishing tool for Hifiasm-assembled *C. lini* genomes was Pilon. Thus, we performed polishing with Pilon for the obtained assemblies. The assemblies’ coverage with trimmed Illumina reads was 25–35× for each strain (the estimated genome length is 55 Mb). The percentage of mapped Illumina reads for each *C. lini* strain was more than 96% (#390-1—96.3%, #757—98.4%, #771—99.0%, #774—99.9%). We evaluated the polished assemblies’ completeness with BUSCO, and there were no changes compared to the unpolished assemblies. The number of mismatches and indels per 100 kbp relative to the reference genome decreased insignificantly (by ~2% and by ~0.5%, respectively). Thus, the obtained Hifiasm assemblies of *C. lini* genomes did not require polishing with Illumina reads.

As a result, we accepted the genome assemblies of *C. lini* strain #390-1 from Dorado-basecalled reads with quality filtration Q10 and minimum read length of 10 kb, *C. lini* strain #757 from Dorado-basecalled reads with quality filtration Q10 and minimum read length of 5 kb, *C. lini* strain #771 from Dorado-basecalled reads with quality filtration Q10 and minimum read length of 5 kb, and *C. lini* strain #774 from Dorado-basecalled reads with quality filtration Q15 and minimum read length of 5 kb as the final assemblies (Table 2). For the *C. lini* strain #774, we filtered small contigs, as all chromosomes were already fully assembled.

Table 2. Statistics of the final genome assemblies produced by Hifiasm (--ont) from Dorado-basecalled reads for *C. lini* strains #390-1, #757, #771, and #774.

Strain	Read Type	Coverage ¹ , ×	Assembly Length, Mb	Number of Contigs	Number of Chromosomes	Number of Complete Chromosomes ²	N50, Mb	L50	GC, %	BUSCO ³	
										C, %	D, %
390-1	Q10, 10 kb	60	54.6	23	13	7	3.4	6	53.99	98.8	0.2
757	Q10, 5 kb	90	53.7	27	12	4	3.6	6	54.03	98.8	0.4
771	Q10, 5 kb	40	54.0	30	13	3	4.5	5	53.94	98.5	0.1
774	Q15, 5 kb	60	54.4	13	12	12	5.8	5	53.95	98.8	0.2

¹ The estimated genome length is 55 Mb. ² Complete chromosomes—fully assembled chromosomes (might be without telomeric repeats). ³ BUSCO: C—complete, D—duplicated.

4. Discussion

The most commonly used strategy for obtaining T2T genome assemblies currently involves the use of PacBio high-fidelity (HiFi) and ultra-long ONT reads, as well as Hi-C data [44–46]. Popular assemblers for ONT reads include Canu and Hifiasm [18,47,48]. In some instances, researchers used Canu to generate separate assemblies from PacBio HiFi and ONT reads and then combined them using Hifiasm [49]. For an assembly from only

ONT reads, Canu was earlier the primary assembler [31,32,50–53]. For Hifiasm, it was necessary to preprocess ONT reads using the Dorado correct module (HERRO) for single-read error correction (<https://github.com/nanoporetech/dorado>, accessed on 12 September 2025), which was suitable only for the R10.4.1 data [35]. The performance of Hifiasm, when used with the `--ont` module, demonstrated a significant improvement in the assembly of *C. lini* genomes. The opportunity to generate near-T2T assemblies using ONT simplex reads from R9.4.1 and R10.4.1 flow cells, without the need for read correction or Illumina polishing, greatly simplifies the assembly process while maintaining high-quality results. Our comparison with previous Canu-produced *C. lini* genome assemblies [33] showed that Hifiasm achieved similar or better results with substantially reduced computational requirements and without the need for multiple polishing iterations.

The comparison between the Dorado and Guppy basecallers revealed important differences in read quality assessment. While both basecallers produced usable data, Dorado consistently generated reads with higher average quality scores, particularly for R10.4.1 data (19.2 vs. 17.1). Comparative analyses of basecallers are regularly conducted. Several years ago, before the Dorado development, the Guppy basecaller was considered the best one [54]. In modern research, Dorado is recognized as the best ONT basecaller [43]. This difference in read quality proved significant for downstream assembly quality, as assemblies from Dorado-basecalled data consistently showed better telomere resolution. The consistent recovery of complete mitochondrial genomes in assemblies from Dorado-basecalled data is another important finding. The presence of fully assembled mitochondrial genomes in a single copy suggests superior assembly continuity and completeness. This has implications for evolutionary studies and population genetics, as mitochondrial genomes provide valuable phylogenetic information [55–57].

