Vol. 18(31), pp. 1031-1043, November, 2019
DOI: 10.5897/AJB2019.16978
Article Number: 9D14A7362342
ISSN: 1684-5315

http://www.academicjournals.org/AJB

ACADEMIC JOURNALS
expand your knowledge

African Journal of Biotechnology

Full Length Research Paper

Whole-genome optical mapping: Improving assembly of *Macrophomina phaseolina* MS6 through spanning of twelve blunt end chromosomes by obviating all errors and misassembles

Quazi Md. Mosaddeque Hossen[1]*, Md. Shahidul Islam[1], Emdadul Mannan Emdad[1], Md. Samiul Haque[1], Md. Monjurul Alam[1] and Maqsudul Alam[1,2]

[1]Basic and Applied Research on Jute (BARJ) Project, Bangladesh Jute Research Institute, Dhaka, Bangladesh.
[2]Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii 96822, USA.

Received 7 October, 2019; Accepted 8 November, 2019

Deciphering genetic information through next-generation sequencing (NGS) is considered as the basic platform to unveil in details of an organism. However, as it produces short reads that lead to difficulties in assembly, we generated long scaffold-based optical mapping (OM) data of previously sequenced devastating fungus, *Macrophomina phaseolina* MS6. In the process, *Kpn*I identified as the most effective restriction endonuclease among tested 13, used to digest high molecular weight (HMW) DNA that generated 270,343 genomic DNA molecules size in more than 250 kb. The molecules were assembled and constructed 12 super-scaffolds (terminated with telomeric blunt-ends and denoted as chromosomes) that were aligned with NGS generated 17 (out of 88 reduced from 94) reference scaffolds. The state-of-the-art technology revealed concordances and different discordances viz., inversions, low-quality assembly, gaps, overlaps followed to correct the NGS misassembles. Based on the results, OM generated improved and validated assembly advance our understanding of the chromosome evolution of fungi. This furnished data might be considered as valuable resources to accelerate the precise planning for the protection of *M. phaseolina* MS6 infected sequenced crops through developing the cross-talk phenomenon between the host and pathogen.

Key words: *Macrophomina phaseolina* MS-6, optical mapping, assembly improvement, assembly validation.

INTRODUCTION

Genome sequencing is the process of determining the full DNA sequences of living organisms including its chromosomal, mitochondrial and chloroplast DNA. Although DNA sequencing seemed not to be an easier and faster process, the rapid development of different pipelines and techniques made the whole genome sequences simple over the last few years (Koboldt et al., 2013). However, NGS produces a large number of short reads restraints the *de novo* assembly due to repeat or complex region of genome that suffers extensive misassembles and comprise gaps (Pendleton et al., 2015; Ganapathy et al., 2014; Ruperao et

*Corresponding author. E-mail: mosaddeque@jutegenome.org.

al., 2014). Therefore, the demand of introducing a new technique was a must to minimize these errors in terms of whole-genome sequencing.

Whole-genome optical mapping, the cutting edge technology offered for resolving the issues through estimating the gap length between the scaffolds and merges them into much longer sequences without introducing new bases (Ghurye and Pop, 2019; Kremer et al., 2017; Zhou et al., 2009; Aston et al., 1999; Samad et al., 1995). It also provides a valuable template for *de novo* genomic sequence assembly where large structural variations in the genome can accurately be detected and quantified (Long et al., 2018; Mak et al., 2016; Shukla et al., 2009; Teague et al., 2010). Furthermore, OM is capable of producing high-resolution, ordered, high-throughput genomic map data that gives information about the structure of a genome (Mukherjee et al., 2018; Schwartz et al., 1993). Though initially, it has been used to construct whole-genome restriction maps of bacteria, parasites, and fungi (Lai et al., 1999; Lim et al., 2001; Lin et al., 1999), recently it has been used for scaffolding contigs and for assembly validation of large-scale sequencing projects including maize, goat and *Amborella* genomes (Chamala et al., 2013; Udall and Dawe, 2018; Dong et al., 2013).

In Bangladesh, Jute (*Corchorus* species) is the most important cash crop considered the second foreign earning resources of the country (BBS, 2011). However, this crop is affected by several pathogens and diseases throughout its growing season and causing severe yield losses (Mamun et al., 2016). Among different agents *Macrophomina phaseolina* MS6, an ascomycetous, necrotrophic, soil-borne fungi, can solely reduce its yield up to 30% (Islam et al., 2012). This pathogen has more than 500 hosts (Lodha and Mawar, 2019; Khan et al., 2017; Islam et al., 2012) including major crops like cotton (Aly et al., 2007), jute (Meena et al., 2015; De et al., 1992), groundnut (Islam et al., 2012), maize (Biemond et al., 2013), sorghum (Su et al., 2001), millet (Lodha and Mawar, 2019), potato (Abbas et al., 2013), sesame (Dinakaran and Mohammed, 2001), soybean (Wyllie, 1993), beans (Mayek-Pérez et al., 2001), sunflower (Khan, 2007), sweet potato (Da Silva and Clark, 2013), tomato (Hyder et al., 2018), and tobacco (Wyllie, 1998). It outbreaks as stem rot (Majumder et al., 2018), seedling blight (Lu et al., 2015), charcoal rot (Majumder et al., 2018), dry root rot (Živanov et al., 2019), wilt (Piperkova et al., 2016), leaf blight (Mahadevakumar and Janardhana, 2016), pre and post-emergence damping-off (Hai et al., 2017), root and stem rot of softwood forest and fruit trees and also in weed species (Singh et al., 1990; McCain and Scharpf, 1989). This fungus forms microsclerotia in the soil and survives up to 15 years without attacking hosts (Kaur et al., 2012). It can also live in extreme environmental conditions like high temperature (30-35°C), low soil moisture, diverse pH, wide-ranging salt state and drought situation (Mengistu et al., 2011).

Considering all aforementioned consequences especially in jute, its genome was sequenced previously to have its mechanisms of attacking crops (Accession: AHHD00000000). But as described earlier, the bottlenecks that suffered the whole genome sequencing project, we addressed optical mapping to furnish and improve the genome. The furnished assembly, considered as valuable resources that might be used to develop a logical strategy for controlling the pathogen by unveiling the host-pathogen interaction within all sequenced crops that are infected by the pathogen.

MATERIALS AND METHODS

Preparation of working sample from *M. phaseolina* MS6

The strain *M. phaseolina* MS6 was taken from a stem rot infected jute plant (*Corchorus capsularis* L). The pathogen was cultured and purified on Potato Dextrose (PD) media maintaining 30°C for 72 h in dark conditions. The grey-brown mycelia were collected and washed with physiological buffer (Na_2HPO_4, pH 7.0: $NaH_2PO_4.H_2O$, pH <7.0) followed by drying under laminar flow Hood (Islam et al., 2012).

