Short Communication

Six De Novo Assemblies from Pathogenic and Nonpathogenic Strains of *Fusarium oxysporum* f. sp. *niveum*

James C. Fulton¹ | Jeremy T. Brawner^{1,†} | Jose C. Huguet-Tapia¹ | Katherine E. Smith^{2,3} | Randy Fernandez⁴ | Nicholas S. Dufault^{1,†}

¹ Department of Plant Pathology, University of Florida, Gainesville, FL 32611

² School of Forest Resources and Conservation, University of Florida, Gainesville, FL 32611

³ USDA Forest Service, Southern Institute of Forest Genetics, Saucier, MS 39574

⁴ Entomology and Nematology Department, University of Florida, Gainesville, FL 32611

† Corresponding authors: N. S. Dufault; nsdufault@ufl.edu, and J. T. Brawner; jeremybrawner@ufl.edu

Accepted for publication 13 July 2021.

Abstract

Fusarium wilt, caused by *Fusarium oxysporum* f. sp. *niveum* (Fon), is a soilborne disease that significantly limits yield in watermelon (*Citrullus lanatus*) and occasionally causes the loss of an entire year's harvest. Reference-quality de novo genomic assemblies of pathogenic and nonpathogenic strains were generated using a combination of next-generation and third-generation sequencing technologies. Chromosomal-level genomes were produced with representatives from three Fon races, facilitating comparative genomic analysis and the identification of chromosomal structural variation. Syntenic analysis between isolates allowed for differentiation of the core and lineage-specific portions of their genomes. This research will support future efforts to refine the scientific understanding of the molecular and genetic factors underpinning the Fon host range, develop diagnostic assays for each of the four races, and decipher the evolutionary history of race 3.

Keywords: watermelon, comparative genomics, syteny, structural variants

Author contributions: This project was conceptualized by JB, ND, and JF. DNA extraction and library preparation were conducted by JF and KS. Sequencing and data analyses were completed by JF. JHT advised on sequencing and bioinformatic analyses. RF illustrated the *Fusarium* life cycle. The manuscript was written by JF and JB.

Funding

This research was funded by the National Institute of Food and Agriculture McIntire–Stennis Grant Number FLA-PLP-005931 and the Research Office of the University of Florida's Institute for Food and Agricultural Science.

e-Xtra: One supplementary figure and two supplementary tables are available online.

The author(s) declare no conflict of interest.

Fusarium wilt disease is the outcome of dynamic interactions between the plant host and its fungal pathogen, most often beginning with the initiation of quiescent chlamydospores present in the soil that germinate in response to root exudates released by susceptible seedlings (Costa et al. 2018; Egel and Martyn 2013; Supplementary Fig. S1). Upon germination, hyphae extend and produce an undifferentiated germ tube that attaches to the root surface (Gordon 2017; Ling et al. 2013). If successful, hyphae penetrate the root surface and invade the host's vascular system (Zhang et al. 2015). In response to infection and *Fusarium oxysporum* f. sp. *niveum* (Fon)-derived effectors, tyloses are produced in the xylem tissue in an attempt to prevent systemic fungal colonization (Niu et al. 2016). However, robust tylose production diminishes the efficiency of transpiration, leading to wilting and eventual death. The outcome of an infection depends on the unique interaction between the Fon isolate and infected host, with some cultivars demonstrating partial or complete resistance to certain groups, or races, of isolates (Keinath et al. 2019; Martyn 2014; Martyn and McLaughlin 1983; Zhou et al. 2010). Currently, four races (0, 1, 2, and 3) are recognized and identified by their pathogenicity to standard test cultivars. The observed pathogenic variability among Fon isolates suggests that genomic variation, especially of effectors and virulence factors,

