Mycologia Iranica 8(2):95–107, 2021 Original Article DOI: 10.22092/MI.2022.359536.1225

Status of SSR, cSSR, iSSR and VNTR motifs in *Leptosphaeria maculans* based on high throughput sequencing data

A. Zamanmirabadi
R. Hemmati✉
Department of Plant Protection, Faculty of Agriculture, University of Zanjan, Zanjan, Iran
A. Dolatabadian
Department of Plant Science, University of Manitoba, Winnipeg, MB R3T 2N2, Canada.
J. Batley
UWA School of Biological Sciences and the UWA Institute of Agriculture, Faculty of Science, The University of Western Australia, Crawley, WA, Australia

Abstract: *Leptosphaeria maculans* is a fungus of the phylum Ascomycota that is a causal agent of blackleg disease on canola (*Brassica napus* L.). Due to the high diversity and worldwide distribution, *L. maculans* has been widely studied as a model phytopathogenic fungus. Simple sequence repeats (SSRs) are robust molecular markers widely used for population diversity research. This study utilized whole-genome sequencing data of four Iranian *L. maculans* isolates (Pk4, Ar3, Ar5, and Alam10). We compared them with the JN3 reference genome to identify and compare different types of SSRs, including perfect (SSRs), compound (cSSR), imperfect (iSSR) and variable tandem number repeats (VNTR) motifs. The average length of SSRs was estimated to be 155.692 kb, accounting for 0.36% of the total genome. An average of 7138 SSR motifs with a frequency of one SSR per 169.5 bp, including an average of 33.86% tri, 25.69% di, 14.48% mono, 10.87% tetra, 8.52% hexa, 6.58% penta-nucleotide repeats, were identified from assembled genomic sequences. Of the total SSRs identified in the Pk4 isolate, 459 motifs were identified in CDS regions. Approximately 13% of the identified SSRs were linked to cSSRs. The average cSSR loci density for four isolates was 487.32 bp/Mb, and C, AG and AC were the most frequent SSR motifs. The assessed isolates' cSSRs lengths ranged from 24 to 295 bp. The largest common cSSRs in four isolates were identified as a motif $(GA)_{26}$-$(CAGAGA)_{15}$ with a length of 142 bp. The tri-nucleotide (AAT) was the most common iSSRs motif, followed by di, tetra, mono, hexa, and penta-nucleotides. About 30% of iSSRs contained the AAT, AT, and AAG motifs. Among the 7 to 30 nucleotide motifs, 7, 8, 9, and 10 motifs showed the most occurrences. In addition, 11 motifs with more than 100 nucleotides were found in the studied isolates and the reference genome. The data demonstrate that these results can be used to characterize *L. maculans* isolates from diverse hosts and geographic locations and are transferable to other isolates of *L. maculans*.

KEYWORDS: Blackleg, SSRs, *Leptosphaeria maculans*, NGS.

INTRODUCTION

Simple sequence repeats (SSRs) or microsatellites are short motifs of one to six base pairs. Wide coverage and high level of polymorphism of SSRs marker (Tautz 1989; Jarne and Lagoda 1996) has remained a potent tool for studying genetic diversity in biology (Robinson et al. 2004; Guichoux et al. 2011). SSRs are inherited as co-dominant markers in a Mendelian fashion (Robinson et al. 2004), which are more abundant in coding regions in eukaryotes (Metzgar et al. 2000). Microsatellites are also one of the most preferred markers for population genetics investigations due to their widespread distribution throughout the genome and high polymorphism rates (Duran et al. 2009). They are classified as pure (perfect), interrupted (imperfect), and compound (complex and interrupted complex) (Chambers and MacAvoy 2000). Perfect SSRs are tandem repetitions of a single motif without interruptions. If one or more base pairs not belonging to the motif occurs between repeats, it is categorized as imperfect SSRs. Compound SSRs composed of two or more individual microsatellites adjacent to each other (Riju 2009). All mentioned microsatellites have been reported in diverse taxa eukaryotes and prokaryotes *(Weber 1990; Bull et al. 1999; Kofler et al. 2008).* Identifying SSRs markers across the fungal genome has been challenging (Dutech et al. 2007a). However, several SSRs markers have been successfully adopted in fungal species, including *Macrophomina phaseolina* (Jana et al. 2005), *Ascochyta rabiei* (Bayraktar et al. 2007), *Puccinia triticina* (Szabo and Kolmer 2007), *Sclerotinia subarctica* (Winton et al. 2007), *Ceratocystis fimbriata* (Rizatto et al. 2009), *Colletotrichum gloeosporioides* (Marulanda et al. 2014), *Botryosphaeria dothidea* (Manawasinghe et al. 2018) and *Leptosphaeria maculans* (Hayden et

Submitted 15 Aug 2021, accepted for publication 18 Dec 2021
✉ Corresponding Author: E-mail: rhemati@znu.ac.ir