Extraction of megabase size DNA

Spheroplasting

Two grams of mycelia was ground into fine powder in liquid nitrogen with a mortar and pestle and immediately transferred into an ice-cold 1000 ml beaker containing 800 to 1000 ml ice-cold 1x Homogenize Buffer (HB) (0.1 M Tris, 0.8 M KCl, O.1 M EDTA, 10 mM Spermidine, 10 mM Spermine) with 0.15% beta-mercaptoethanol and 0.5% Triton X-l00. The contents were swirled gently for 10 minutes on ice and filtered by two layers of cheesecloth followed to one layer of Miracloth (Sigma-Aldrich, USA). The homogenate was taken into a centrifuge to have the pellet with a fixed-angle rotor at 1,800 g at 4°C for 20 minutes. The supernatant was discarded and approximately 1 ml of ice-cold Wash Buffer (WB) (1x HB, 20% TritonX-100, 0.15% beta-mercaptoethanol) was added to each tube. The pellet was resuspended gently with a small paintbrush soaked in ice-cold wash buffer. The nuclei were pelleted by centrifugation at 1,800 g at 4°C for 15 minutes in a swinging bucket centrifuge. The pellets were washed additional three times by resuspending in washing buffer using a paintbrush followed by centrifugation at 1,800 g at 4°C for 15 minutes. The pelleted nuclei were resuspended again in a small amount (1 ml) of 1x HB without beta-mercaptoethanol followed by counting the nuclei (approx. 5×10^7 nuclei/ml), under the contrast phase microscope with addition of the 1x HB without beta-mercaptoethanol and stored on ice.

Embedding cells

Low-melting-point (LMP) agarose (1%) was prepared in 1x HB without beta-mercaptoethanol and Triton X-100 followed by cooling down to 45°C and maintained in a 45°C water bath before use. The nuclei were pre-warmed to 45°C in a water bath (5 minutes) and mixed with an equal volume of the pre-warmed 1% LMP agarose in 1x HB without beta-mercaptoethanol and Triton X-100 using a cut-off pipette tip. The mixture was aliquoted into ice-cold plug molds on ice with the same pipette tip at 100 ml per plug.

Lysis

The solidified gel was sliced into pieces and incubated in 50 ml of digestion buffer (0.5 M EDTA, 7.5% β-mercaptoethanol) at 37°C for overnight. The buffer was replaced with NDSK buffer (0.5 M EDTA, 1% (w/v) N-lauroylsarcosine, 1 mg/ml proteinase K) for downstream work.

Plugs washing

The plugs were placed in a new 50 ml conical tube and added 45 ml of 1X TE buffer. The conical tube was capped with a clean green sieve and a regular cap and rock on a platform rocker at low speed for 1 h. 1X TE buffer was decanted from the conical tube and added fresh 1X TE buffer followed by rocking on a platform rocker at low speed for another 1 h. It was repeated for three times.

Melting plugs

The plugs were transferred into a sterile petri-dish and cut using a sterile scalpel in half and transferred to a separate 2.0 ml microcentrifuge tube. The microcentrifuge tubes were taken into a heat block at previously maintained 70°C for 7 minutes followed by pipetting 50 µl of the pre-warmed β-Agarose-TE solution (mixing 48 µl of 1X TE with 2 µl of 1 U/µl β-Agarose). The tube was incubated at 42°C heat block for overnight. Loading buffer was added to the DNA solution and stored at 4°C.

Restriction enzyme selection

The Enzyme.pl script (In-house script) was used to select optimal restriction endonuclease that generates restriction fragment statistics for different restriction enzymes. The optimal restriction enzyme was selected using this script by considering average fragment size (kb), fragment greater than 100 kb, maximum fragment size and the highest percentage of average fragment size underlie within 5 to 20 kb size fragments.

MapCard setup and data collection

The high molecular weight DNA was placed on optical chips to make them linear. This immobilized DNA was digested randomly with a restriction endonuclease and subsequently stained with jojo-1TM dye (Life Technologies Corporation, USA) for image capturing and fragment size measurement. The mapset (total data sets generated from all runs) was put together for assembly by using the Argus system embedded Gentig map assembler to create a consensus optical map. The mapsets were considered for assembly after filtering them following minimum molecule length (>150 kb) with minimum fragments per molecule (>12) and minimum molecule quality score (0.2).

Optical mapping assembly

The data from each MapCard were combined for the final assembly. In the assembly process, the filtered mapsets were taken for aligning them to form contigs by overlapping and keeping some extending fragments for resolving both side blunt ends (telomeric end protection). The final restriction maps were obtained by fulfilling criteria like coverage, depth, genome complexity of primary and final draft contigs. Different errors like low occurrence, low depth, and potential misassembles, as well as potential problems like false cut, missing cut, and false fragments were obviated using find hits techniques or removing the errors through Argus optical mapping embedded Gentig software packages. Assembly was conducted also with the removal of default circularization parameters. Partial assembly results were saved when 12 contigs became apparent by having >50 molecules each. Contigs were split off and reassembled against the original mapset individually using the "Find Hits" feature. Contigs were considered "finished" when no additional molecules were added by subsequent reassembles. Chromosome ends that were not blunt were visually inspected and any questionable molecule was removed from the final map (Figure 1).

Alignment by MapSolver

MapSolver software uses a dynamic programming algorithm to find the optimal placement location of each supplied sequence scaffolds in the Optical Map. The algorithm applies user-provided settings toward generating local alignments between each scaffold and the optical map. Scaffolds are aligned in both forward and reverse directions. MapSolver determines an alignment score for each comparison, where a higher score implies greater confidence in the alignment. Alignments with scores that meet or exceed the minimum score for local parameters are evaluated for placement. The number of aligned fragments must also meet or exceed the number specified by the Minimum aligned fragments parameter. In this study, default parameters were used.

RESULTS

Restriction endonuclease selection

A total of 13 restriction enzymes were evaluated to select the most efficient and suitable restriction endonuclease to consider different parameters including average and maximum size of DNA fragment and number of fragments. The effects of different restriction endonuclease are shown in Table 1. In the case of long size fragment, *Kpn*I enzyme showed the highest followed by *Nde*I and *Xba*I. However, none of the enzymes were found to produce more than 100 kb fragment size. Based on the rest parameters *Kpn*I was found as the most feasible and effective restriction enzyme to have the maximum MapSets for OM process (Table 1).

