Optical mapping of the *Fusarium oxysporum* f. sp. *melongenae* genome

Duygu Ates¹ | Hacer H. Altinok² | Muhammed Bahattin Tanyolac¹

¹Molecular Genetic Laboratory, Department of Bioengineering, Ege University, Izmir, Turkey
²Department of Plant Protection, Faculty of Agriculture, Erciyes University, Kayseri, Turkey

Correspondence
Muhammed Bahattin Tanyolac,
Molecular Genetic Laboratory, Department of Bioengineering, Ege University, 35040 Izmir, Turkey.
Email: bahattin.tanyolac@ege.edu.tr

Funding information
The Scientific and Technological Research Council of Turkey (TUBITAK), Grant/Award Number: COST-114O866

Abstract
Optical mapping approaches are widely preferred and applied in different branches of genomic studies because of their accuracy, low cost, and high efficiency. In the current study, a sequence orientation of the *Fusarium oxysporum* f. sp. *melongenae* (FOMG) genome that is deposited in GenBank National Center for Biotechnology Information under accession number MPIL0000000 was used as the reference genome, which we checked with Bionano Genomics optical mapping approaches. The optical mapping produced 103 contigs, the longest of which was 3.05 Mb. The N50 value of optical map contigs is 0.85 Mb. The sequences of the FOMG reference genome and optical map mainly match each other. Results obtained in the current study indicate that optical mapping can be used to construct complete and gapless assemblies of the FOMG genome. It also can be applied to validate a previous genome assembly.

KEYWORDS
bionano, eggplant, *Fusarium oxysporum* f. sp. *melongenae*, genome sequencing, optical mapping

1 | INTRODUCTION

Eggplant (*Solanum melongena*, 2II = 24) is an economically important vegetable crop, especially in warm and tropical climate locations. Comparable to tomatoes in total nutritional value, it is a rich source of various vitamins and minerals [1]. *Fusarium oxysporum* f. sp. *melongenae* (FOMG) is the causal agent of fusarium wilt on eggplant, and it is one of the most important and widespread fungal diseases that limit the production of eggplant around the world [1-4]. This disease reported in Japan [5], Kenya [6], The Netherlands [7], United States of America [8,9], Italy [10], Greece [11], Korea [12], Spain [13], China [14], and Turkey [1,2,15]. This disease arises under suitable environmental conditions and eggplants that are affected by this fungus present with characteristic vascular wilt symptoms that primarily consist of slight vein clearing and chlorosis of the leaf on the outer leaflets. The leaves turn yellow and eventually fall off. Discoloration of the xylem is observed in the following period and finally, the above-ground parts of the eggplant die [1]. Control of this disease is very difficult, and the fungus can spread widely and remain in the soil for years. Furthermore, suppression of this fungus is difficult due to the limited availability of recommended registered fungicides [1]. Therefore, control of fusarium wilt requires the development of FOMG-resistant eggplant varieties.

Having sufficient information about the genetic structure of a pathogen is the most important

Abbreviation: FOMG, *Fusarium oxysporum* f. sp. *melongenae*.
requirement for developing resistant varieties. Recent genetic research on FOMG has benefited from a partially ordered draft genome sequence assembly that was uploaded to GenBank from the National Center for Biotechnology Information (NCBI) since in 2017 [16]. However, there are some limitations in these currently available data, and these limitations are usually the direct results of two typical genomic properties: (i) too many repetitive and transposable elements [17], and (ii) significantly decreased meiotic recombination frequency in pericentromeric regions [18]. The F. oxysporum genome contains core and lineage-specific regions [19]; core regions are conserved in all strains of F. oxysporum, while lineage-specific regions generally include high numbers of repetitive and transposable elements [17,20,21]. These elements can cause severe misorientation of the genome sequences from the raw data sequences generated from the sequencing process [22]. They also restrict the contiguity of whole-genome assemblies to kilobase-sized sequences that originate from low copy regions of the FOMG genome. Therefore, detailed research on the repetitive and transposable elements of the FOMG genome has not been conducted. To address these genome assembly problems while improving the quality of sequence data, an approach that combines long reads and optical mapping can be used.