Our analysis of coverage requirements revealed that moderate coverage ($40\text{--}65\times$) consistently produced optimal results, while excessively high coverage ($>90\times$) did not necessarily improve assembly quality and sometimes led to chromosome merging artifacts. A similar study was conducted for Hifiasm when working with PacBio HiFi data, and the researchers obtained a similar range of genome coverage levels, which helped to create the best assemblies in terms of contiguity and completeness [58]. This finding challenges the conventional inference that higher coverage always improves assembly statistics and suggests that beyond a certain threshold, read quality and length become more critical factors than the coverage itself. This has practical implications for sequencing experiment design, potentially reducing the cost and time requirements for generating high-quality fungal genome assemblies.

While our study focused specifically on *C. lini*, the methodologies and insights gained have broader applicability to other fungi with comparable genome lengths. Despite the existing diversity, they share general structural characteristics and present similar challenges during the assembly process. The principles of choosing the optimal coverage, read filtration thresholds, and assembly algorithm can be applied in genome assembly projects for other fungi. This is particularly relevant for species with genome architectures similar to *Colletotrichum*, including other plant pathogens and medically important fungi [59,60].

5. Conclusions

The performed research identified that the ONT R10.4.1 flow cell and Dorado-basecalled data with a moderate genome coverage ($40\text{--}65\times$) and thresholds of a Q15 average read quality and 5 kb minimum read length was the optimal configuration for achieving complete, contiguous, and accurate assemblies of *C. lini* genomes using Hifiasm with the `--ont` module. For the *C. lini* genome assembly from the ONT R9.4.1 data, we could recom-

mend Dorado-basecalled reads with Q10 quality and 5 kb minimum read length thresholds and genome coverage of 40–90×.

These findings have important implications for fungal genome assembly projects, particularly for applications requiring complete T2T assemblies to resolve complex genomic regions.

Author Contributions: Conceptualization, E.A.I., E.M.D., N.V.M., and A.A.D.; performing experiments, E.A.I., E.M.D., A.M.K., T.A.R., L.P.K., N.M.B., O.D.M., K.A.K., A.A.A. (Andrei A. Artamonov), E.V.B., D.A.K., and E.N.P.; data analysis, E.A.I., E.M.D., A.A.A. (Alexander A. Arkhipov), T.A.R., L.P.K., F.D.K., N.V.M., and A.A.D.; writing, E.A.I., E.M.D., N.V.M., and A.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Russian Science Foundation, grant number 22-16-00169-P, <https://rscf.ru/en/project/22-16-00169/>, accessed on 12 September 2025.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data generated in this study can be found in the NCBI database under the BioProject PRJNA929545 (linked to BioSample SAMN13621712).