Optical maps construction and assembly

Based on the efficiency of the enzyme, HMW DNA was digested with *Kpn*I (New England Biolab, USA) restriction enzyme on the optical chips followed by subsequent dyeing that generated 71 GB raw data from 19 MapCards (Table 2). A total of 270,343 Single-Molecule Restriction Fragments (SMRMs) with an average size of 263.22 kb were produced from the optical chips analysis (Table 2). Within total molecules, 5,007,936 fragments were found having an average size of 14.209 kb. In addition to this, assembly of the molecules produced 12 unambiguous super-scaffolds (denoted as chromosomes) that are terminated with telomeric blunt ends (Figure 2). The

Table 1. Evaluation of restriction enzyme for OM compatibility of *Macrophomina phaseolina* MS 6 genome.

Enzyme	5-20 kb (%)	6-15 kb (%)	6-12 kb (%)	AFS (kb)	Frags>100 kb	Max. frag. size (kb)
*Af*1II	92.57	77.95	52.92	4.94	0	87.74
*Bam*H	21.92	3.87	3.87	4.77	0	39.95
***Kpn*I**	**98.69**	**79.79**	**74.42**	**7.80**	**0**	**63.25**
*Nco*I	0.00	0.00	0.00	3.23	0	31.09
*Nhe*I	87.55	55.88	55.88	5.29	0	64.05
*Spe*I	61.46	32.89	16.44	6.00	3	120.01
*Bg*1II	14.26	5.08	5.08	3.70	0	46.13
*EcoR*I	36.76	10.36	10.36	4.54	0	36.13
*Mlu*I	58.76	21.97	21.97	4.83	0	52.64
*Nde*I	95.10	70.03	68.78	7.55	0	55.90
*Pvu*II	0.00	0.00	0.00	2.41	0	31.09
*Xba*I	94.60	86.07	71.24	3.48	0	67.38
*Xho*I	0.00	0.00	0.00	2.70	0	25.67

Table 2. Optical map construction statistics.

Parameter	Maps
Number of molecules	270,343
Average molecule size (kb)	263.216
Number of fragments	5,007,936
Average fragment size (kb)	14.209
Total size (Mb)	71,158,564
Average quality score	0.517

chromosomes ranged from 1.6 to 6.7 Mb in sizes and spanned a total 49.723 Mb by joining all SMRMs through Argus^{TM} optical mapping system. It was also observed that optical mapping reduced the number of scaffolds from 94 to 88 where the largest scaffold increased by ~2 Mb in size. The indications in terms of contiguity along with quality and improvement, N50 placed on scaffold number 5 of OM instead of scaffold 6 of NGS. It increased by 4.25 Mb from 3.39 of NGS. Correspondingly, N90 also changed over its place on scaffold 11 instead of 14 by increasing size 2.9 Mb from 1.4. Although the N rate increased by 0.09%, still the GC content was unchanged (Table 3).

These results clearly pointed out the improvement of the assembly quality of *M. phaseolina* MS6 genome.

Alignment features between optical maps and reference maps

In alignment matrix, among all scaffolds only 17 were anchored on 11 chromosomes that spanned over 93.31% of the genome while none were on chromosome 12 (Figures 3 and 4). The sizes of aligned and non-aligned scaffolds were 46.35 and 3.3 Mb, respectively. Optical mapping deciphered 107 gaps and 4 overlaps size totaling 9.98 Mb and 26,040 kb, respectively, while 18% of the gaps can be closeable (Table 4).

A total number of 12 inversions (map is in reverse orientation) having 17.07 Mb in size were identified and made corrected for sequence finishing (Supplementary Files 1 and 2).

DISCUSSION

We have constructed whole-genome optical maps of M. phaseolina MS6 genome based on HMW DNA shearing by KpnI restriction endonuclease using OpGenTM optical mapping approach.. This technology is used to pick out the different types of incongruity between sequence generated *in silico* map and optical map along with current sequence validation of total spanned assembly. These issues were achieved by a series of action like-(i) alignment of optical map with *in silico* restriction map to find out all types of error, (ii) characterization of sequence contigs in respect to finding out the gaps, and (iii) the sequences were validated and placed on the optical map resulting in an explicit sequence validation. In the optical mapping process, restriction endonuclease is used to

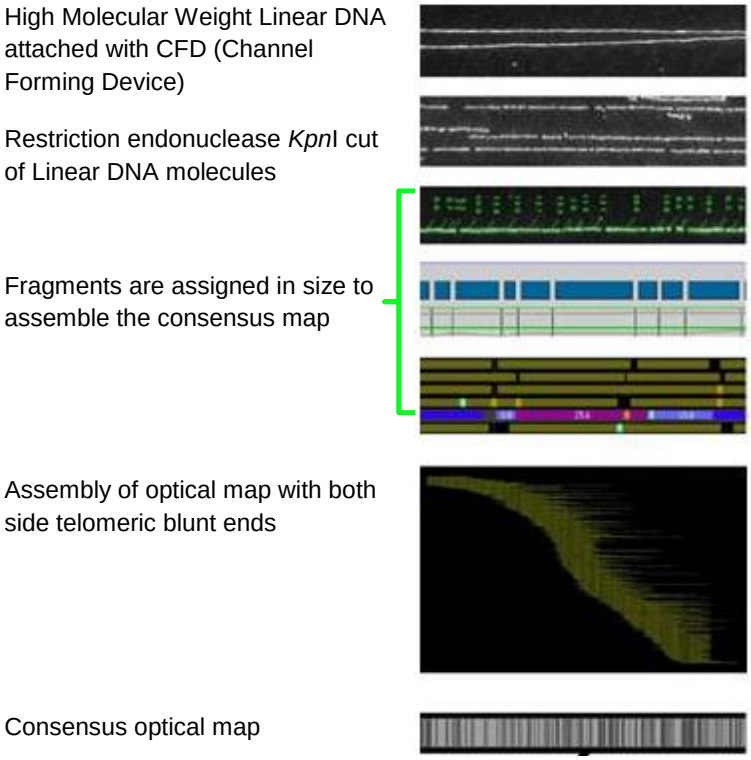

Figure 1. Generation of optical map from High Molecular Weight (HMW) DNA of *Macrophomina phaseolina* MS6.