Optical mapping involves the physical mapping of a genome. It produces high-resolution restriction maps that generate the correct genome orientation by comparing genome sequences. To construct a high-resolution sequence motif map, this approach uses the distance between restriction sites [22]. Optical mapping is also more precise because its output is based on a physical consequence. Optical mapping, combined with de novo assembly, can determine the actual orientation of genome contigs. Thus, this combination of techniques can predict the position and size of gaps [22]. By using optical mapping techniques, regions of unknown orientation and local order can be corrected and located, and mis-assembled sequences such as misorientations, pseudoduplications, or chimeras can be repaired [22]. Therefore, proper editing of existing sequence data is as significant as uncovering new genome sequences in the developing genome references process. On the other hand, a translocation is a genomic rearrangement in which a particular sequence is moved from one region in the genome to another. Similar to translocations, duplications, and inversions also alter the location, orientation, and certain sequence’s copy number, creating differences in genome structure [21]. Previous studies have indicated that developments in the techniques of optical mapping can ease de novo genome assembly when utilized with parallel sequencing [22]. Optical mapping approaches have also been applied in the analysis of the structural variation of the human genome [23,24]. Therefore, the objective of the current study was to check the sequence orientation of the FOMG genome using optical mapping approaches.

2 | MATERIALS AND METHODS

2.1 | Fungal isolate and high-quality megabase-sized DNA extraction

Pathogenic strains of FOMG were isolated from infected vascular tissues of eggplant collected from Menemen. A highly virulent single-spore isolate (FOMG109) was selected for the study. This single-spore isolate was placed on potato dextrose agar (PDA; Detroit) in a plate and these PDA plates were incubated at 27 ± 1°C until the FOMG109 colony grew (~5–7 days). Then, FOMG109 grown on PDA was inoculated aseptically into the Erlenmeyer flasks (250 ml) containing potato dextrose broth (Detroit) media of 55 ml each. To mycelial growth, these Erlenmeyer flasks were placed on an orbital shaker at 110 rpm for 6 days at 28°C. We followed the protocol of Zhang et al. [25], with some minor modifications, to extract high-quality megabase-sized DNA from FOMG109. Fungal DNA was extracted from a total of 20 g mycelia and the mycelia were separated by vacuum filtration through sterilized filter paper in a Büchner funnel. After 20 g of FOMG109 mycelium is obtained, intact nuclei from this fungal tissue was prepared following the method of option-A in the protocol of Zhang et al. [25], and this nuclei was embedded in agarose plugs (low-melting-point) [25]. Finally, high-quality megabase-sized DNA was extracted from these agarose plugs [25] and stored in Tris-EDTA buffer solution for further analyses. The purity of this extracted DNA was visually assessed by electrophoresis using 1% agarose gel.

2.2 | Map assembly via BioNano genomics

The nicking endonuclease enzyme, Nt.BspQ1 (GCTCTTC; New England BioLabs) was preferred to nick DNA of FOMG at particular sequence motifs. These nicked molecules of FOMG were first labeled and then stained following the steps of IrysPrep Reagent Kit (BioNano Genomics) as defined in the protocol of Luo et al [26]. Irys platform of BioNano Genomics was used to construct an optical map of the FOMG genome. Prepared samples were loaded onto the IrysChip nanochannel array (BioNano Genomics) and run. This loaded
DNA was first linearized then imaged automatically by this platform (BioNano Genomics). To construct single molecule maps and following to de novo assemble them into consensus physical maps, the software package of the IrysView (BioNano Genomics) was used. AutoDetect software was utilized to provide data about length information of DNA and also basic labeling. Using this software, the data of the raw DNA molecules that smaller than 20 kb were converted into " `.bnx " files. Assembly pipeline of BioNano Genomics [27] was used to aligned, clustered, and assembled all DNA molecules that passing quality control and also >180 kb. The thresholds of \( p \)-value were used as \( 2 \times 10^{-8} \) for pairwise assembly stages, \( 1 \times 10^{-9} \) for extension/refinement stages, and \( 1 \times 10^{-15} \) for final refinement stages [27]. Finally, constructed optical maps of FOMG were controlled if there are any potential chimeric contigs of optical maps.