https://mij.areeo.ac.ir

al. 2007; Raman et al. 2012). Formerly, the development of SSRs markers was expensive and time-consuming compared to the other markers (Taheri et al. 2018). However, these SSRs sequences that can be used to study genetic diversity have been made available by reducing the cost of DNA sequencing through complete genome sequencing (Singh et al. 2015). Recently, whole genome sequencing as a high-throughput and cost-effective method has been widely used to study microsatellites through a genome-wide diversity of some fungal species like *Colletotrichum gloeosporioides* (Moges et al. 2016) *Phytophthora cinnamomi* (Engelbrecht et al. 2017), *Tilletia indica* (Gurjar et al. 2019) *Didymella pisi* (Owati et al. 2019), *Ustilago hordei* (Kashyap et al. 2020). Typically, the genomes of fungi have a great deal of diversity in terms of abundance and density (Srivastava et al. 2019). However, we still know little about the many kinds of satellites found in fungi, particularly *L. maculans* species, which contain several pathogenic genes (*Avr*) (Stachowiak et al. 2006; Liban et al. 2016; Neik et al. 2017). *L. maculans* causes significant damage to global canola production (Howlett 2004; Fitt et al. 2006). *L. maculans* is present in rapeseed fields in Iran (Zamanmirabadi et al. 2008). Our knowledge of the genetic alterations in the *L. maculans* population in Iran and their involvement in evolutionary processes is limited. Only one research in Iran used housekeeping genes such as *B-tubulin*, *tef*, and *act* gene sections as molecular markers to identify *L. maculans* (Zamanmirabadi et al. 2021, 2022). Given the importance of SSRs in the study of genomic data, it is necessary to introduce the diversity of these markers as a powerful tool for genome analysis. In this study, we refer to introduce the different types of SSRs, including: pure or perfect (SSR), compound SSR(cSSRs), imperfect SSR (iSSR) and VNTR microsatellites and to get the help of genome NGS data of four isolates of *L. maculans*, in the north of Iran comparing the reference genome of *L. maculans* (strain v23.1.3) (Rouxel et al. 2011).

MATERIAL AND METHODS

Fungal sample, isolation, and DNA extraction

The whole-genome sequence of four *L. maculans* isolates collected from the northern regions of Iran and the JN3 (23-1-3) reference genome https://www.genoscope.cns.fr/leptolife/datasets.html) were used in this study (Table 1).

Tissue samples of canola showing blackleg disease symptoms were collected from different northern regions of Iran. Briefly, leaves showing blackleg disease symptoms were surface sterilized by dipping into 10% sodium hypochlorite for 1 min, rinsed in sterile distilled water for 2 min, soaked in 70% ethanol for 30 s, and washed again with sterile distilled water for another 2 min. The disinfected sections were blot dried on sterile filter paper and placed on Potato Dextrose Agar (PDA) (Zamanmirabadi et al. 2022). We isolated 91 isolates of the fungus. The pathogenicity tests showed that only four isolates were placed in the pathogenic group of *L. maculans*. The rest were identified as a non-pathogenic group and *L. biglobosa*; therefore, only these four isolates were used in this study.

A growing mycelium disc of these pathogenic isolates (5-mm diameter) was cultured in a flask containing Potato Carrot Broth (PCB) (40 ml) for ten days at 25±2°C to grow enough fungal tissue for DNA extraction. The total harvested mycelia were frozen and ground to a fine powder in liquid nitrogen, and the total DNA was extracted according to the CTAB method (Doyle and Doyle 1990). The quantity and quality of the extracted DNA were determined with a Qubit® 3.0 Fluorometer (Thermo Fisher Scientific, Q33216) and LabChip GX Touch Nucleic Acid Analyzer (PerkinElmer, CLS138162), respectively.

Table 1. Sampling regions of *Leptosphaeria maculans* in Mazandaran province in 2018

Isolate	Province	City/Region	Longitude	Latitude	Sampling date
Pk4	Mazandaran	Kiasar road	53.09186	36.46446	24 Feb 2018
AR3	Mazandaran	Sari/Arabkhail	53.04467	36.69189	23 Feb 2018
AR5	Mazandaran	Sari/Arabkhail	53.04467	36.69189	23 Feb 2018
Alam10	Mazandaran	Alamdardeh	53.25651	36.36606	24 Feb 2018

Library preparation and whole genome sequencing data

Whole-genome sequencing libraries of four *L. maculans* isolates (Pk4, Ar3, Ar5 and Alam10) were constructed using the Nextra DNA Flex Library Prep kit, October 2018 (Illumina, San Diego, CA, USA). Five stages, including tagment of genomic DNA, post tagmentation clean up, amplifying tagmented DNA, libraries clean up and finally, libraries pooling, were performed according to the manufacturer's protocol. Briefly, 100 to 250 ng DNA taken from each isolate was used to construct the DNA library. The final concentrations of the libraries were measured using the Qubit double-stranded DNA (dsDNA). The library was sequenced in paired-end format under the Illumina Hiseq2500

platform (AGRF, Queensland, Australia). Quality control, trimming of sequence data, and mapping reads to reference (JN3) were completed using CLC Genomics Workbench version 20.0.0 (CLC bio, Aarhus, Denmark).

Microsatellites identification

The assembled sequences of whole-genome sequences of four *L. maculans* isolates and references genome (JN3) (Rouxel et al. 2011) were screened for the different types of SSR motifs using Krait software (Du et al. 2018) which is a robust and ultrafast software for genome-wide investigation of microsatellites. The different types of repeated motifs are presented in Table 2.

The minimum repeats for each SSR type were set to 12, 7, 5, 4, 4, and 4 for Mono, di, tri, tetra, penta and hexa- nucleotides, respectively, and the motif standardization level was set to Level 3 and at least four tandem copies for four bp to six bp nucleotide motifs. The maximum distance between adjacent SSRs to account as compounded was ten bp. The minimum seed length and seed repeats for iSSRs were set to 8 and 3, respectively. The maximum consecutive edits (including substitutions and InDels) allowed were specified as 3. The penalty cost was set to 1 for mismatch and 2 for InDels (gaps). The minimum required score for identified iSSR was set to 10, and the motif standardisation level was set to Level 3. The minimum and maximum motif lengths for VNTR were set to 7 and 30, respectively. The minimum repeats were set to 2.