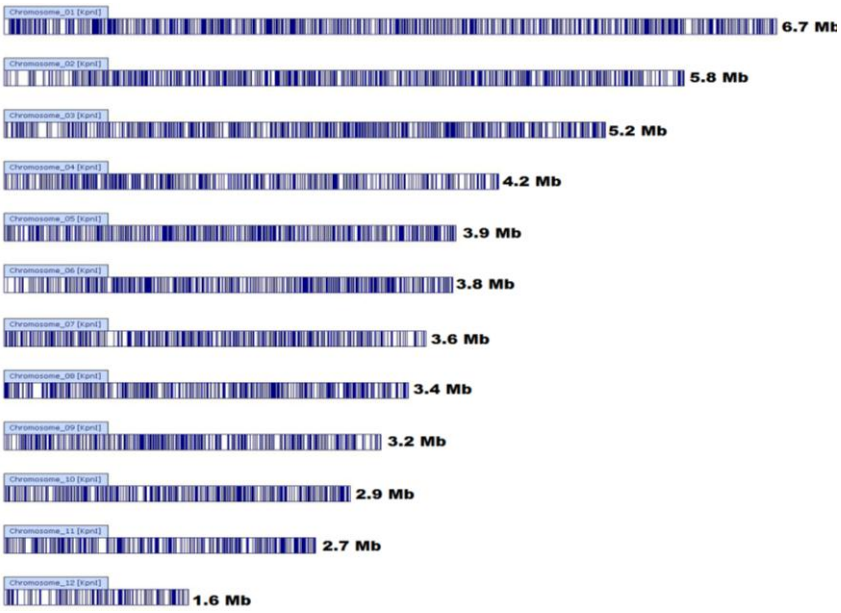

Figure 2. Optical maps of 12 chromosomes. Each vertical line represents the *Kpn*I restriction site. Chromosomes are ranked based on size.

digest total genomic DNA as a single molecule of >200 kb in size that was assembled into an ordered high-resolution restriction map possessing all fundamental genomic bases (Reslewic et al., 2005). This iterative and

Table 3. Assembly improvement statistics.

Parameter	Before (NGS)	After (OM)
No. of scaffolds	94	88
Largest scaffold (bp)	5,651,736	7,569,616
Bases in scaffolds (bp)	49,679,705	49,723,705
N50 scaffold	6	5
N50 length (bp)	3,400,455	4,259,500
N90 scaffold	14	11
N90 length (bp)	1,465,625	2,920,785
GC content	52.43%	52.43%
N rate	2.33%	2.42%

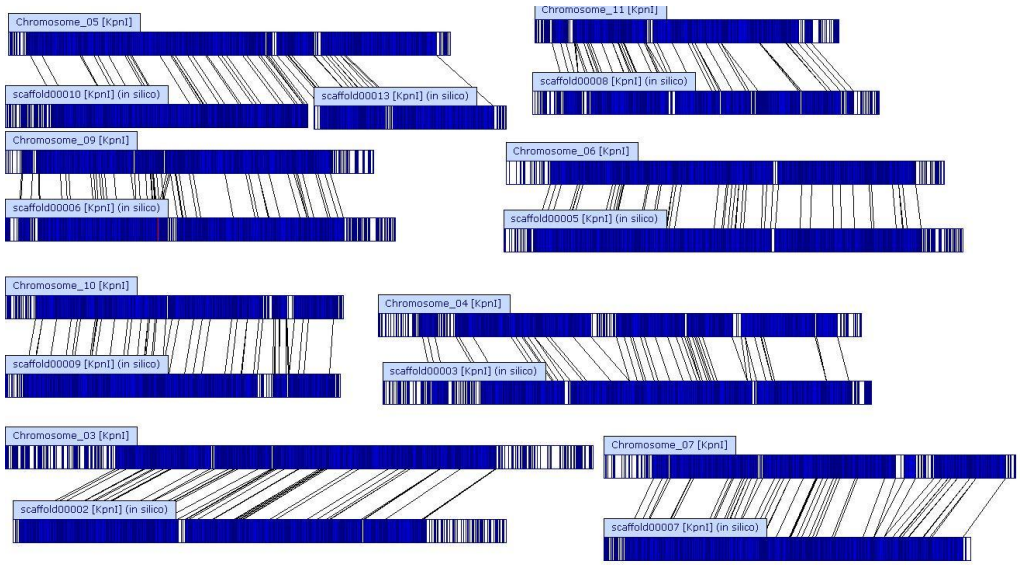

Figure 3. Alignment between optical maps and NGS sequence generated scaffolds. The blue-shaded regions of each map represent regions of the genomes that are similar whereas white areas are different. The alignment lines (lines connecting maps) connect regions of similarity from one map to the other.

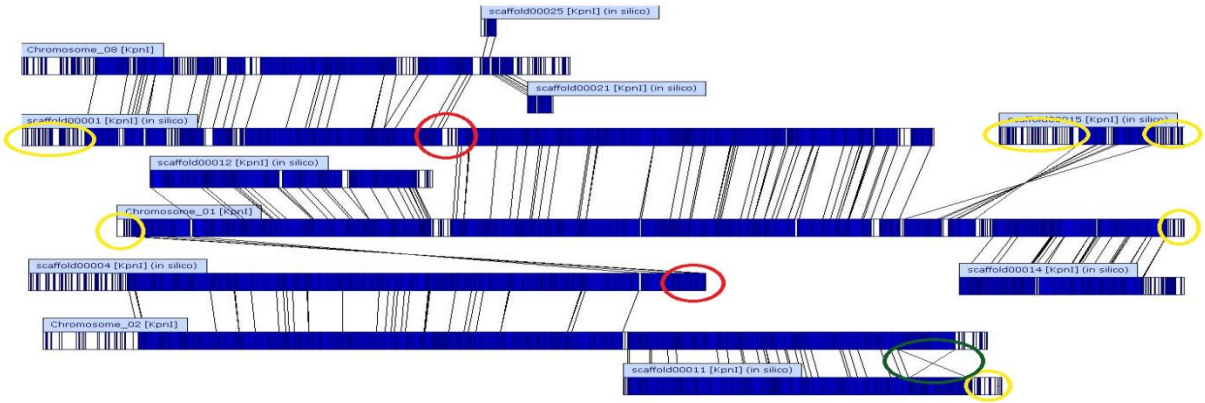

Figure 4. Alignment derived discordances of *de novo* chromosomes and NGS scaffolds. Regions encircled in red, crossed green and yellow indicates misassembled, inversion and low-quality assembly, respectively.

Table 4. Sequence alignment statistics.

Parameter	Alignment
Number of aligned scaffold	17
The total size of aligned scaffolds	46356277
Genome covered (%)	93.31
Number of unaligned scaffold	77
Unaligned sequences (%)	6.69
Total size of unaligned contigs	3323428
Number of total gaps	107
Number of gaps over 2 kb	91
Closeable gaps (%)	18
The total number of contig overlaps	4

computational assembly process joined all the SMRMs into super scaffolds by accomplishing sufficient representation across the chromosomes by coverage, the sufficient number of molecule maps covering each restriction fragment of the chromosome (depth >30X) and represents the genome complexity in terms of composition and structure (Ghurye and Pop, 2019). Finally, the optical mapping process generated 12 chromosomes terminating with both chromosomal telomeric blunt ends (blunt ends are not enzyme cut sites rather than the true end of a chromosome where the SMRMs ended at the same sequence). Furthermore, the telomeric end sequence (TTAGGG) of filamentous fungi within optical mapping organized ordered chromosomal sequence and found every chromosome possess their telomeric repetitive nucleotide sequence in between last SMRMs's (Average length 263 kb). The whole-genome restriction map consists of the chromosomes with dispersed arranged gaps. The similar results were observed in rice (Zhou et al., 2007) where the physical map consists of 14 contigs, covering its 12 chromosomes. In *Ganoderma lucidum* (Chen et al., 2012), 82 scaffolds were ordered and oriented onto 13 chromosome-wide optical maps that are very similar to our optical mapping results. The finished current NGS assembly is 49.295 Mb, that is, very close to our estimated 49.723 Mb *Kpn*I optical restriction map. The difference between the two assemblies was around ~1% which is identified as map error denoted as gaps, misassembles, and inversions. The improved and furnished assembly along with chromosome evolution mannered study of the fungus might be used for valuable resources along with fixation of control measures by biotechnological manner.