Copyright © 2022 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

contributes to an isolate's ability to cause disease. This hypothesis has been supported by research reporting the successful differentiation of races based on one or a few individual genes or loci (Hudson et al. 2021). For example, Niu et al. (2016) demonstrated that the *SIX6* effector was present in races 0 and 1 but absent in race 2. Furthermore, multiple effectors can be analyzed together to create effector profiles that allow for the identification of more genetically dissimilar populations, as shown by van Dam et al. (2016), who reported distinguishing five different *Cucurbitaceae*-infecting *Fusarium oxysporum* formae speciales based on their unique effector assemblages (van Dam et al. 2016, 2018). In the case of *Fusarium oxysporum* f. sp. *lycopersici*, these suites of effectors have been identified as belonging to unique "pathogenicity" chromosomes. Typically, these pathogenic, or "lineage-specific," chromosomes are specialized to confer pathogenicity by preferentially encoding effectors and virulence factors (Schmidt et al. 2013). Ma et al. (2010) showed that nonpathogenic isolates with structural variants of these chromosomes could become pathogenic through horizontal chromosome transfers of the wild-type strains' corresponding chromosome.

One approach to identifying these lineage-specific chromosomes is through syntenic analysis, which quantifies the similarity between genomes (Kanapin et al. 2020). As the conserved core of the *Fusarium oxysporum* genome has been determined (Ayhan et al. 2018; Ma et al. 2010), the discovery of corresponding regions with high degrees of similarity in de novo assemblies provides an initial filter that can be used to identify noncore regions of the genome that are more likely to include co-located pathogenicity factors. Furthermore, syntenic analysis in combination with the detection of structural variants (genomic insertions, deletions, and inversions typically larger than 1 kilobase) can be a powerful approach to identify effectors and virulence factors that regulate host range (Massonnet et al. 2018). Discovering these variants can be a useful step in identifying pathogenicity factors, as they often encompass regions rich in transposable elements, which have been previously described as factors contributing to pathogenicity in *Fusarium* (Chalvet et al. 2003; Daboussi and Capy 2003; Möller and Stukenbrock 2017).

An improved genetic understanding of Fon is valuable to increase understanding of the pathogen's virulence, as well as providing race identification tools that are quicker, cheaper, and possibly more accurate than the current techniques. These tools will help with enhancing *Fusarium* wilt management through

varietal resistance and other integrated, but costly, management techniques (e.g., fumigation, grafting, and planting date). We present six de novo assemblies we present six de novo assemblies from pathogenic and non-pathogenic isolates from previously described research (Fulton et al. 2021). Isolates were cultured and mycelia harvested as described by Cassago et al. (2002). Samples were then homogenized by grinding in a mortar and pestle with liquid nitrogen, and DNA was extracted following the method described by Vaillancourt and Buell (2019).

Samples were processed for Illumina sequencing (MiSeq v. 3) at the Interdisciplinary Center for Biotechnology Research, University of Florida. Sequencing libraries were prepared with 125 nanograms using the sparQ DNA Library Prep Kit (Quantabio, Beverly, MA) and IDT Unique Dual Indexes (Integrated DNA Technologies, Coralville, IA) following the PCR-free sparQ protocol. A yield of approximately 100 ng of library was obtained with an average length of ~650 bp, ranging from 250 to 1,000 bp. DNA sequencing libraries were barcoded, normalized, and pooled equimolarly for sequencing on a single Illumina MiSeq 2 × 300 cycles run, using a 14-pM loading concentration and a 1% PhiX spike-in control. Approximately 60 million high-quality (% ≥ Q30 reads of ~90%) paired-end reads were obtained for a total yield of approximately 35 Gb of sequencing data.

Short-read assemblies were produced as follows. Adaptor sequences were removed using Trim Galore v. 0.6.5 (Martin 2011), and the first eight bases were trimmed from forward and reverse paired reads with the FASTX toolkit v. 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads with 200 nucleotide bases or longer were kept, and the last 10 nucleotide bases were trimmed from forward reads, whereas the last 30 nucleotide bases were trimmed from reverse reads. FASTQC v. 0.11.7 was used for quality control (Andrews 2010). Reads were assembled into contigs with SPAdes v. 3.13.0 with k-mer values of 21, 33, 55, 71, 91, 111, and 127 (Bankevich et al. 2012). Reads were aligned with Bowtie 2 v. 2.3.5 against the filtered contigs, resulting in SAM format alignment files (Langmead and Salzberg 2012). Using SAMtools v. 1.10, alignments were converted into BAM files and polished with Pilon v. 1.22 to output FASTA files (Li et al. 2009; Walker et al. 2014). Unless otherwise specified, default parameters were used.