2.3 | Sequences of FOMG from NCBI

The most recent draft genome of FOMG [16] was obtained from GenBank NCBI under the accession number MPIL00000000 was used as a reference genome.

2.4 | Comparison between the sequence–map and the map–map

To compare the previously published FOMG draft genome with the optical map, sequences from both maps were digested in silico with Nt. BspQ1 using Knickers software. The recognition sequence of BspQI is 5’ GCTCTTC N1/N4 3’ and it is a thermostable Type IIS restriction endonuclease (REase). This recognition sequence of BspQI was searched at NCBI and labeled where this sequence was detected. Sequence assemblies of the optical maps were aligned using RefAligner and the alignments were visualized using BioNano Genomics software. Finally, snapshots of these alignments were saved using IrysView software (https://bionano genomics.com/).

3 | RESULTS

The optical map of the FOMG genome was generated using Bio Nano Genomics (BNG) technology. FOMG DNA molecules were nicked using Nt.BspQ1 nickase, and the resulting nicks were fluorescently labeled. An electric force was applied to stretch the DNA molecules linearly, and then finally, all images were assembled into optical map contigs and maps. The assembly consists of 103 contigs, and the longest optical map contig measures 3.05 Mb. The \( N_{50} \) value of the optical map contig, which is 0.85 Mb.

The completed FOMG optical map encompasses 15 chromosomes, corresponding to the haploid chromosome (chr) number of FOMG. All contigs measure 72.27 Mb, and after removing all the overlaps, the optical maps map of the FOMG genome is 57.7 Mb long. Table 1 presents the length/bp of each chromosome. The largest chromosome is chr1 (6,854,690 length/bp), while the smallest chromosome is chr14 (1,646,465 length/bp; Table 1). The number of nucleotides in the overlapping regions were 356,780 N/bp (on chr13; Table 1) and 3,513,368 N/bp (on chr3; Table 1). A previous sequence assembly [16] contains 15 chromosomes and other supercontigs, in which 11 chromosomes (chr 1, 2, 4, 5, 7–13) were aligned and validated with optical map contigs. We were unable to align the remaining four chromosomes (chr 3, 6, 14, and 15) to any optical map contig (Table 1). The \( N\% \) values of chr 3, 6, 14, and 15 are over or close to 50 \( N\% \) (Table 1). Meanwhile, the lengths of chr 8, 13, and the supercontig of the optical maps were longer than their corresponding sequences of the reference genome, and the lengths of the other chromosomes were shorter (Figure 1).

The sequences in the optical map are compared in Figure 1. In general, there are 15 chr and other
FIGURE 1  (See caption on next page)
supercontigs in the reference sequence assembly of Dong et al. [16] of which 11 chromosomes can be aligned with a high degree of similarity and validated with optical map contigs, while the other four chromosomes (chr 3, 6, 14, and 15) could not be aligned with any other optical map contigs (Figure 1).

4 | DISCUSSION

Short-read sequencing techniques have made it affordable to study the genetics and the genomes of various species [28]. However, the short sequence lengths obtained using these techniques make it difficult to assemble sequencing reads of the most complex and repetitive regions of the genome, resulting in collapsed and fragmented genome assemblies [28]. Therefore, assemblies based on short-read sequencing are limited in the accuracy of downstream analyses such as the identification of genomic variations [29]. Several methodologies, such as next-generation sequencing (NGS) and recently developed algorithms, have been developed to overcome this problem, it has been difficult to completely eliminate the shortcomings of short-read length sequencing [24]. The power of NGS lies in its capacity to generate a huge volume of reads. However, because these reads are rather short, it is impossible to resolve the assembly of many families of repetitive DNA elements that populate fungal genomes. Moreover, fungal genome sequencing is also hampered by the wrong orientation of assembled contigs.