Conclusion

Here we presented an improved assembly of *M. phaseolina* MS6 genome with chromosomal level analyses using optical mapping data was presented. *In silico* analyzed 12 chromosomes with congruence and discordance makes the assembly error-free. The improved non-erroneous longer scaffold based chromosomal spanned assembly might be considered as the milestone to researchers for searching precisely its pestilential tools as well as survival dimensions in diverse environmental cues. The furnished assembly can also be used for future chromosomal re-arrangements and evolution studies in other fungi along with its control measures by developing the cross-talk phenomena between host and the pathogen.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

The authors thank Professor Dr. Wang Lei, Dr. Bin Liu and Dr. Yamin San of TEDA School of Biological Sciences and Biotechnology, Nankai University, China for their comprehensive help in the experiment. The research was funded by Basic and Applied Research on Jute Project of Bangladesh Jute Research Institute, Dhaka, Bangladesh.

REFERENCES

Abbas MF, Farah N, Gulshan I (2013). Important fungal diseases of potato and their management - a brief review. Mycopathology 11(1):45-50.

Aly AA, Abdel SMA, Omar MR, Abd-Elsalam KA (2007). Differential antagonism of *Trichoderma* sp. against *Macrophomina phaseolina*. Journal of Plant Protection Research 47(2):91-102.

Aston C, Mishra B, Schwartz DC (1999). Optical mapping and its potential for large-scale sequencing projects. Trends in Biotechnology 17(7):297-302.

Bangladesh Bureau of Statistics (2011). Government of the People's Republic of Bangladesh, Bangladesh Bureau of Statistics, Agriculture Wing, Parishankhyan Bhaban E-27/A, Agargaon, Dhaka-1207.