Extracted DNA to be sequenced by a third-generation platform (Oxford Nanopore Technologies, Ltd., Oxford, United Kingdom) was first quantified by Qubit 3.0 fluorescence (Thermo Fisher Scientific, Waltham, MA) and sized on the 2200 Agilent

TABLE 1
Fusarium oxysporum f. sp. *niveum* genome assembly summary statistics

Statistic	110407-3-1-1	140508B	150523	150524	150319	150515-1
Bioassay-defined race	Nonpathogenic	0	1	3	3	3
Sequencing technology	Nanopore + Illumina	Illumina	Nanopore + Illumina	Nanopore + Illumina	Illumina	Illumina
Coverage (Illumina)	224 (52)	(28)	166 (30)	84 (39)	(37)	(41)
Number of contigs	33	5,581	45	35	5,405	3,298
Genome size (bp)	49,704,699	55,633,892	66,197,694	61,374,323	55,977,163	52,618,203
GC content (%)	47.39	47.48	47.63	47.65	47.46	47.38
N ₅₀ (bp), Ragtag (SPADES)	5,058,071	4,676,359 (162,590)	4,777,472	4,668,666	5,299,414 (197,527)	4,658,551 (517,561)
L ₅₀	4	5	4	4	5	4
Longest contig (bp) (Ragtag)	4,847,234 (6,473,700)	728,959 (6,307,866)	6,945,410 (7,938,720)	6,493,961 (6,945,849)	959,376 (6,535,239)	1,500,413 (6,048,710)
Contigs > 50 kb	32	238	45	35	220	142
BUSCO completeness (%)	99.8	99.7	99.7	99.7	99.7	99.8
Transposable element (%)	5.59	7.92	17.38	14.50	7.97	6.83
Number of predicted genes	16,177	16,783	21,213	19,688	16,878	15,707
BioSample	SAMN16725394	SAMN16725403	SAMN16727278	SAMN16727288	SAMN16727311	SAMN16727322

TapeStation (Agilent Technologies, Santa Clara, CA). Libraries from quantified DNA were then prepared using the ligation sequencing kit SQK-LSK109 according to the manufacturer's published protocol. One microgram of extracted genomic DNA was repaired using NEBNext FFPE Repair Mix (NEB cat no. M6630) and NEBNext Ultra II End-Repair/dA-tailing Module (NEB cat no. E7546). The repair mix was incubated and purified with AMPure beads. Adaptors were ligated to repaired pure DNA with NEBNext Quick T4 DNA Ligase. Prepped DNA was then sequenced using a FLO-MIN106 (R9.4) flowcell and minION sequencer for 72 h and monitored with ONT's MinKNOW v. 2.0. Data in FAST5 files were converted to nucleotide sequences with Guppy (v. 4.0.11) on the UF computing cluster, and FASTQ files were then concatenated into a single file.

Long-read assemblies were produced as follows. Adaptor sequences present within the concatenated FASTQ file were removed with Porechop v. 0.2.4 (<https://github.com/rrwick/Porechop>) and filtered using Filtlong v. 0.2.0 with a minimum quality score of 90 and nucleotide length of 10,000. SMARTdenovo (<https://github.com/ruanjue/smartdenovo>) was used for de novo genome assembly, with a dot matrix over-lapper engine and k-mer size set to 16. The Burrows-Wheeler Aligner (BWA) v. 0.7.17 program indexed and aligned the reads with the "-x ont2d" parameter selected (Li and Durbin 2009). Aligned reads were then polished against the original FASTQ file using Racon v. 20161013 (Vaser et al. 2017). Parameters chosen included "-m 8, x -6, -g -8, -w 500." Corresponding MiSeq Illumina reads were aligned using BWA to polish the long-read assembly with Pilon v. 1.22.