The recent development of long-read sequencing methods, such as optical mapping, have enabled the spanning of most repeats and the generation more complete and correct genome assemblies [30]. Optical mapping techniques have been performed extensively in genomic studies on human diseases, animals, plants, and microorganisms [24]. In the original version of the method, high molecular weight DNA molecules were cleaved on an open glass surface and then imaged on fluorescence microscopy [31]. The image from the fluorescence microscopy help to order the partially digested molecules, which is called an optical map. The genomic sequence of an organism from de novo sequencing is align on to the image and therefore sequence information is placed in the correct order of the optical map [32]. Using these techniques, it is possible to detect errors in the arrangement and structure of contigs, even if the organism’s genome contains distinct repetitive regions [33]. Therefore, optical maps are an ideal means of finishing fungal sequence assembly because in most cases, the restriction regions cover repetitive regions [18] such as retrotransposons [20]. Relying on the currently published complete genome sequences that are stored at NCBI may be risky because such sequences likely contain assembly errors [22]. In the current study, BNG technology was used to generate an optical map of FOMG and verify contigs. We demonstrated that short-read sequencing assembly could generate some assembly errors within the FOMG genome (Table 1, Figure 1). Thus, by utilizing more uniform linearization, BNG technology highly improves the throughput and accuracy of genome length prediction. Moreover, by using nicking enzymes to generate only single-strand breaks, this approach preserves the contiguity of molecules more than other optical mapping technologies [24].

Dong et al. [16] reported that the size of the FOMG reference genome is 53.9 Mb, with an N50 value of 0.56 Mb. However, optical mapping indicates a genome size of 57.7 Mb with an N50 value of 0.85 Mb (Table 1). The difference in sequence length may be due to errors in the assembly of the draft genome, which was produced by the Genome Québec platform [16]. These errors were corrected by optical mapping to ensure the efficient use of the data. Support of our results, previous studies have indicated that the length of the optical map of FOMG is slightly different from that of the assembled genome [23,34]. Although the length differences is often attributed to repetitive regions of the genome, they may also result from the poor quality of the assemblies using both sides of the comparison [23]. The supercontig optical map was obtained from a few genomic sequences that were aligned onto the supercontig. This is likely because previously sequenced FOMG DNA was not derived from high molecular weight DNA (Figure 1).

We obtained a high N50 value of 0.85 Mb, indicating that a high percentage of the optical map aligned with the reference map. The N50 value defines how much of an assembly is consists of segments larger than a given size where “N” is the size of the scaffold or contig and “50” is the assembly length percentage [23]. Solely genome assemblies with mega-base scale N50s can be

**Figure 1** Comparisons of sequences to the optical map. The green box is a BNG contig (~6.5 Mb), blue boxes are sequences, the red lines within boxes are restriction sites, and the red lines between green and blue boxes indicate matches. The potential chimeric points in each sequence are indicated by red arrows and the original sequence name is the figure title. (a)-(l) are optical mapping of chromosome aligning with sequence. BNG, Bio Nano Genomics.
verified performing optical mapping technology because contigs/scaffolds smaller than 100 kb usually do not have sufficient nick knowledge to be securely aligned [23].

In the current study, each DNA sequence was converted into a restriction map format to match the sequence with the optical map contigs. As seen in Figure 1, correct matches between the nick patterns of the optical map contigs and the reference DNA sequence confirms the base pair distances between nick sites in the DNA sequence. Mismatched regions of the genome may be due to the poor quality of the assemblies of the reference genome.

Optical mapping of the FOMG genome was completed in the current study. The genome sequence of FOMG that deposited into the GenBank NCBI under the accession number MPIL00000000 and optical map mainly match each other. On the other hand, it is still unsequenced genomic regions exist in the FOMG genome. The current sequence order of the FOMG genome deposited into GenBank NCBI is not in the right order. This error is likely due to the presence of highly repetitive elements distributed throughout the genome, causing wrong genome assembly. Usage of megabase-sized DNA as a sequencing material is important to enable the coverage of the entire genome. After completing the assembly of genome sequences, the completed genome should be aligned with the optical map.