Biemond C, Oluwole O, Tjeerd JS, Kumar L, Aad T, Struik PC (2013). Health of farmer-saved maize seed in north-east Nigeria. European Journal of Plant Pathology 137(3):563-572.
Chamala S, Chanderbali AS, Der JP, Lan T, Walts B, Albert VA, dePamphilis CW, Leebens-Mack J, Rounsley S, Schuster SC, Wing RA, Xiao N, Moore R, Soltis PS, Soltis DE, Barbazuk WB (2013). Assembly and validation of the genome of the non-model basal angiosperm *Amborella*. Science 342(6165):1516-1517.
Chen S, Xu J, Liu C, Zhu Y, Nelson DR, Zhou S, Li C, Wang L, Guo XZ, Sun Y, Luo H, Li YR, Song J, Henrissat B, Levasseur A, Qian J, Li J, Luo X, Shi L, He L, Li X, Xu X, Niu Y, Li Q, Han MV, Yan H, Zhang J, Chen H, Lv A, Wang Z, Liu M, Schwartz DC, Sun C (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. Nature Communications 3(1):913.
Da Silva WL, Clark CA (2013). Infection of Sweet potato by *Fusarium solani* and *Macrophomina phaseolina* Prior to Harvest. Plant Disease 97(12):1636-1644.
Dinakaran D, Mohammed N (2001). Identification of resistant sources to root rot of sesame caused by *M. phaseolina* (Tassi.) Goid. Sesame and Safflower Newsletter 16:68-71.
De BK, Chattopadhya SB, Arjunan G (1992). Effect of potash on stem rot diseases of jute caused by *Macrophomina phaseolina*. Journal of Mycopathological Research 30(1):51-55.
Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, Gwenola TK, Jinhuan W, Shuang Y, Jie L, Wenbin C, Jing C, Peng Z, Yong H, Chao B, Shengkai P, Yuxiang L, Xin L, Wenliang W, Bertrand S, Brian S, Bin Z, Deacon S, Rich M, Wenhui N, Yongyi S, Ruoping Z, Guojie Z, Jinquan L, Thomas F, James W, Yaping Z, James K, Noelle C, Xun X, Shuhong Z, Jun W, Wen W, Wang W (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nature Biotechnology 31(2):135-141.
Ganapathy G, Howard JT, Ward JM, Li J, Li B, Li Y, Xiong Y, Zhang Y, Zhou S, Schwartz DC, Schatz M, Aboukhalil R, Fedrigo O, Bukovnik L, Wang T, Wray G, Rasolonjatovo I, Winer R, Knight JR, Koren S, Warren WC, Zhang G, Phillippy Jarvis AM (2014). High-coverage sequencing and annotated assemblies of the budgerigar genome. GigaScience 3:11 doi: 10.1186/2047-217X-3-11.
Ghurye J, Pop M (2019). Modern technologies and algorithms for scaffolding assembled genomes. PLOS Computational Biology 15(6):e1006994.
Hyder S, Gondal AS, Ahmed R, Sahi ST, Rehman A, Hannan A (2018). First Report of Charcoal Rot in Tomato Caused by *Macrophomina phaseolina* (Tassi) Goid From Pakistan. Plant Disease 102(7):1459.
Hai KMA, Ali AAE, El-Metwally MA (2017). Down-regulation of Damping-off and Root Rot Diseases in Lentil Using Kinetin and *Trichoderma*. International Journal of Agricultural Research 12(1):41-51.
Islam MS, Haque S, Islam MM, Emdad EM, Halim A, Hossen QM, Hossain MZ, Ahmed B, Rahim S, Rahman MS, Alam, MM, Hou S, Wan X, Saito JA, Alam M (2012). Tools to kill: Genome of one of the most destructive plant pathogenic fungi Macrophomina phaseolina. BMC Genomics 13(1):493.
Kaur S, Dhillon GS, Brar SK, Vallad GE, Chand R, Chauhan VB (2012). Emerging phytopathogen *Macrophomina phaseolina*: biology, economic importance, and current diagnostic trends. Critical Reviews in Microbiology 38(2):136-151.
Khan AN, Faluk S, Kamran M, Zafar H, Ayub KM, Yusuf HF, Nadeem HM (2017). Molecular Identification and Genetic Characterization of *Macrophomina phaseolina* Strains Causing Pathogenicity on Sunflower and Chickpea. Frontiers in Microbiology 8:1309
Khan SN (2007). *M. phaseolina* as causal agent for charcoal rot of sunflower. Mycopathologia 5(2):111-118.
Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis, ER (2013). The next-generation sequencing revolution and its impact on genomics. Cell 155(1):27-38.
Kremer FS, McBride A, Pinto LS (2017). Approaches for in silico finishing of microbial genome sequences. Genetics and Molecular Biology 40(3):553-576.
Lai Z, Jing J, Aston C, Clarke V, Apodaca J, Dimalanta ET, Carucci DJ, Gardner MJ, Mishra B, Anantharaman TS, Paxia S, Hoffman SL, Craig Venter J, Huff EJ, Schwartz DC (1999). A Shotgun Optical Map of the Entire *Plasmodium Falciparum* Genome. Nature Genetics 23(3):309-313.
Lim A, Dimalanta ET, Potamousis KD, Yen G, Apodoca J, Tao C, Schwartz DC (2001). Shotgun optical maps of the whole Escherichia coli O157:H7 genome. Genome Research 11(9):1584-1593.
Lin J, Lin J, Qi R, Aston C, Jing J, Anantharaman TS, Mishra B, White O, Daly MJ, Minton KW, Venter JC, Schwartz DC (1999). Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. Science 285(5433):1558-62.
Lodha S, Mawar R (2019). Population dynamics of *Macrophomina phaseolina* in relation to disease management: A review. Journal of Phytopathology 00:1-17.
Long E, Evans C, Chaston J, Udall JA (2018). Genomic Structural Variations within Five Continental Populations of *Drosophila melanogaster*. G3-Genes Genomes Genetics 8(10):3247-3253.
Lu C, Song B, Zhang H, Wang Y, Zheng X (2015). Rapid Diagnosis of Soybean Seedling Blight Caused by *Rhizoctonia solani* and Soybean Charcoal Rot Caused by *Macrophomina phaseolina* Using LAMP Assays. Phytopathology105 (12):1612-7.
Mak AC, Lai YY, Lam ET, Kwok TP, Leung AK, Poon A, Kwok PY(2016). Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. Genetics 202(1):351-362.
Majumder S, Datta K, Sarkar C, Saha SC, Datta SK (2018). The Development of *Macrophomina phaseolina* (Fungus) Resistant and Glufosinate (Herbicide) Tolerant Transgenic Jute. Frontiers in Plant Science 9:920
Mahadevakumar S, Janardhana GR (2016). First report of leaf blight caused by *Macrophomina phaseolina* on *Jasminium multiflorum* in india. Journal of Phytopathology 98(1):120-125.
Mamun MA, Shamsi S, Bashar MA (2016). Estimation of interrelationships among some quality factors of jute seeds. Dhaka University Journal of Biological Science 25(10):9-17.
Mayek-Pérez N, López-Castañeda C, López-Salinas E, Cumpián-Gutiérrez J, Acosta-Gallegos JA (2001): *Macrophomina phaseolina* resistance in common bean under field conditions in Mexico. Agrociencia 35:649-661.
McCain AH, Scharpf RF (1989). Effect of inoculum density of *Macrophomina phaseolina* on seedling susceptibility of six conifer species. European Journal of Forest Pathology 19(2):119-123.
Meena PN, Rajib D, Roy A, Bhimashankar G, Sabyasachi M (2015). Evaluation of stem rot disease in jute (*Corchorus olitorius*) germplasm caused by *Macrophomina phaseolina* (Tassi) Goid. Journal of Applied and Natural Science 7(2):857-859.
Mukherjee K, Washimkar D, Muggli MD, Salmela L, Boucher C (2018). Error correcting optical mapping data. Gigascience 7(3):61.
Mengistu A, Smith JR, Ray JD, Bellaloui N (2011). Seasonal progress of charcoal rot and its impact on soybean productivity. Plant Disease 95(9):1159-1166.
OpGen Inc. U.S. 708 Quince Orchard Road Gaithersburg, Maryland 20878 USA https://www.bloomberg.com/profile/company/OPGN:US
Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stütz AM (2015). Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nature Methods12(8):780-786.
Piperkova N, Zarkova M, Ahmed B (2016). Characterization of *Macrophomina Phaseolina* and *Fusarium* Spp. isolates from sunflower. Agricultural Sciences VIII (19):95-100.
Reslewic S, Zhou S, Place M, Zhang Y, Briska A, Goldstein S, Churas C, Runnheim R, Forrest D, Lim A, Lapidus A, Han CF, Roberts GP, Schwartz DC (2005). Whole-Genome Shotgun Optical Mapping of *Rhodospirillum rubrum*. Applied and Environmental Microbiology 71(9):5511-5522.
Ruperao P, Chan CK, Azam S, Karafiátová M, Hayashi S, Cížková J, Saxena RK, Simková H, Song C, Vrána J, Chitikineni A, Visendi P, Gaur PM, Millán T, Singh KB, Taran B, Wang J, Batley J, Doležel J, Varshney RK, Edwards D (2014). A chromosomal genomics approach to assess and validate the desi and Kabuli draft chickpea genome assemblies. Plant Biotechnology Journal 12(6):778-786.
Samad A, Huff EF, Cai W, Schwartz DC (1995). Optical mapping: a novel, single-molecule approach to genomic analysis. Genome Research 5(1):1-4.
Schwartz DC, Li X, Hernandez I, Ramnarian SP, Huff EJ, Wang YK (1993). Ordered restriction maps of *Saccharomyces cerevisiae*

chromosomes constructed by optical mapping. *Science* 262(5130):110-114.
Singh SK, Nene YL, Reddy MV (1990). Influence of cropping systems on *Macrophomina phaseolina* populations in soil. Plant Disease 74(10):812-814.
Shukla SK, Kislow J, Briska A, Henkhaus J, Dykes C (2009). Optical Mapping Reveals a Large Genetic Inversion between Two Methicillin-Resistant *Staphylococcus aureus* Strains. Journal of Bacteriology 191(18):5717-5723.
Su G, Suh SO, Schneider RW, Russin JS (2001). Host specialization in the charcoal rot fungus, *Macrophomina phaseolina*. Phytopathology 91(2):120-126.
Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM, Kohn S, Runnhei R, Lamers C, Forrest D, Newton MA, Eichler EE, Kent-First M, Surti U, Livny M, Schwartz DC (2010). High-resolution human genome structure by single-molecule analysis. Proceedings of the National Academy of Sciences 107(24):10848-10853.
Udall JA, Dawe RK (2018). Is It Ordered Correctly? Validating Genome Assemblies by Optical Mapping. The Plant Cell 30(1):7-14.
Wyllie TD (1998). Soybean Diseases of the North Central Region. In: Wyllie TD, Scott DH, editors. Charcoal rot of soybean-current status: American Phytopathological Society, St. Paul, pp. 106-113.
Wyllie TD (1993). Charcoal rot. Compendium of Soybean Diseases. Edited by: Sinclair JB, Backman PA., American Phytopathological Society, St. Paul, pp. 30-33.
Zhou S, Bechner M, Place, M, Churas C, Pape L, Leong SA, Runnheim R, Forrest DK, Goldstein S, Livny M, Schwartz DC (2007). Validation of rice genome sequence by optical mapping. BMC Genomics 15(8):278.
Zhou S, F. Wei J, Nguyen M, Bechner K, Potamousis S, Goldstein L, Pape MR, Mehan C, Churas S, Pasternak DK, Forrest R, Wise D, Ware RA, Wing MS, Waterman ML, Schwartz DC (2009). A single molecule scaffold for the maize genome. PloS Genetics 5(11):e1000711
Živanov D, Tančić Živanov S, Nagl N, Savić A, Katanski S, Milić D (2019). First Report of *Macrophomina phaseolina* on chickpea *(Cicer arietinum L.)* in Serbia. Plant Disease 103(10):2685.