Acknowledgments: This work was performed using the equipment of the EIMB RAS “Genome” center (http://www.eimb.ru/ru1/ckp/ccu_genome_ce.php, accessed on 12 September 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- da Silva, L.L.; Moreno, H.L.A.; Correia, H.L.N.; Santana, M.F.; de Queiroz, M.V. *Colletotrichum*: Species complexes, lifestyle, and peculiarities of some sources of genetic variability. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 1891–1904. [CrossRef]
- Damm, U.; Baroncelli, R.; Cai, L.; Kubo, Y.; O’Connell, R.; Weir, B.; Yoshino, K.; Cannon, P.F. *Colletotrichum*: Species, ecology and interactions. *IMA Fungus* **2010**, *1*, 161–165. [CrossRef]
- De Silva, D.D.; Crous, P.W.; Ades, P.K.; Hyde, K.D.; Taylor, P.W.J. Life styles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol. Rev.* **2017**, *31*, 155–168. [CrossRef]
- Islam, S.; Hasan, M.B.; Karim, F.-E.; Kodrić, M.; Islam, M.R.; Khatun, M.M.; Motaleb, K.Z.M.A. Thermoset and thermoplastic polymer composites reinforced with flax fiber: Properties and application—A review. *SPE Polym.* **2025**, *6*, e10172. [CrossRef]
- Ma, W.; Yan, L.; Kasal, B. Intensive exploration: The application potential of flax fibre/textiles as reinforcement in cementitious composites. *J. Clean. Prod.* **2025**, *507*, 145457. [CrossRef]
- Gao, Z.; Cao, Q.; Deng, Z. Unveiling the power of flax lignans: From plant biosynthesis to human health benefits. *Nutrients* **2024**, *16*, 3520. [CrossRef]
- Dhiman, A.; Halde, P. Industrial application of linseed oil. In *Dynamics of Linseed and Its Valorization*; Chopra, R., Singh, A., Oberoi, H.S., Eds.; Springer Nature: Singapore, 2025; pp. 183–200.
- Kunutsor, S.K.; Jassal, D.S.; Ravandi, A.; Lehoczk, A. Dietary flaxseed: Cardiometabolic benefits and its role in promoting healthy aging. *GeroScience* **2025**, *47*, 2895–2923. [CrossRef]
- Nie, Y.; Wang, Y.; Hui, J.; Shao, D.; Chen, R.; Deng, Q.; Chen, Y.; Wang, X.; Zhu, D. The impact of flaxseed (*Linum usitatissimum* L.) oil supplementation on human health: A human-centric evidence-graded approach. *Nutrients* **2025**, *17*, 1791. [CrossRef]
- Stepień, A.E.; Trojaniak, J.; Tabarkiewicz, J. Anti-Oxidant and Anti-Cancer Properties of Flaxseed. *Int. J. Mol. Sci.* **2025**, *26*, 1226. [CrossRef] [PubMed]
- Moyse, J.; Lecomte, S.; Marcou, S.; Mongelard, G.; Gutierrez, L.; Höfte, M. Overview and management of the most common eukaryotic diseases of flax (*Linum usitatissimum*). *Plants* **2023**, *12*, 2811. [CrossRef]
- Rashid, K.Y. Principal diseases of flax. In *Flax*; CRC Press: Boca Raton, FL, USA, 2003; pp. 104–135.
- Vineeth, V.K.; Reshma, T.R.; Babu, S.; Philip, S.; Mahadevan, C. Comprehensive whole-genome sequencing reveals genetic characteristics of *Colletotrichum fruticola* (Nara gc5) the causative organism of circular leaf spot disease of rubber (*Hevea brasiliensis*). *J. Plant Pathol.* **2024**, *106*, 579–591. [CrossRef]
- Witte, T.E.; Villeneuve, N.; Boddy, C.N.; Overy, D.P. Accessory chromosome-acquired secondary metabolism in plant pathogenic fungi: The evolution of biotrophs into host-specific pathogens. *Front. Microbiol.* **2021**, *12*, 664276. [CrossRef]