ACKNOWLEDGMENTS
The study was supported by The Scientific and Technological Research Council of Turkey (TUBITAK), project no. COST-1140866. This study was funded by The Scientific and Technological Research Council of Turkey (TUBITAK), Grant/Award Number: COST-1140866.

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT
Optical mapping data presented in the manuscript are available from the corresponding author upon reasonable request.

ORCID
Muhammed Bahattin Tanyolac ♦ http://orcid.org/0000-0002-4368-0988

REFERENCES
[1] Altink HH, Can C. Characterization of Fusarium oxysporum f. sp. melongenae isolates from eggplant in Turkey by pathogenicity, VCG and RAPD analysis. Phytoparasitica. 2010;38:149–57.
[2] Altink HH, Can C, Altink MA. Characterization of Fusarium oxysporum f. sp. melongenae isolates from Turkey with ISSR markers and DNA sequence analyses. Eur J Plant Pathol. 2018;150:609–21.
[3] Sahar P, Sahi ST, Jabbar A, Rehman A, Riaz K, Hannan A. Chemical and biological management of Fusarium oxysporum f. sp. melongenae. Pak J Phytopathol. 2013;25:155–9.
[4] Kalloo G. Eggplant: Solanum melongena L. In: Kalloo G, Bergh BO, editors. Genetic Improvement of Vegetable Crops. Oxford, New York: Pergamon Press; 1993. p. 587–604.
[5] Matuo T, Ishigaki M. On the wilt of Solanum melongena L. and its causal fungus Fusarium oxysporum f. melongenae nf. Jpn J Phytopathol. 1958;23:189–92.
[6] Nattrass RM. Host lists of Kenya fungi and bacteria. Mycol Papers. 1961;81:1–46.
[7] van Steekelenburg N. Fusarium wilt of eggplant in the Netherlands. Neth J Plant Pathol. 1976;82:191–2.
[8] Alfersi S Jr. Index of plant diseases in Florida. Fla Dept Agric Cons Serv Div Plant Ind Bull. 1984;11:1–389.
[9] Alfersi S. Diseases and disorders of plants in Florida, Florida Department of Agriculture and Consumer Services, Division of Plant Industry. 1994. p. 14.
[10] Cappelli C, Stravato V, Rotino G, Buonaurio R. Sources of resistance among Solanum spp. to an Italian isolate of Fusarium oxysporum f. sp. melongenae. Meeting on Genetics and Breeding on Capsicum and Eggplant of EUCARPIA, 9th Budapest, Aug 21-25, Hungary HU. 1995.
[11] Holevas C, Chitzanidis A, Pappas A, Tzamos E, Elena K, Psallidas PG, et al. Disease agents of cultivated plants observed in Greece from 1981 to 1990. Ann Inst Phytopathol Benaki. 2000;19(1):1–96.
[12] Cho WD, Shin HD. List of plant diseases in Korea. 4th ed. Seoul, Korea: Hanrinwon Publishing Company; 2004.
[13] Urrutia Herrada M, Gomez Garcia V, Tello Marquina J. Fusarium wilt on eggplant in Almeria (Spain). Boletin de Sanidad Vegetal, Plagas. 2004;30:85–92.
[14] Zhuang W. Fungi of northwestern China. Ithaca, NY, USA: Mycotaxon, Ltd; 2005.
[15] Altunok H. First report of Fusarium wilt of eggplant caused by Fusarium oxysporum f. sp. melongenae in Turkey. Plant Pathol. 2005;54:577.
[16] Dong Z, Hsiang T, Luo M, Xiang M. Draft genome sequence of an isolate of Fusarium oxysporum f. sp. melongenae, the causal agent of Fusarium wilt of eggplant. Genome Announc. 2017;5:e01597–16.
[17] Rao S, Sharda S, Oddi V, Nandineni MR. The landscape of repetitive elements in the refined genome of chilli anthracnose fungus Colletotrichum truncatum. Front Microbiol. 2018;9:2367.
[18] Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc Lond B Biol Sci. 2017;372:20160455.
[19] Deltour P, Franca S, Heyman L, Pereira O, Hofte M. Comparative analysis of pathogenic and nonpathogenic Fusarium oxysporum populations associated with banana on a farm in Minas Gerais, Brazil. Plant Pathol. 2018;67:707–18.
[20] Ates D, Altink HH, Ozkuru F, Ferik F, Erdoganus S, Can C, Tanyolac MB. Population structure and linkage...
disequilibrium in a large collection of *Fusarium oxysporum* strains analysed through iPBS markers. J Phytopathol. 2019; 167:576–90.