SUPPLEMENTARY MATERIALS

File 1. Optical mapping gaps and overlaps

Optical Map	Type	Start	End	Length
Chromosome_01 [KpnI]	Gap	1	79358	79358
Chromosome_01 [KpnI]	Overlap	447422	460046	12625
Chromosome_01 [KpnI]	Gap	747528	752122	4595
Chromosome_01 [KpnI]	Gap	945314	950401	5088
Chromosome_01 [KpnI]	Gap	1290151	1292842	2692
Chromosome_01 [KpnI]	Gap	1923102	2049281	126180
Chromosome_01 [KpnI]	Gap	2538197	2541665	3469
Chromosome_01 [KpnI]	Gap	3196744	3213241	16498
Chromosome_01 [KpnI]	Gap	3657504	3660489	2986
Chromosome_01 [KpnI]	Gap	4010741	4012829	2089
Chromosome_01 [KpnI]	Gap	4150231	4179968	29738
Chromosome_01 [KpnI]	Gap	4634834	4677569	42736
Chromosome_01 [KpnI]	Gap	4795755	4818973	23219
Chromosome_01 [KpnI]	Gap	5050756	5094873	44118
Chromosome_01 [KpnI]	Gap	5258586	5362643	104058
Chromosome_01 [KpnI]	Gap	5726673	5732010	5338
Chromosome_01 [KpnI]	Gap	5996370	6005240	8871
Chromosome_01 [KpnI]	Gap	6434333	6581957	147625
Chromosome_02 [KpnI]	Gap	1	576503	576503
Chromosome_02 [KpnI]	Gap	3566337	3574343	8007
Chromosome_02 [KpnI]	Gap	4666099	4667788	1690
Chromosome_02 [KpnI]	Gap	4706528	4708564	2037
Chromosome_02 [KpnI]	Overlap	5236138	5240657	4520
Chromosome_02 [KpnI]	Gap	5598673	5790302	191630
Chromosome_03 [KpnI]	Gap	1	831165	831165
Chromosome_03 [KpnI]	Gap	1327249	1329064	1816
Chromosome_03 [KpnI]	Gap	1694143	1696119	1977
Chromosome_03 [KpnI]	Gap	2788280	2802124	13845
Chromosome_03 [KpnI]	Gap	3296156	3337159	41004
Chromosome_03 [KpnI]	Gap	3741990	3743910	1921
Chromosome_03 [KpnI]	Gap	3901113	3913498	12386
Chromosome_03 [KpnI]	Gap	4176741	5119287	942547
Chromosome_04 [KpnI]	Gap	1	387731	387731
Chromosome_04 [KpnI]	Gap	499599	678302	178704

File 1. Contd.

Chromosome_04	[KpnI]	Gap	1160359	1194969	34611
Chromosome_04	[KpnI]	Gap	1249413	1251114	1702
Chromosome_04	[KpnI]	Gap	1853551	2077590	224040
Chromosome_04	[KpnI]	Gap	2285405	2287395	1991
Chromosome_04	[KpnI]	Gap	2669156	2688624	19469
Chromosome_04	[KpnI]	Gap	3080799	3166130	85332
Chromosome_04	[KpnI]	Gap	3331794	3338912	7119
Chromosome_04	[KpnI]	Gap	3807654	3817821	10168
Chromosome_04	[KpnI]	Gap	4005328	4212496	207169
Chromosome_05	[KpnI]	Gap	1	113512	113512
Chromosome_05	[KpnI]	Gap	1126000	1189809	63810
Chromosome_05	[KpnI]	Gap	1528950	1562392	33443
Chromosome_05	[KpnI]	Gap	1601451	1610620	9170
Chromosome_05	[KpnI]	Gap	1769933	1772111	2179
Chromosome_05	[KpnI]	Gap	3211560	3228069	16510
Chromosome_05	[KpnI]	Gap	3664142	3849511	185370
Chromosome_06	[KpnI]	Gap	1	384283	384283
Chromosome_06	[KpnI]	Gap	639443	641244	1802
Chromosome_06	[KpnI]	Gap	2001087	2031642	30556
Chromosome_06	[KpnI]	Gap	2323247	2374952	51706
Chromosome_06	[KpnI]	Gap	3453531	3454828	1298
Chromosome_06	[KpnI]	Gap	3579468	3820145	240678
Chromosome_07	[KpnI]	Gap	1	72040	72040
Chromosome_07	[KpnI]	Gap	714355	874226	159872
Chromosome_07	[KpnI]	Gap	967769	1051336	83568
Chromosome_07	[KpnI]	Gap	1852132	1854424	2293
Chromosome_07	[KpnI]	Gap	2207992	2254218	46227
Chromosome_07	[KpnI]	Gap	3011557	3036031	24475
Chromosome_07	[KpnI]	Gap	3157829	3590256	432428
Chromosome_08	[KpnI]	Gap	1	543109	543109
Chromosome_08	[KpnI]	Gap	737294	793992	56699
Chromosome_08	[KpnI]	Overlap	904383	906268	1886
Chromosome_08	[KpnI]	Gap	1021335	1152669	131335
Chromosome_08	[KpnI]	Gap	1252737	1304792	52056
Chromosome_08	[KpnI]	Gap	1463069	1558040	94972
Chromosome_08	[KpnI]	Gap	2380415	2514290	133876

File 1. Contd.