Isolate 150524 was then aligned to the *Fusarium oxysporum* f. sp. *lycopersici* (Fol) reference strain 4287 (Fol4287) using Minimap2 (Li 2018), and aligned contigs were scaffolded using Ragtag v. 1.0.0 (Alonge et al. 2019). The remaining Fon assemblies were then aligned and scaffolded against isolate 150524 similarly. Any gaps between contigs were filled with 100 'N' characters. Chromosomes in the resulting 150524 genome were identified based on homology with Fol4287, and the remaining scaffolds were recorded separately.

Repetitive elements in the de novo assemblies were annotated using Repeatmodeler v. 2.0 (Smit and Hubley 2008) to generate custom repeat libraries. Annotation of LTR retrotransposons was performed using the LTRharvest option included with Repeatmodeler (Ellinghaus et al. 2008). Repeat sequences were then classified into superfamilies (Wicker et al. 2007) by application of the RepeatClassifier program included with Repeatmodeler. These repeat libraries were then fed into RepeatMasker v. 4.0.9 (Smit et al. 2015) to predict repeat sequences.

Statistics of the de novo assemblies were assessed with BBmap v. 38.44 (Bushnell 2014), and the degree of completeness was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 4.0.6 (Seppey et al. 2019; Simão et al. 2015; Waterhouse et al. 2018) training datasets for Eukaryota and Hypocreales. Satsuma2 was used to assess synteny between scaffolded assemblies using default parameters (Grabherr et al. 2010). Syntenic relationships were visualized using Circos (v. 0.69-9), which represents regions of synteny between isolates with a pictorial ribbon (Krzywinski et al. 2009). Alignments greater than or equal to 10,000 kilobases were shown. Structural variants were detected by first aligning de novo non-scaffolded assemblies to isolate 150524 using the MUMmer (v. 3.23) nucmer program with minimum cluster length (-c) set to 500 and minimum length of an exact match (-l) set to 100 (Delcher et al. 1999, 2002). Aligned assemblies were analyzed using Assemblytics with the default parameters (Nattestad and Schatz 2016). Genomes and unassembled sequence reads were deposited

in the National Center for Biotechnology Information GenBank repository under BioProject number PRJNA675848. BioSample numbers for the sequenced isolates are shown in Table 1.

Assembly sizes ranged from 49.7 Mb in the race 0 isolate to 66.2 Mb in the race 1 isolate. The average size was 56.9 Mb, close to Fol strain 4287's estimated size of 59.9 Mb. Nanopore sequencing coverage ranged from 84 to 224×, whereas the Illumina sequencing coverage range was between 28 and 52×. GC percent content was consistent with a small range from 47.38 to 47.65%. The N₅₀ for scaffolded assemblies was as high as 6.5 Mb in isolate 150524 and as low as 4.7 Mb in isolate 150515-1. There were fewer contigs greater than or equal to 50 kilobases (32, 35, and 45 kilobases from isolates 110407-3-1-1, 150524, and 150523, respectively) synthesized from Nanopore reads, as they were considerably longer than those produced from Illumina reads (142, 220, and 238 kilobases, respectively). Nevertheless, the L₅₀ for scaffolded assemblies was either 4 or 5, regardless of sequencing technology. The BUSCO completeness scores for all de novo assemblies were greater than or equal to 99.8%. There was a considerably higher percentage of repetitive elements detected in the isolates sequenced using Nanopore. For example, the assemblies for isolates 150523 and 150524 comprised 17.38 and 14.50% repetitive elements, respectively. In comparison, the assemblies generated using Illumina sequencing comprised 7.92, 7.97, and 6.83% for isolates 140508B, 150319, and 150515-1, respectively. Lastly, the number of predicted genes ranged from 16,202 to 21,408, with an average of 18,134.