15. Wang, H.; Huang, R.; Ren, J.; Tang, L.; Huang, S.; Chen, X.; Fan, J.; Li, B.; Wang, Q.; Hsiang, T.; et al. The evolution of mini-chromosomes in the fungal genus *Colletotrichum*. *mBio* **2023**, *14*, e00629–00623. [[CrossRef](#)]
16. Gan, P.; Hiroyama, R.; Tsushima, A.; Masuda, S.; Shibata, A.; Ueno, A.; Kumakura, N.; Narusaka, M.; Hoat, T.X.; Narusaka, Y.; et al. Telomeres and a repeat-rich chromosome encode effector gene clusters in plant pathogenic *Colletotrichum* fungi. *Environ. Microbiol.* **2021**, *23*, 6004–6018. [[CrossRef](#)] [[PubMed](#)]
17. Becerra, S.; Baroncelli, R.; Bouffleur, T.R.; Sukno, S.A.; Thon, M.R. Chromosome-level analysis of the *Colletotrichum graminicola* genome reveals the unique characteristics of core and minichromosomes. *Front. Microbiol.* **2023**, *14*, 1129319. [[CrossRef](#)]
18. Lapalu, N.; Simon, A.; Lu, A.; Plaumann, P.-L.; Amselem, J.; Pigné, S.; Auger, A.; Koch, C.; Dallery, J.-F.; O'Connell, R.J. Complete genome of the Medicago anthracnose fungus, *Colletotrichum destructivum*, reveals a mini-chromosome-like region within a core chromosome. *Microb. Genom.* **2024**, *10*, 001283. [[CrossRef](#)]
19. Menicucci, A.; Iacono, S.; Ramos, M.; Fiorenzani, C.; Peres, N.A.; Timmer, L.W.; Prodi, A.; Baroncelli, R. Can whole genome sequencing resolve taxonomic ambiguities in fungi? The case study of *Colletotrichum* associated with ferns. *Front. Fungal Biol.* **2025**, *6*, 1540469. [[CrossRef](#)]
20. Ma, Z.; Liu, F.; Tsui, C.K.M.; Cai, L. Phylogenomics and adaptive evolution of the *Colletotrichum gloeosporioides* species complex. *Commun. Biol.* **2025**, *8*, 593. [[CrossRef](#)] [[PubMed](#)]
21. Mostafa, H.H. An evolution of Nanopore next-generation sequencing technology: Implications for medical microbiology and public health. *J. Clin. Microbiol.* **2024**, *62*, e00246–00224. [[CrossRef](#)]
22. Ratcliff, J.D.; Merritt, B.; Gooden, H.; Siegers, J.Y.; Srikanth, A.; Yann, S.; Kol, S.; Sin, S.; Tok, S.; Karlsson, E.A.; et al. Improved resolution of avian influenza virus using Oxford Nanopore R10 sequencing chemistry. *Microbiol. Spectr.* **2024**, *12*, e0188024. [[CrossRef](#)] [[PubMed](#)]
23. Sigova, E.A.; Dvorianinova, E.M.; Arkhipov, A.A.; Rozhmina, T.A.; Kudryavtseva, L.P.; Kaplun, A.M.; Bodrov, Y.V.; Pavlova, V.A.; Borkhert, E.V.; Zhernova, D.A.; et al. Nanopore data-driven T2T genome assemblies of *Colletotrichum lini* strains. *J. Fungi* **2024**, *10*, 874. [[CrossRef](#)]
24. Burger, N.F.V.; Nicolis, V.F.; Botha, A.-M. Evaluating long-read assemblers to assemble several aphididae genomes. *Brief. Bioinform.* **2025**, *26*, bbaf105. [[CrossRef](#)]