Chromosome_08	[KpnI]	Gap	2847035	2906972	59938
Chromosome_08	[KpnI]	Gap	2964044	2971980	7937
Chromosome_08	[KpnI]	Overlap	3012351	3019359	7009
Chromosome_08	[KpnI]	Gap	3111589	3444310	332722
Chromosome_09	[KpnI]	Gap	1	125892	125892
Chromosome_09	[KpnI]	Gap	235466	283656	48191
Chromosome_09	[KpnI]	Gap	524756	526744	1989
Chromosome_09	[KpnI]	Gap	550524	552683	2160
Chromosome_09	[KpnI]	Gap	1097161	1127720	30560
Chromosome_09	[KpnI]	Gap	1219548	1220850	1303
Chromosome_09	[KpnI]	Gap	1271646	1287701	16056
Chromosome_09	[KpnI]	Gap	1361684	1408860	47177
Chromosome_09	[KpnI]	Gap	2094379	2096211	1833
Chromosome_09	[KpnI]	Gap	2515039	2531652	16614
Chromosome_09	[KpnI]	Gap	2703610	2705653	2044
Chromosome_09	[KpnI]	Gap	2724149	2725923	1775
Chromosome_09	[KpnI]	Gap	2839562	3209463	369902
Chromosome_10	[KpnI]	Gap	1	262536	262536
Chromosome_10	[KpnI]	Gap	632659	652206	19548
Chromosome_10	[KpnI]	Gap	795495	803422	7928
Chromosome_10	[KpnI]	Gap	1183685	1185635	1951
Chromosome_10	[KpnI]	Gap	1398164	1453469	55306
Chromosome_10	[KpnI]	Gap	1711265	1713480	2216
Chromosome_10	[KpnI]	Gap	2242393	2335381	92989
Chromosome_10	[KpnI]	Gap	2385276	2386923	1648
Chromosome_10	[KpnI]	Gap	2451049	2518758	67710
Chromosome_10	[KpnI]	Gap	2862713	2948255	85543
Chromosome_11	[KpnI]	Gap	1	234895	234895
Chromosome_11	[KpnI]	Gap	315293	352704	37412
Chromosome_11	[KpnI]	Gap	461982	469634	7653
Chromosome_11	[KpnI]	Gap	1260397	1282191	21795
Chromosome_11	[KpnI]	Gap	1618703	1682423	63721
Chromosome_11	[KpnI]	Gap	1704229	1748358	44130
Chromosome_11	[KpnI]	Gap	1872927	1874668	1742
Chromosome_11	[KpnI]	Gap	1897448	1899532	2085
Chromosome_11	[KpnI]	Gap	2062058	2067777	5720
Chromosome_11	[KpnI]	Gap	2186940	2200017	13078
Chromosome_11	[KpnI]	Gap	2223138	2224929	1792
Chromosome_11	[KpnI]	Gap	2307985	2377282	69298
Chromosome_11	[KpnI]	Gap	2503819	2652226	148408

File 2. Alignment information between optical maps and NGS sequence derived insilico maps

Chromosome	Start	End	Contig	Start	End	Orientation
Chromosome_01 [KpnI]	84599	447421	scaffold00004 [KpnI] (in silico)	3754787	4138122	-1
Chromosome_01 [KpnI]	460047	1919768	scaffold00012 [KpnI] (in silico)	6677	1630180	1
Chromosome_01 [KpnI]	2058751	4621849	scaffold00001 [KpnI] (in silico)	2739187	5428339	1
Chromosome_01 [KpnI]	4681571	4793752	scaffold00001 [KpnI] (in silico)	5520902	5639258	1
Chromosome_01 [KpnI]	4821388	5046981	scaffold00015 [KpnI] (in silico)	708429	946091	-1
Chromosome_01 [KpnI]	5096608	5252281	scaffold00015 [KpnI] (in silico)	488655	651614	-1
Chromosome_01 [KpnI]	5369160	6431276	scaffold00014 [KpnI] (in silico)	6517	1129657	1
Chromosome_02 [KpnI]	584978	3553334	scaffold00004 [KpnI] (in silico)	614339	3738177	1
Chromosome_02 [KpnI]	3583285	5236137	scaffold00011 [KpnI] (in silico)	23765	1746602	1
Chromosome_02 [KpnI]	5240658	5591437	scaffold00011 [KpnI] (in silico)	1754726	2110127	-1
Chromosome_03 [KpnI]	838070	4161621	scaffold00002 [KpnI] (in silico)	52587	3570054	-1
Chromosome_04 [KpnI]	389545	494988	scaffold00003 [KpnI] (in silico)	3722076	3833917	-1
Chromosome_04 [KpnI]	680721	4002755	scaffold00003 [KpnI] (in silico)	196387	3397557	-1
Chromosome_05 [KpnI]	121869	1514309	scaffold00013 [KpnI] (in silico)	51660	1560437	-1
Chromosome_05 [KpnI]	1563606	3660695	scaffold00010 [KpnI] (in silico)	443499	2633246	-1
Chromosome_06 [KpnI]	386252	3571691	scaffold00005 [KpnI] (in silico)	284719	3635033	1
Chromosome_07 [KpnI]	88203	3155720	scaffold00007 [KpnI] (in silico)	187125	3128052	-1
Chromosome_08 [KpnI]	544757	733814	scaffold00001 [KpnI] (in silico)	451658	650302	1
Chromosome_08 [KpnI]	796028	904382	scaffold00001 [KpnI] (in silico)	694057	808998	1
Chromosome_08 [KpnI]	906269	1016062	scaffold00001 [KpnI] (in silico)	825646	937173	1
Chromosome_08 [KpnI]	1157124	1250126	scaffold00001 [KpnI] (in silico)	1057655	1156833	1
Chromosome_08 [KpnI]	1344489	1455367	scaffold00001 [KpnI] (in silico)	1235932	1348893	1
Chromosome_08 [KpnI]	1560874	2844475	scaffold00001 [KpnI] (in silico)	1436060	2636783	1
Chromosome_08 [KpnI]	2909815	2955214	scaffold00025 [KpnI] (in silico)	37991	85335	1
Chromosome_08 [KpnI]	2971981	3012350	scaffold00021 [KpnI] (in silico)	1	42763	1
Chromosome_08 [KpnI]	3019360	3100473	scaffold00021 [KpnI] (in silico)	62711	144842	1
Chromosome_09 [KpnI]	139286	233651	scaffold00006 [KpnI] (in silico)	118062	217210	1
Chromosome_09 [KpnI]	287540	2833740	scaffold00006 [KpnI] (in silico)	296554	2946891	1
Chromosome_09 [KpnI]	1291111	1357055	scaffold00006 [KpnI] (in silico)	1321185	1392070	1
Chromosome_10 [KpnI]	264954	2857186	scaffold00009 [KpnI] (in silico)	174737	2825737	1
Chromosome_11 [KpnI]	245190	300565	scaffold00008 [KpnI] (in silico)	2739572	2797367	-1
Chromosome_11 [KpnI]	357185	2499966	scaffold00008 [KpnI] (in silico)	280656	2680529	-1