Isolate 150524 was then used as the reference for syntenic comparisons with the other isolates. Isolate 150524 was selected because it was assembled using both long- and short-read technologies and was one of the longest assembled genomes. After 150524 was scaffolded into a pseudo-molecule using Fol4287 as a reference, syntenic analysis of previously identified core chromosomes was completed (Table 2). Chromosomes 1, 2, 4, 5, and 7 to 13 were highly syntenic (≥83.9%) with the corresponding chromosomes in the Fol4287 genome. The 15 remaining scaffolds were assessed for syntenic relationships with variable chromosomes 3, 6, 14, and 15. Of these scaffolds, nine were found to have variable degrees of synteny with chromosomes 3, 6, and 14; however, no scaffolds were detected to have any significant synteny with chromosome 15 (Fig. 1). There were six scaffolds present without a corresponding alignment to Fol4287; four of these were shown to have corresponding matches with scaffolds belonging to at least one of the other de novo assemblies.

Synteny among the Fon isolates was then evaluated using race 3 isolate 150524 as the Fon reference genome. Core chromosomes 1, 2, 5, 7, 8, 9, 12, and 13 were similar (≥70%) to the corresponding chromosomes in the 150524 reference

TABLE 2

Percent syntenic similarity between core chromosomes of isolate 150524 and *Fusarium oxysporum* f. sp. *lycopersici* isolate 4287^z

150524 scaffold	Fol4287 chromosome	Synteny (percent)
CM010335	1	84.81
CM010336	2	89.65
CM010337	4	90.22
CM010338	5	94.01
CM010339	7	94.49
CM010340	8	93.24
CM010341	9	83.95
CM010342	10	83.85
CM010343	11	89.52
CM010344	12	84.72
CM010345	13	90.10

^z Genome data related to isolate Fol4287 from Ayhan et al. (2018).

(Supplementary Table S1). Core chromosomes 4, 10, and 11 were generally syntenic, with some notable exceptions. Chromosome 4 in isolate 110407-3-1-1 shared only 64.1% similarity with the corresponding reference chromosome. Isolates 150319, 150523, and 150515-1 were slightly less syntenic, with 65.1, 68.6, and 69.6% similarity, respectively. Synteny between chromosome 11 in isolate 110407-3-1-1 and the reference was only 67.4%.

Lineage-specific chromosomes were much more variable than core chromosomes. Similarity between chromosome 3 in isolate 110407-3-1-1 was observed to be 3.6%, whereas it was 52.0, 77.4, 56.8, and 21.7% for isolates 140508B, 150523, 150319, and 150515-1, respectively. Chromosome 6 was absent in isolate

110407-3-1-1 and dissimilar in isolates 140508B, 150319, and 150515-1 (8.1, 5.5, and 3.2%, respectively). Synteny of chromosome 14 between the reference and isolates 140508B, 150523, and 150319 was moderate (60.9, 67.5, and 61.2%, respectively) and considerably lower in isolates 150515-1 (25.8%) and 110407-3-1-1 (1.8%). Chromosome 15 was absent in all other *Fon* isolates. In addition to the aforementioned principal *Fusarium oxysporum* chromosomes, four other scaffolds were present in reference 105524 with a corresponding scaffold in at least one of the other isolates. Scaffold 5 similarity with matching scaffolds in isolates 140508B, 150319, and 150515-1 was 77.0, 60.2, and 66.6%, respectively, and was absent in isolates 110407-3-1-1 and

FIGURE 1 Synteny between *Fusarium oxysporum* f. sp. *lycopersici* strain 4287 (blue band) and de novo assembly *Fusarium oxysporum* f. sp. *niveum* (orange band) isolate 150524. Syntenic relationships are indicated by links between homologous genomic regions using Circos v. 0.69-9 (Krzywinski et al. 2009).

150523. The similarity of scaffold 106 was variable across isolates 140508B (47.6%), 150523 (71.9%), 150319 (47.1%), and 150515-1 (37.7%), and the scaffold was absent in isolate 110407-3-1-1. We observed highly syntenic matches to scaffold 362 in isolates 140508B and 150319, with 76.3 and 86.0% similarity, respectively, and this scaffold was missing in isolates 110407-3-1-1, 150523, and 150515-1. Lastly, scaffold 400 was observed in reference 150524, with similar scaffolds in isolates 140508B (39.5%), 150319 (37.5%), and 150515-1 (38.7%); however, no matches were found in isolates 110407-3-1-1 and 150523. The genome fraction, or the total number of bases that align with the reference, was calculated for each isolate. Isolates 110407-3-1-1 and 150515-1 were the most dissimilar at 68.1 and 71.2%,

respectively. Isolates 140508B, 150523, and 150319 were very similar to the reference, with percent similarities of 85.6, 94.6, and 86.0%, respectively.