Table 2. The different types of microsatellites include SSRs, cSSRs, iSSRs, and VNTRs

Microsatellite types	Motifs	Microsatellite types	Motifs
Perfect SSRs	$(A)_6$	Compound SSRs	$(T)_4\ (A)_4$
	$(AG)_6$		$(CT)_3\ (GA)_3$
	$(AGT)_5$		$(CTG)_2\ (AGC)_2$
Imperfect SSRs	$(C)_3\ T(C)_3$	VNTRs	$(AGCCTTG)_2$
	$(AG)_3\ CCC\ (AG)_4$		$(AGGGGCTG)_3$
	$(AGG)_2\ CCC\ (AGG)_3$		

RESULTS

Genome Sequencing

A summary of whole genome sequencing data statistics of Pk4, Ar3, Ar5 and Alam10 isolates is presented in Table 3. An average of 97% of reads were mapped to the JN3 reference genome for the four isolates chosen for microsatellite extraction. The average mapping coverage and GC% content for various isolates ranged from X27 to X42 and 45-47%, respectively.

SSRs detection

The characteristics associated with SSRs, including the number, length, average length, distribution, frequency, and relative density of the sequences of four isolates compared to the reference genome, are presented separately for each isolate in Table 4 and Table S1. Compared to the reference genome, the studied genomes showed no significant differences in the relevant data, except for a few changes in the frequency and density of these sequences. The average length and percentage of SSRs in all genomes were 21.85 bp and 0.35%, respectively. Either reference genome or *L. maculans* isolates contained all known types of SSRs. The Tri, di, mono, tetra, hexa, and penta-nucleotide motifs showed the highest abundance in the genome with an average of 33.9, 25.70, 14.52, 10.89, 8.48, and 6.56 percent, respectively (Fig. 1 and Fig. S2). There was no significant difference between the isolates regarding motif type percentage. The most common motifs with a high frequency percentage were C/A, AC/AG, ACG/ACC, AAAG/AACG, AAAAG/AAAAC and AAAAAG/AAACAC.

In this investigation, even though motifs with a frequency of ≥20bp were more common than motifs with a frequency of 20bp. Motifs with a frequency of more than 20bp in the reference genome were observed more than other study isolates. (Fig. 2).

Out of the total SSRs identified in the Pk4 isolate (Table 4), approximately 459 types of SSRs were identified in the CDS regions, of which, 20 motifs with the highest frequency percentage in the gene regions are shown in Fig. 3. Some motifs with their protein codes in UniProt (www. https://www.uniprot.org) are including E4ZVB0,E4ZZA6,E5AD95,E4ZFR9,E4ZG26,E4ZH16,E4ZHJ0,E4ZI48,E4ZI81,E4ZJ95,E4ZJN2,E4ZJT3,E4ZK76,E4ZLH8,E4ZMP2,E4ZMR9,E4ZN07,E4ZNC4,E4ZNI4,E4ZPA8,E4ZPK8,E4ZPX6,E4ZPZ6,E4ZQK3,E4ZS10,E4ZTU1,E4ZU55,E4ZU88,E4ZUM0,E4ZV99,E4ZWL3,E4ZXC6,E4ZXN1,E4ZXS3,E4ZY73,E4ZZ82,E4ZZI3,E5A0R7,E5A129,E5A192,E5A1H6,E5A2V7,E5A321,E5A8C0,E5A8Y1,E5AAE3,E5AC31,E5ACK8,E5ADC9,E5R5E8,.

Table 3. Resequencing assembly statistics of four isolates of *Leptosphaeria maculans*

No	Isolate	Sample	Total Read	Total PE reads	Read Length (bp)	Insert Size median (bp)	Total # bp	Coverage	%GC	Genome size estimation
13	Pk4	lepto13	5,118,460	10,236,920	150	321	772,887,460	34.53	47	39.74
22	Ar3	lepto22	6,346,472	12,692,944	150	321	958,317,272	42.72	44	46.21
71	Ar5	lepto71	5,056,729	10,113,458	150	317	763,566,079	33.93	45	43.83
94	Alam10	lepto94	4,059,351	8,118,702	150	302	612,962,001	27.51	45	46.19

Table 4. The summary information of perfect microsatellites in Pk4, Ar3, Ar5, Alam10 and Reference genome

Item	PK4	Ar3	Ar5	Alam10	JN3	Average
Total number of perfect SSRs (Counts)	7125	7123	7100	7164	7177	7138
The total length of perfect SSRs (bp)	155669	155394	155381	156326	156871	155928
The average length of SSRs (bp)	21.85	21.82	21.89	21.83	21.86	21.85
SSRs per sequence (Counts)	94	94	93	94	94	94
The percentage of sequences covered by SSRs (%)	0.39	0.33	0.35	0.33	0.35	0.35
Relative abundance of SSRs (loci/Mb)	171	168	168	171	163	168
Relative density (bp/Mb)	3741	3664	3675	3731	3566	3675

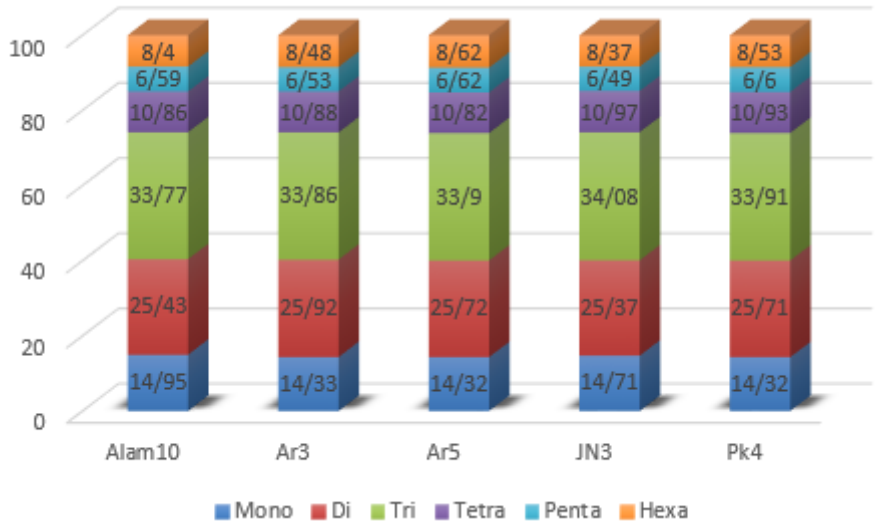

Fig. 1. Frequency percentage of SSRs types in four blackleg isolates of this study and the reference genome