[21] Zhang Y, Yang H, Turra D, Zhou S, Ayhan DH, Deiulio GA, et al. The genome of opportunistic fungal pathogen *Fusarium oxysporum* carries a unique set of lineage-specific chromosomes. Commun Biol. 2020;3:50.

[22] de Jesus Sousa T, Parise D, Profeta R, Parise MTD, Gomide ACP, Kato RB, et al. Re-sequencing and optical mapping reveals misassemblies and real inversions on *Corynebacterium pseudotuberculosis* genomes. Sci Rep. 2019; 9:1–11.

[23] Udall JA, Dawe RK. Is it ordered correctly? Validating genome assemblies by optical mapping. Plant Cell. 2018;30:7–14.

[24] Yuan Y, Chung CYL, Chan TF. Advances in optical mapping for genomic research. Comput Struct Biotechnol J. 2020;18: 2051–62.

[25] Zhang M, Zhang Y, Scheuring CF, Wu CC, Dong JJ, Zhang HB. Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. Nat Protoc. 2012;7:467–78.

[26] Luo MC, Deal KR, Murray A, Zhu T, Hastie AR, Stedman W, Sadowski H, Saghbini M. Optical nano-mapping and analysis of plant genomes. Methods Mol Biol. 2016;1429:103–17.

[27] Zhu T, Hu Z, Rodríguez JC, Deal KR, Dvorak J, Vogel JP, Liu Z, Luo MC. Analysis of *Brachypodium* genomes with genome-wide optical maps. Genome. 2018;61:559–65.

[28] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2012;13:36–46.

[29] Cameron DL, di Stefano L, Papenfuss AT. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. Nat Commun. 2019;10: 1–11.

[30] Chen P, Jing X, Liao B, Zhu Y, Xu J, Liu R, Zhao Y, Li X. BioNano genome map resource for *Oryza sativa* ssp. *japonica* and *indica* and its application in rice genome sequence correction and gap filling. Mol Plant. 2017;10:895–8.

[31] Meng XJ, Paul PS, Halbur PG, Morozov I. Sequence comparison of open reading frames 2 to 5 of low and high virulence United States isolates of porcine reproductive and respiratory syndrome virus. J GenVirol. 1995;76:3181–8.

[32] Staňková H, Hastie AR, Chan S, Vrana J, Tulápová Z, Kubaláková M, Visendi P, Hayashi S, Luo M, Batley J, Edwards D, Doležel J, Šimková H. BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnol J. 2016;14:1523–31.

[33] Latreille P, Norton S, Goldman BS, Henkhaus J, Miller N, Babaszak B, Bode HB, Darby C, Du Z, Forst S, Gaudriault S, Goodner B, Goodrich-Blair H, Slater S. Optical mapping as a routine tool for bacterial genome sequence finishing. BMC Genom. 2007;8:1–6.

[34] van Kan JA, Stassen JH, Mosbach A, van der Lee TA, Faino L, Farmer AD, Papasotiriou DG, Zhou S, Seidl MF, Cottam E, Edel D, Hahn M, Schwartz DC, Dietrich RA, Widdison S, Scalliet G. A gapless genome sequence of the fungus *Botrytis cinerea*. Mol Plant Pathol. 2017;18:75–89.