Structural variation was assessed for all *Fon* genomes (Supplementary Table S2). The total number of variants ranged from a minimum of 121 to a maximum of 1,007, with an average of 527 variants per assembly (Fig. 2). There were more deletions than insertions for isolates 110407-3-1-1 (292 versus 167), 140508B (82 versus 12), 150523 (145 versus 54), 150319 (80 versus 15), and 150515-1 (285 versus 196). Although present, both tandem expansions and contractions constituted only a fraction of the total number of variants, with 1.4, 8.3, 3.0, 10.9, 1.2, and 0.2% for isolates 110407-3-1-1, 140508B, 150523, 150319, 150515-1,

FIGURE 2

Structural variation (deletion [blue], insertion [red], repeat expansion [green], repeat contraction [purple], tandem expansion [orange], and tandem contraction [brown]) between isolates: **A**, 110407-3-1-1; **B**, 140508B; **C**, 150523; **D**, 150319; **E**, 150515-1, and reference genome 150524. Structural variation between 150524 and *Fusarium oxysporum* f. sp. *lycopersici* isolate 4287 is shown in **F**.

and 150524, respectively. Isolates 110407-3-1-1, 150515-1, and 150524 had the largest numbers of repeat expansions and contractions, which constituted 42.6, 39.6, and 46.5% of the total observed variations. These were also present in isolates 140508B, 150523, and 150319 but at a lower percentage of the total (14.05, 21.29, and 15.5%, respectively).

Prior to our research, 12 Fon assemblies were publicly available in sequence databases (van Dam et al. 2016); however, only three of these assemblies were defined to a race-level classification (Hudson et al. 2021). Our efforts increase the number of available assemblies by 50%, increase the number of assemblies with race-level classification by 200%, and present the first completely assembled genomes from multiple isolates sampled throughout the state of Florida. As assemblies for isolates 110407-3-1-1, 150523, and 150524 were developed with a long-read sequencing platform with later error correction through the addition of short-read technology, they provide a reliably accurate reference for future genomics studies, including of repetitive regions of the genome, often sources of novel genes including effectors and pathogenicity factors, which otherwise are typically poorly represented by short-read platforms (Jurka et al. 2007; Kanapin et al. 2020). Importantly, the 150524 assembly showed significant synteny with the previously identified core chromosomes (1, 2, 4, 5, 7 to 13) in Fol isolate 4287.

These core chromosomes are evolutionarily conserved and stable across isolates of the same species, as they are most often associated with basic metabolism and cellular function. On the contrary, lineage-specific chromosomes (3, 6, 14, 15) are more variable and determine pathogenicity and host range. Consistent with this trend, synteny was significantly lower for these chromosomes, and there was no clear chromosome in the de novo assemblies corresponding to chromosome 15 in Fol4287. Interestingly, scaffold 106, approximately 1 million base pairs in length, was observed with moderate synteny in all isolates except for the nonpathogenic isolate 110407-3-1-1. It is possible that this scaffold was in fact chromosome 15 but sufficiently divergent from the Fol4287 reference chromosome 15 that it was not identified. In addition to the absence of scaffold 106, isolate 110407-3-1-1 was distinguished by other qualitative differences, including its genome size (49,704,699 bp) compared with the average (56,917,662 bp), the absence of any chromosomes corresponding to chromosome 6, and especially poor synteny to chromosomes 6 and 14. The described research suggests that future efforts should be focused on chromosomes 3, 6, 14, and 15 to identify virulence factors that will allow for unambiguous classification of Fon isolates into the races that have traditionally been identified using differentials in cultivar susceptibility trials. Furthermore, this research suggests significant genomic differences between pathogenic and nonpathogenic strains, and these differences may be used for future diagnostic assay development and possible biological management compounds.