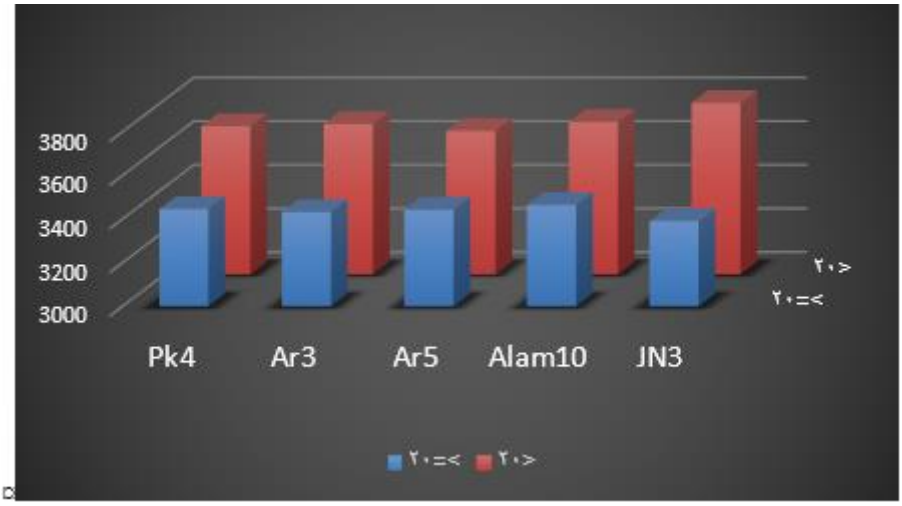

Fig. 2. Frequency of motifs ≥ 20 bp and <20 bp in the Pk4, Ar3, Ar5, and Alam10 isolates, as well as the reference genome

cSSRs detection

The cSSR sequence counts, frequency percentages, and shortest and longest motif lengths are shown in Tables 5 and 6, respectively. About 13 percent of the total SSRs were related to cSSRs. Motifs containing at least five nucleotides are detailed in Table 7. Motifs containing 120 nucleotides are presented in Fig. 4. Only three motifs out of more than 120 nucleotides, including $(GA)_{26}$-$(CAGAGA)_{15}$, $(CA)_8$-$(CA)_{8}$-$(CA)_8$-$(CA)_8$-$(CA)_8$-$(CA)_8$-$(CA)_8$-$(CA)_8$ and $(ACT)_6$-$(TAC)_{22}$-$(ACT)_6$-$(ACT)_6$ were shared among the isolates. Most cSSR (≈ 87%) contain only two cSSRs. It founded the largest compound microsatellite as $(AG)_8$-$(AG)_7$-$(AG)_9$-$(AG)_9$-$(AG)_{12}$-$(AG)_7$-$(AG)_9$-$(AG)_9$-$(AG)_{10}$ in *Ar3 isolate*, having nine cSSRs. In the whole genome of all five isolates, the most abundant cSSR-couple motifs were CA/CT and AC/TC.

Thirty-eight cSSR motifs were found in the gene regions from the total number of identified cSSRs. The majority of these cSSRs were contained $(ACA)_n$ and $(CCA)_n$, $(AG)_n$, (GA)n, $(AGTGC)_n$, $(GACT)_n$, $(ATG)_n$, $(ACG)_n$, $(CA)_n$, $(AC)_n$, $(CGA)_n$, $(GAC)_n$, $(CT)_n$, $(TC)_n$, $(TGT)_n$ and $(TTC)_n$.

ISSRs detection

The total number, length, and relative abundance of iSSRs of four isolates and reference genome are shown in Table 7.