25. Meuser, A.V.; Pitura, A.R.; Mandeville, E.G. A high-quality reference genome for the common creek chub, *Semotilus atromaculatus*. *G3 Genes Genomes Genet.* **2023**, *14*, jkad283. [[CrossRef](#)]
26. Xu, Y.; Wang, C.; Li, Z.; Zheng, X.; Kang, Z.; Lu, P.; Zhang, J.; Cao, P.; Chen, Q.; Liu, X. A chromosome-level haplotype-resolved genome assembly of oriental tobacco budworm (*Helicoverpa assulta*). *Sci. Data* **2024**, *11*, 461. [[CrossRef](#)]
27. Yu, W.; Luo, H.; Yang, J.; Zhang, S.; Jiang, H.; Zhao, X.; Hui, X.; Sun, D.; Li, L.; Wei, X.-q. Comprehensive assessment of 11 de novo HiFi assemblers on complex eukaryotic genomes and metagenomes. *Genome Res.* **2024**, *34*, 326–340. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, J.; Guo, Y.; Su, J.; Yu, G.; Ma, Z.; Qin, C. The first high-quality genome assembly and annotation of *Anthracidaris crassispina*. *Sci. Data* **2024**, *11*, 866. [[CrossRef](#)]
29. Arkhipov, A.A.; Pushkova, E.N.; Bolsheva, N.L.; Rozhmina, T.A.; Borkhert, E.V.; Zhernova, D.A.; Rybakova, T.Y.; Barsukov, N.M.; Moskalenko, O.D.; Sigova, E.A.; et al. Nanopore data-driven chromosome-level assembly of flax genome. *Plants* **2024**, *13*, 3465. [[CrossRef](#)] [[PubMed](#)]
30. Krasnov, G.S.; Pushkova, E.N.; Novakovskiy, R.O.; Kudryavtseva, L.P.; Rozhmina, T.A.; Dvorianinova, E.M.; Povkhova, L.V.; Kudryavtseva, A.V.; Dmitriev, A.A.; Melnikova, N.V. High-quality genome assembly of *Fusarium oxysporum* f. sp. *lini*. *Front. Genet.* **2020**, *11*, 959. [[CrossRef](#)]
31. Sigova, E.A.; Pushkova, E.N.; Rozhmina, T.A.; Kudryavtseva, L.P.; Zhuchenko, A.A.; Novakovskiy, R.O.; Zhernova, D.A.; Povkhova, L.V.; Turba, A.A.; Borkhert, E.V.; et al. Assembling quality genomes of flax fungal pathogens from Oxford Nanopore Technologies data. *J. Fungi* **2023**, *9*, 301. [[CrossRef](#)]
32. Sigova, E.A.; Dvorianinova, E.M.; Rozhmina, T.A.; Kudryavtseva, L.P.; Zhernova, D.A.; Kaplun, A.M.; Pavlova, V.A.; Bodrov, Y.V.; Arkhipov, A.A.; Borkhert, E.V.; et al. Complete annotated genome assembly of flax pathogen *Colletotrichum lini*. *J. Fungi* **2024**, *10*, 605. [[CrossRef](#)]
33. Dvorianinova, E.M.; Sigova, E.A.; Mollaev, T.D.; Rozhmina, T.A.; Kudryavtseva, L.P.; Novakovskiy, R.O.; Turba, A.A.; Zhernova, D.A.; Borkhert, E.V.; Pushkova, E.N.; et al. Comparative genomic analysis of *Colletotrichum lini* strains with different virulence on flax. *J. Fungi* **2024**, *10*, 32. [[CrossRef](#)]
34. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A Cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **2016**, *11*, e0163962. [[CrossRef](#)]
35. Stanojević, D.; Lin, D.; Nurk, S.; Florez de Sessions, P.; Šikić, M. Telomere-to-telomere phased genome assembly using HERRO-corrected simplex Nanopore reads. *bioRxiv* **2024**, 2024.05.18.594796. [[CrossRef](#)]

36. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [[CrossRef](#)]
37. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
38. Cheng, H.; Jarvis, E.D.; Fedrigo, O.; Koepfli, K.-P.; Urban, L.; Gemmell, N.J.; Li, H. Haplotype-resolved assembly of diploid genomes without parental data. *Nat. Biotechnol.* **2022**, *40*, 1332–1335. [[CrossRef](#)]
39. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
40. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997. [[CrossRef](#)]
41. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
42. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)] [[PubMed](#)]
43. Kušmirek, W. Estimated nucleotide reconstruction quality symbols of basecalling tools for Oxford Nanopore Sequencing. *Sensors* **2023**, *23*, 6787. [[CrossRef](#)]
44. Rajesh, M.K.; Budhwar, R.; Shukla, R.; Oraon, P.K.; Goel, S.; Paul, B.; Thomas, R.J.; Dinesh, A.; Jayasekhar, S.; Chandran, K.P.; et al. Chromosome scale genome assembly and annotation of coconut cultivar Chowghat Green Dwarf. *Sci. Rep.* **2024**, *14*, 28778. [[CrossRef](#)]
45. Han, H.; Salinas, N.; Barbey, C.R.; Jang, Y.J.; Fan, Z.; Verma, S.; Whitaker, V.M.; Lee, S. A telomere-to-telomere phased genome of an octoploid strawberry reveals a receptor kinase conferring anthracnose resistance. *GigaScience* **2025**, *14*, g1af005. [[CrossRef](#)]
46. Wang, Z.-Q.; Zhang, S.; Chang, X.-Q.; Yang, X.-L.; You, J.-M.; Zhou, Y.; Luo, C.-X.; Lv, L. Complete genome sequence of *Colletotrichum jinshuiense*, the causal agent of goldthread anthracnose disease. *Microbiol. Resour. Announc.* **2025**, *14*, e0038024. [[CrossRef](#)]
47. Xie, L.; Liu, L.; Luo, Y.; Rao, X.; Di, Y.; Liu, H.; Qian, Z.; Shen, Q.; He, L.; Li, F. Complete genome sequence of biocontrol strain *Bacillus velezensis* YC89 and its biocontrol potential against sugarcane red rot. *Front. Microbiol.* **2023**, *14*, 1180474. [[CrossRef](#)]
48. Nurk, S.; Walenz, B.P.; Rhie, A.; Vollger, M.R.; Logsdon, G.A.; Grothe, R.; Miga, K.H.; Eichler, E.E.; Phillippy, A.M.; Koren, S. HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **2020**, *30*, 1291–1305. [[CrossRef](#)]
49. Li, W.-S.; Xiao, Y.-D.; Liu, J.-Q.; Li, S.-L.; Chen, Y.; Xu, Y.-J.; Yang, X.; Wang, Y.-J.; Li, Z.-Q.; Xia, Q.-Y. The T2T genome of the domesticated silkworm *Bombyx mori*. *Int. J. Mol. Sci.* **2024**, *25*, 12341. [[CrossRef](#)] [[PubMed](#)]
50. Bhadauria, V.; Li, G.; Gao, X.; Laborda, P. Near-complete genome and infection transcriptomes of the maize leaf and sheath spot pathogen *Epicoccum sorghinum*. *Sci. Data* **2025**, *12*, 261. [[CrossRef](#)]
51. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* **2017**, *3*, e000132. [[CrossRef](#)] [[PubMed](#)]
52. Dvorianinova, E.M.; Bolsheva, N.L.; Pushkova, E.N.; Rozhmina, T.A.; Zhuchenko, A.A.; Novakovskiy, R.O.; Povkhova, L.V.; Sigova, E.A.; Zhernova, D.A.; Borkhert, E.V. Isolating *Linum usitatissimum* L. nuclear DNA enabled assembling high-quality genome. *Int. J. Mol. Sci.* **2022**, *23*, 13244. [[CrossRef](#)] [[PubMed](#)]
53. Dvorianinova, E.M.; Pushkova, E.N.; Bolsheva, N.L.; Borkhert, E.V.; Rozhmina, T.A.; Zhernova, D.A.; Novakovskiy, R.O.; Turba, A.A.; Sigova, E.A.; Melnikova, N.V. Genome of *Linum usitatissimum* convar. *crepitans* expands the view on the section *Linum*. *Front. Genet.* **2023**, *14*, 1269837. [[CrossRef](#)] [[PubMed](#)]
54. Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **2019**, *20*, 129. [[CrossRef](#)]
55. Nie, R.-e.; Andújar, C.; Gómez-Rodríguez, C.; Bai, M.; Xue, H.-J.; Tang, M.; Yang, C.-T.; Tang, P.; Yang, X.-K.; Vogler, A.P. The phylogeny of leaf beetles (Chrysomelidae) inferred from mitochondrial genomes. *Syst. Entomol.* **2020**, *45*, 188–204. [[CrossRef](#)]
56. Chen, Q.; Chen, L.; Liao, C.-Q.; Wang, X.; Wang, M.; Huang, G.-H. Comparative mitochondrial genome analysis and phylogenetic relationship among lepidopteran species. *Gene* **2022**, *830*, 146516. [[CrossRef](#)]
57. Mackiewicz, P.; Matosiuk, M.; Świsłocka, M.; Zachos, F.E.; Hajji, G.M.; Saveljev, A.P.; Seryodkin, I.V.; Farahvash, T.; Rezaei, H.R.; Torshizi, R.V.; et al. Phylogeny and evolution of the genus *Cervus* (Cervidae, Mammalia) as revealed by complete mitochondrial genomes. *Sci. Rep.* **2022**, *12*, 16381. [[CrossRef](#)]
58. Zhang, X.; Liu, C.-G.; Yang, S.-H.; Wang, X.; Bai, F.-W.; Wang, Z. Benchmarking of long-read sequencing, assemblers and polishers for yeast genome. *Brief. Bioinform.* **2022**, *23*, bbac146. [[CrossRef](#)]

59. Thambugala, K.M.; Daranagama, D.A.; Tennakoon, D.S.; Jayatunga, D.P.W.; Hongsan, S.; Xie, N. Humans vs. fungi: An overview of fungal pathogens against humans. *Pathogens* **2024**, *13*, 426. [[CrossRef](#)]
60. Reddy, G.K.K.; Padmavathi, A.R.; Nanchaiah, Y.V. Fungal infections: Pathogenesis, antifungals and alternate treatment approaches. *Curr. Res. Microb. Sci.* **2022**, *3*, 100137. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.