LITERATURE CITED

- Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., and Schatz, M. C. 2019. RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* 20:1-17.
- Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.
- Ayhan, D. H., López-Díaz, C., Di Pietro, A., and Ma, L.-J. 2018. Improved assembly of reference genome *Fusarium oxysporum* f. sp. *lycopersici* strain Fol4287. *Microbiol. Resour. Announc.* 7:e00910-18.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., and Pribelski, A. D. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477.
- Bushnell, B. 2014. BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Lab (LBNL), Berkeley, CA.
- Cassago, A., Panepucci, R. A., Baião, A. M. T., and Henrique-Silva, F. 2002. Cellophane based mini-prep method for DNA extraction from the filamentous fungus *Trichoderma reesei*. *BMC Microbiol.* 2:14.
- Chalvet, F., Grimaldi, C., Kaper, F., Langin, T., and Daboussi, M.-J. 2003. Hop, an active mutator-like element in the genome of the fungus *Fusarium oxysporum*. *Mol. Biol. Evol.* 20:1362-1375.
- Costa, A. E. S., da Cunha, F. S., Honorato, A. D., Capucho, A. S., Dias, R. D. S., Borel, J. C., and Ishikawa, F. H. 2018. Resistance to *Fusarium* wilt in watermelon accessions inoculated by chlamydospores. *Sci. Hortic.* 228:181-186.
- Daboussi, M.-J., and Capy, P. 2003. Transposable elements in filamentous fungi. *Annu. Rev. Microbiol.* 57:275-299.
- Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. 1999. Alignment of whole genomes. *Nucleic Acids Res.* 27:2369-2376.
- Delcher, A. L., Phillippy, A., Carlton, J., and Salzberg, S. L. 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res.* 30:2478-2483.
- Egel, D. S., and Martyn, R. D. 2013. *Fusarium* wilt of watermelon and other cucurbit crops. *Plant Health Instr.* Online publication. <http://dx.doi.org/10.1094/PHI-I-2007-0122-01>.
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinform.* 9:18.
- Fulton, J. C., Amaradasa, B. S., Ertek, T. S., Iriarte, F. B., Sanchez, T., Ji, P., Paret, M. L., Husdon, O., Md. Emran, A., and Dufault, N. S. 2021. Phylogenetic and phenotypic characterization of *Fusarium oxysporum* f. sp. *niveum* isolates from Florida-grown watermelon. *PLoS One* 16:e0248364. <https://doi.org/10.1371/journal.pone.0248364>
- Gordon, T. R. 2017. *Fusarium oxysporum* and the *Fusarium* wilt syndrome. *Annu. Rev. Phytopathol.* 55:23-39.
- Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., and Lindblad-Toh, K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26:1145-1151.
- Hudson, O., Waliullah, S., Fulton, J. C., Ji, P., Dufault, N. S., Keinath, A., and Ali, M. E. 2021. Marker development for differentiation of *Fusarium oxysporum* f. sp. *niveum* race 3 from races 1 and 2. *Int. J. Mol. Sci.* 22:822.
- Jurka, J., Kapitonov, V. V., Kohany, O., and Jurka, M. V. 2007. Repetitive sequences in complex genomes: Structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8:241-259.
- Kanapin, A., Samsonova, A., Rozhmina, T., Bankin, M., Logachev, A., and Samsonova, M. 2020. The genome sequence of five highly pathogenic isolates of *Fusarium oxysporum* f. sp. *lini*. *Mol. Plant-Microbe Interact.* 33:1112-1115.
- Keinath, A. P., Coolong, T. W., Lanier, J. D., and Ji, P. S. 2019. Managing *Fusarium* wilt of watermelon with delayed transplanting and cultivar resistance. *Plant Dis.* 103:44-50.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. 2009. Circos: An information aesthetic for comparative genomics. *Genome Res.* 19:1639-1645.
- Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357.
- Li, H. 2018. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094-3100.
- Li, H., and Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079.
- Ling, N., Zhang, W., Wang, D., Mao, J., Huang, Q., Guo, S., and Shen, Q. 2013. Root exudates from grafted-root watermelon showed a certain contribution in inhibiting *Fusarium oxysporum* f. sp. *niveum*. *PLoS One* 8:e63383.
- Ma, L.-J., Van Der Does, H. C., Borkovich, K. A., Coleman, J. J., Daboussi, M.-J., Di Pietro, A., Dufresne, M., Freitag, M., Grabherr, M., and Henrissat, B. 2010. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464:367-373.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17:10-12.
- Martyn, R. D. 2014. *Fusarium* wilt of watermelon: 120 years of research. *Hortic. Rev.* 42:349-442.
- Martyn, R. D., and McLaughlin, R. J. 1983. Effects of inoculum concentration on the apparent resistance of watermelons to *Fusarium oxysporum* f. sp. *niveum*. *Plant Dis.* 67:493-495.