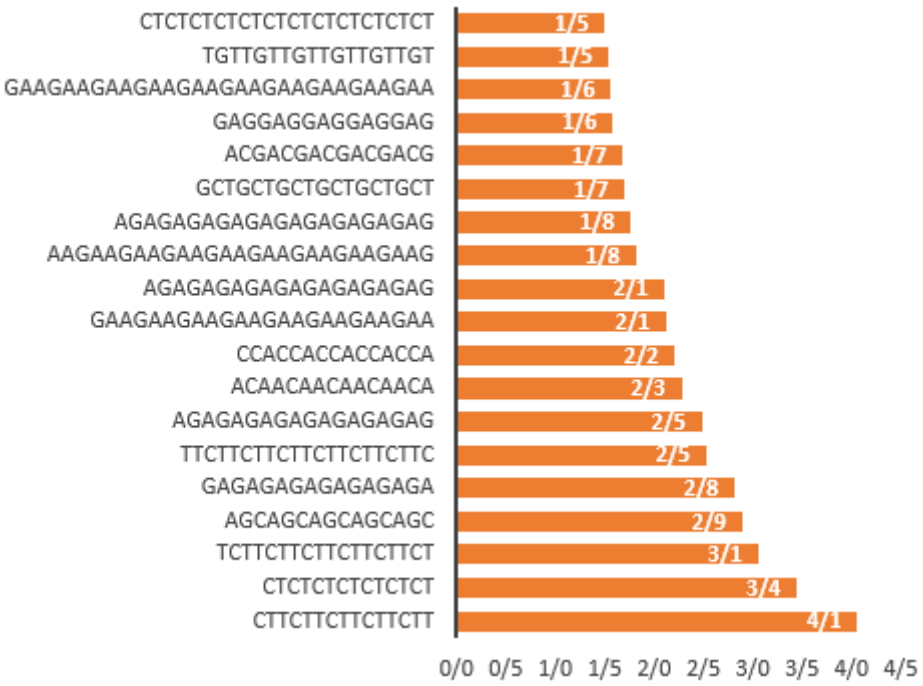

Fig 3. Twenty motifs with a high frequency in the Pk4 genome's CDS region

Table 4. The summary information of compound cSSRs in four *Leptosphaeria maculans* isolates and reference genome

Description	Pk4	Ar3	Ar5	Alam10	JN3	Average
Total number of compounds SSRs (Counts)	419	431	414	428	400	418
Total number of cSSRs(individual SSRs being part of a compound SSRs)	915	944	907	946	884	919
Percentage of cSSRs(total cSSR counts/total SSR counts (%))	12.85	13.26	12.78	13.21	12.32	13
Relative abundance (total CM counts/total valid length (loci/Mb))	10.07	10.16	9.79	10.22	9.09	10
Relative density (total CM length/total valid length (bp/Mb)	487.4	490.45	475.4	496.06	447.93	479

Table 5. The shortest and longest cSSR motifs in Pk4, Ar3, Ar5 and Alam10 isolates and the reference genome

Motifs	Max length	Motifs	Min length	Gap No	Total cSSr	Isolate
(A)105-(A)47-(A)25-(A)112	289	$(CA)_{21}$-$(CA)_{10}$ $(CCA)_6$-$(CTA)_6$	25	205	419	**Pk4**
(A)106-(A)47-(A)25-(A)113	291	$(A)_{12}$-$(A)_{12}$	24	197	431	**Ar3**
(A)104-(A)72-(A)118	294	$(C)_{13}$-$(A)_{12}$ $(A)_{13}$-$(A)_{12}$	25	209	414	**Ar5**
(A)96-(A)47-(A)25-(A)127	295	$(A)_{12}$-$(A)_{12}$	24	196	428	**Alam10**
(A)106-(A)47-(A)25-(A)379	557	$(G)_{12}$-$(T)_{12}$	24	201	400	**JN3**

Table 6. Unique motifs containing a minimum of five nucleotides in *Leptosphaeria maculans* isolates

Isolates	Motifs
Pk4	$(AGTGC)_6$-$(GACT)_4$; (GGATG)4-$(GATGG)_8$-$(ATGGG)_6$ $(GTT)_5$-$(GCTGTT)_8$
Ar3	$(GAGGTG)_6$-$(TGGAGG)_5$;(GT)7-$(GCT)_6$-$(CGAGA)_4$ $(CAACAG)_{11}$-$(CAACAG)_{11;}$ $(TCGTCT)_4$-$(TTC)_5$ $(GGGAGA)_5$-$(GGAAGA)_6$;$(ACACCC)_4$-$(AC)_8$
Ar5	$(GGAAGG)_4$-$(AGGGGA)_5$; $(AAGGGA)_4$-$(AAGGGA)_4$ $(TCATCT)_5$-$(ATCTTC)_4$;$(GGGAGA)_5$-$(GGAAGA)_8$- $(AGAGGA)_{10}$;$(GTCCTC)_4$-$(TCCTCG)_4$
Alam10	$(GGAAGG)_4$-$(AGGGGA)_5$; $(CCATCT)_5$-$(C)_{18}$ $(GGATG)_4$-$(GATGG)_6$; $(GGCT)_5$-$(GGTAG)_7$ $(TCT)_8$-$(TCT)_{18}$-$(TCGTCT)_4$-$(TCT)_5$ $(CCTAAC)_7$-$(CCTAAC)_4$
JN3	$(GGGAGA)_5$-$(GGAAGA)_5$-$(GGGAGA)_4$-$(GGAAGA)_{10}$ $(TCGCAA)_4$-$(CGCAGT)_4$;$(GTCCTC)_5$-$(CTC)_6$ $(TACTGC)_5$-$(TGC)_7$; $(TCT)_{16}$-$(TCGTCT)_4$-$(TCT)_5$ $(CATTAT)_6$-$(TCATTA)_9$; $(CACTCT)_4$-$(TCTT)_4$ $(CCTTCT)_5$-$(TCCTTC)_5$-$(TCCTTC)_4$; $(GTT)_5$-$(GCTGTT)_6$-$(GTGGTT)_5$; $(TTTTC)_{12}$-$(TTCTT)_9$

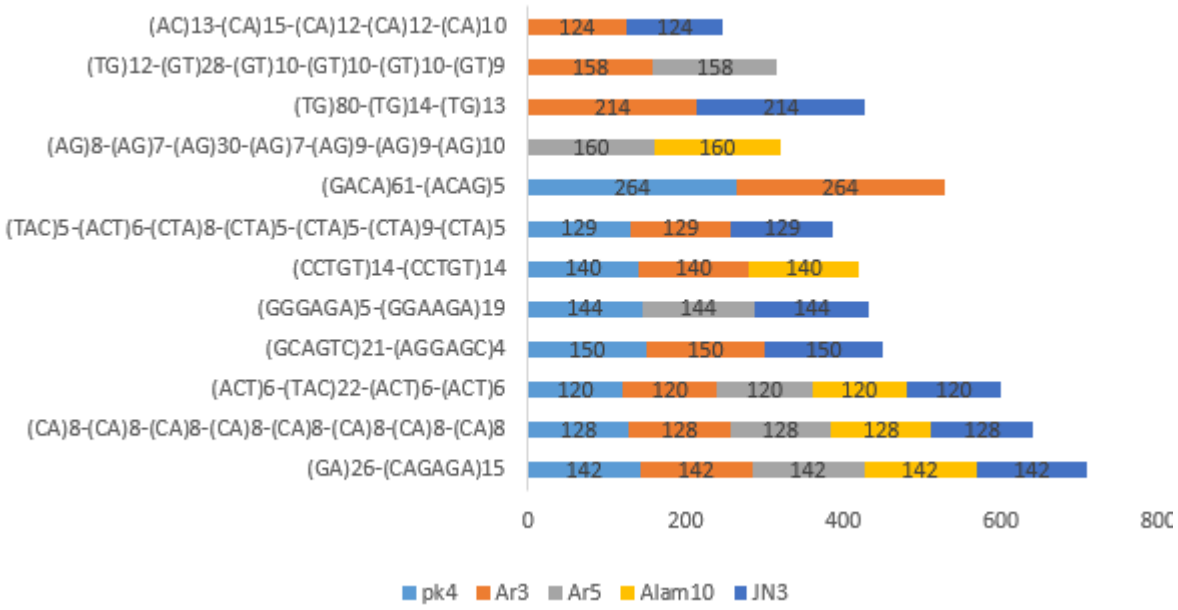

Fig. 4. cSSR motif structure with a length of more than 120 bp and a high similarity between isolates

The isolate Ar3 showed the highest frequency of iSSRs among the isolates, but not as much as the reference genome. The frequency percentage of different iSSRs is presented in Fig. 5. According to the result, most frequencies were related to the tri, di, tetra, mono, hexa, and penta motifs. A summary of the number, density, and length of iSSRs for each isolate and reference genome is given in Table 8. AAT, AT, and AAG motifs accounted for approximately 29% of the total number of motifs in the genome of all isolates.

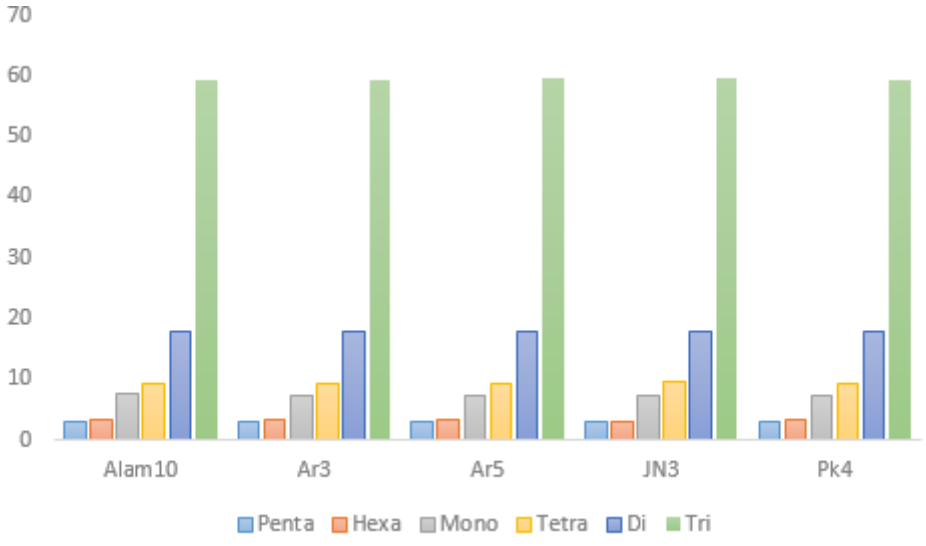

Fig. 5. iSSR type frequency percentage in the reference genome and Iranian *Leptosphaeria. maculans* isolates

VNTRs detection

In this study, the abundance of motifs with more than six nucleotides is shown in Fig. 6. Hepta, Octa, nona, and deca-nucleotide motifs showed the greatest number of occurrences among the 7-to-30 nucleotide motifs. Also, 11 motifs with more than 100 nucleotides were found either in the isolates or the reference genome, whereas most were found in the reference genome. There was no similarity between these motifs (Table 9).

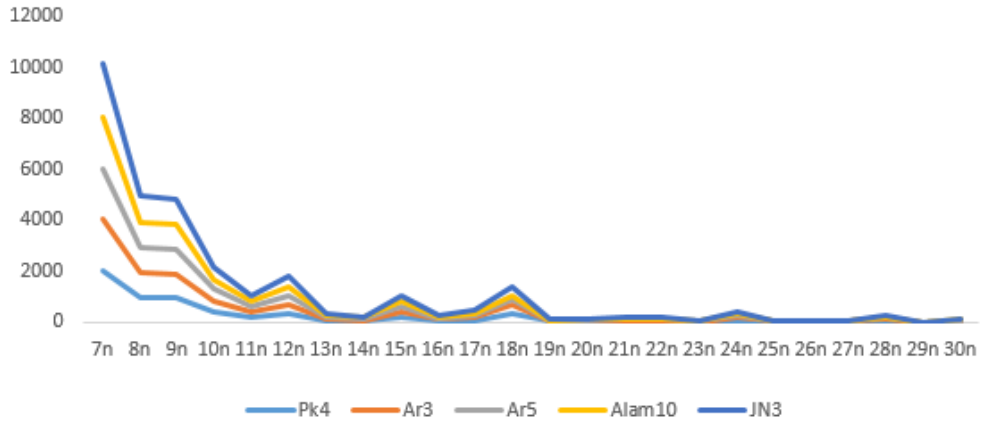

Fig. 6. The frequency of VNTRs motifs in different isolates of *Leptosphaeria maculans*