- Massonnet, M., Morales-Cruz, A., Minio, A., Figueroa-Balderas, R., Lawrence, D. P., Travadon, R., Rolshausen, P. E., Baumgartner, K., and Cantu, D. 2018. Whole-genome resequencing and pan-transcriptome reconstruction highlight the impact of genomic structural variation on secondary metabolite gene clusters in the grapevine esca pathogen *Phaeoacremonium minimum*. *Front. Microbiol.* 9:1784.
- Möller, M., and Stukenbrock, E. H. 2017. Evolution and genome architecture in fungal plant pathogens. *Nat. Rev. Microbiol.* 15:756.
- Nattestad, M., and Schatz, M. C. 2016. Assemblytics: A web analytics tool for the detection of variants from an assembly. *Bioinformatics* 32:3021-3023.
- Niu, X., Zhao, X., Ling, K.-S., Levi, A., Sun, Y., and Fan, M. 2016. The *FonSIX6* gene acts as an avirulence effector in the *Fusarium oxysporum* f. sp. *niveum*-watermelon pathosystem. *Sci. Rep.* 6:1-7.
- Schmidt, S. M., Houterman, P. M., Schreiver, I., Ma, L., Amyotte, S., Chellappan, B., Boeren, S., Takken, F. L., and Rep, M. 2013. MITes in the promoters of effector genes allow prediction of novel virulence genes in *Fusarium oxysporum*. *BMC Genom.* 14:119.
- Seppy, M., Manni, M., and Zdobnov, E. M. 2019. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol Biol.* 1962:227-245.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Smit, A., Hubley, R., and Green, P. 2015. RepeatMasker Open-4.0, 2013-2015.
- Smit, A. F., and Hubley, R. 2008. RepeatModeler Open-1.0. <http://www.repeatmasker.org>.
- Vaillancourt, B., and Buell, C. R. 2019. High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing. *BioRxiv*: 783159.
- van Dam, P., de Sain, M., Ter Horst, A., van der Gragt, M., and Rep, M. 2018. Use of comparative genomics-based markers for discrimination of host specificity in *Fusarium oxysporum*. *Appl. Environ. Microbiol.* 84: e01868-17.
- van Dam, P., Fokkens, L., Schmidt, S. M., Linmans, J. H., Kistler, H. C., Ma, L. J., and Rep, M. 2016. Effector profiles distinguish formae speciales of *Fusarium oxysporum*. *Environ. Microbiol.* 18: 4087-4102.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27:737-746.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., and Young, S. K. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9: e112963.
- Waterhouse, R. M., Seppy, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35:543-548.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., and Panaud, O. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8:973-982.
- Zhang, M., Xu, J., Liu, G., Yao, X., Li, P., and Yang, X. 2015. Characterization of the watermelon seedling infection process by *Fusarium oxysporum* f. sp. *niveum*. *Plant Pathol.* 64:1076-1084.
- Zhou, X. G., Everts, K. L., and Bruton, B. D. 2010. Race 3, a new and highly virulent race of *Fusarium oxysporum* f. sp. *niveum* causing fusarium wilt in watermelon. *Plant Dis.* 94:92-98.