Table 7. The number, length, frequency, and density of iSSRs in four *Leptosphaeria maculans* isolates and the reference genome

Relative density (bp/Mb)	Relative abundance (loci/Mb)	Average length (bp)	Length (bp)	Counts	Type	Isolate
5750.67	209.74	27.42	240917	8787	Di	
1734.39	37.79	45.9	72660	1583	Hexa	
2697.35	88.72	30.4	113002	3717	Mono	**Alam10**
1278.54	34.59	36.97	53563	1449	Penta	
3425.4	109.28	31.35	143503	4578	Tetra	
21093.56	698.91	30.18	883688	29280	Tri	
5732.41	208.88	27.44	243146	8860	Di	
1737.6	37.72	46.06	73702	1600	Hexa	
2627.99	86.17	30.5	111469	3655	Mono	**Ar3**
1266.01	34.52	36.68	53699	1464	Penta	
3471.45	110.41	31.44	147245	4683	Tetra	
21192.26	701.1	30.23	898891	29738	Tri	
5729.23	208.5	27.48	242222	8815	Di	
1734.18	37.68	46.03	73318	1593	Hexa	
1275.34	34.72	36.73	53919	1468	Penta	**Ar5**
3434.11	109.91	31.24	145188	4647	Tetra	
21124.57	700.64	30.15	893110	29622	Tri	
5782.75	211.48	27.34	254414	9304	Di	
1696.79	36.48	46.51	74651	1605	Hexa	
2532.93	85.17	29.74	111437	3747	Mono	
1249.04	34.12	36.61	54952	1501	Penta	
3510.94	112.03	31.34	154465	4929	Tetra	
21346.32	706.23	30.23	939139	31071	Tri	**JN3**
1763.24	38.11	46.26	73376	1586	Hexa	
2621.1	86.58	30.27	109075	3603	Mono	
1285.74	34.6	37.16	53505	1440	Penta	
3456.05	109.48	31.57	143821	4556	Tetra	
21108.43	698.05	30.24	878412	29049	Tri	

Table 8. Specific VNTRs motifs of more than 100 nucleotides in the *Leptosphaeria maculans* isolates

102	GCGCGTTCTATAACTTT	Pk4
119	AAGTTATAGAACGCGCA	Ar3
135	TAGCAAAGAGGA	Ar5
120	TGTTTAGAATA	Alam10
120	AACGAGGCTGGAGAGGAAGCTCAG	
120	ACATCGTCACTGCACACACA	
102	ATTGTAGCCGTGTGAGT	
184	CTGTAGGTATACCATTGCGGTGC	JN3
224	GCAGGAAACGGGTTTTGGGTCTGGGTTT	
168	GCGTCAACAGCGTGGACAAGGCAGGCAG	
108	GGTTCTTCTGGTCTTGGT	

DISCUSSION

Much research has not been done to discover different forms of SSRs in fungal populations. (Dutech et al. 2007b). Furthermore, according to our findings, no equivalent research has been conducted to introduce distinct types of SSrs into the *L. maculans* genome. Using whole genome sequencing technology, this study offers the first comprehensive study of SSRs, cSSR, iSSRs, and VNTRs microsatellites in four L. maculans genomes. A large amount of data provides an excellent resource to investigate the diversity and polymorphism of *L. maculans* isolates. This diversity in SSRs makes them suitable for population studies due to their mutation rate (from 10^{-4} to 10^{-3} mutations per base per) (Kelkar et al. 2008). Based on the findings of this research and other studies (Srivastava et al. 2019a), SSRs had the highest amount compared with other types of microsatellites in the genome, and most of these sequences were observed in the non-coding region(Luo et al. 2015).

In this study, the average length of SSRs for the isolates was 0.35% of the genomic sequence. At the same time, this rate in *Ustilago maydis*, *Magnaporthe grisea* and *Neurospora crassa* was reported 0.79%, 0.82%, and 0.95% of the genomic sequence, respectively (Li et al. 2009). In this study, 7138 SSRs were identified. This number of identified SSRs was more than the other species such as *Puccinia striiformis* (4792) (Luo et al. 2015), *Aspergillus nidulans* (2410), *Fusarium graminearum* (2896), *U. maydis* (3033) and *Saccharomyces cerevisiae* (3618)) and less than some other species such as *M. grisea* (11642) and *N. crassa* (14319) (Karaoglu et al. 2005). The longest SSR in the *L. maculans* isolates was AAAT motif 48 with 192 nucleotides. This was observed in *Trichoderma reesei*, *Trichoderma virens* and *Cryptococcus neoformans* species, although with less repetition of the AAAT motif (Murat et al. 2011). The average density of SSRs in these isolates was one SSR per 3.6 kb sequence. This density is equivalent to that described in fungi (Li et al. 2009) and plants (Katti et al. 2001).

The cSSRs are less than SSRs but are more stable because they are less affected by mutations ((Brandstrom and Ellegren 2008). According to this research, the total percentage of SSRs trinucleotide motifs was higher than other motifs, but the highest frequency was related to the C mononucleotide motif, which confirms the results of this study in terms of the high amount of mononucleotides in eukaryotes ((Sharma, P.C. et al. 2007; Liu et al. 2017). Most motifs found in SSRs and iSSRs were tri-nucleotides, whereas dinucleotides were more prevalent in cSSRs. In this study and many other fungi, including; *U. maydis* and *F. graminearum*, the ACG trinucleotide was the most common tri-nucleotide (Srivastava et al. 2019). SSRs are divided into two categories based on their length: class I, which are ≥ 20 bp repeats, while Class II SSRs contained less than 20 bp (Wang et al. 2018). Motifs with more than 20bp are more likely to exhibit hyper-polymorphism (Temnykh et al. 2001). In this study, the number of SSRs motifs ≥20 nucleotides exceeded the number of motifs with fewer than 20 nucleotides. Given the high diversity of SSRs in different regions of the genome, including the expression regions in this study and previous studies regarding the role of these regions in influencing biological functions (Bagshaw 2017), it would be suggested to validate these data in *Leptosphaeria* population using local and global isolates.

DATA AVAILABILITY

The data used in this study were deposited at NCBI and can be accessed with the following accession numbers. BioProject: PRJNA878667 and BioSample: SAMN30747262 (Pk4),

SAMN30800508 (Ar3), SAMN30800932 (Ar5), SAMN30800934 (Alam10).

REFERENCES

Bayraktar, H., Sara Dolar, F., and Tör, M. 2007. Determination of genetic diversity within Ascochyta Rabiei (Pass.) Labr., the cause of Ascochyta blight of Chickpea in Turkey Determinatıon Of Fungi That Cause Cluster and Fruıt Diseases On Hazelnut in Ordu Giresun and Trabzon Provinces and Reactions of C. JSTOR. [Online] Available: https://www.researchgate.net/publication/24275 8035 [2020 Mar. 30].

Bull, L.N., Pabón-Peña, C.R., and Freimer, N.. 1999. Compound Microsatellite Repeats: Practical and Theoretical Features. Genome Res. 9: 830-838.

Chambers, G.K., and MacAvoy, E.S. 2000. Microsatellites: consensus and controversy. Comp Biochem Physiol B Biochem Mol Biol. 126: 455–76.

Doyle, J.J., and Doyle, J.L. 1990. A rapid total DNA preparation procedure for fresh plant tissue. Focus (Madison). 12: 13–15.

Du, L., Zhang, C., Liu, Q., Zhang, X., and Yue, B. 2018. Krait: an ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics 34: 681–683.

Duran, C., Appleby, N., Edwards, D., and Batley, J. 2009. Molecular Genetic Markers: Discovery, Applications, Data Storage and Visualisation. Curr. Bioinform. 4: 16–27. doi:10.2174/157489309787158198.

Dutech, C., Enjalbert, J., Fournier, E., Delmotte, F., Barrès, B., Carlier, J., Tharreau, D., and Giraud, T. 2007a. Challenges of microsatellite isolation in fungi. Fungal Genet. Biol. 44: 933–949. doi:10.1016/j.fgb.2007.05.003.

Dutech, C., Enjalbert, J., Fournier, E., Delmotte, F., Barrès, B., Carlier, J., Tharreau, D., Giraud, T., Biology, F.D.-… G. and, 2007, undefined, Delmotte, F., Barrès, B., Carlier, J., Tharreau, D., and Giraud, T. 2007b. Challenges of microsatellite isolation in fungi. Fungal Genet. Biol. 44: 933–949.

Engelbrecht, J., Duong, T.A., Reports, N. vd B.-S., 2017, U., and Berg, N. V.D. 2017. New microsatellite markers for population studies of *Phytophthora cinnamomi*, an important global pathogen. Sci. Rep. 7: 1–10. Springer US. doi:10.1038/s41598-017-17799-9.

Fitt, B.D.L.L., Brun, H., Barbetti, M.J., and Rimmer, S.R. 2006. Worldwide importance of phoma stem canker (*Leptosphaeria maculans* and *L. biglobosa*) on oilseed Rape (*Brassica napus*). Eur. J. Plant Pathol. 114: 3–15. doi:10.1007/s10658-005-2233-5.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., and Petit, R.J. 2011. Current trends in microsatellite genotyping. Molecular Ecology Resources. 4: 591–611.

Gurjar, M.S., Aggarwal, R., Jogawat, A., Kulshreshtha, D., Sharma, S., Solanke, A.U., Dubey, H., and Jain, R.K. 2019. De novo genome sequencing and secretome analysis of *Tilletia indica* inciting Karnal bunt of wheat provides pathogenesis-related genes. 3 Biotech 9: 1–11. Springer International Publishing. doi:10.1007/s13205-019-1743-3.

Hayden, H.L., Cozijnsen, A.J., and Howlett, B.J. 2007. Microsatellite and Minisatellite Analysis of *Leptosphaeria maculans* in Australia Reveals Regional Genetic Differentiation. Phytopathology 97: 879–887. Scientific Societies. doi:10.1094/phyto-97-7-0879.

Howlett, B.J. 2004. Current knowledge of the interaction between *Brassica napus* and *Leptosphaeria maculans*. Can. J. Plant Pathol. 26: 245–252.

Jana, T., Sharma, T.R., and Singh, N.K. 2005. {SSR}-based detection of genetic variability in the charcoal root rot pathogen Macrophomina phaseolina. Mycol. Res. 109: 81–86. Elsevier {BV}. doi:10.1017/s0953756204001364.

Jarne, P., and Lagoda, P.J.L. 1996. Microsatellites, from molecules to populations and back. Trends Ecol. Evol. 11: 424–429.

Karaoglu, H., Lee, C.M.Y., and Meyer, W. 2005. Survey of simple sequence repeats in completed fungal genomes. Mol. Biol. Evol. 22: 639–649. doi:10.1093/molbev/msi057.

Kashyap, P.L., Kumar, S., Kumar, R.S., Tripathi, R., Sharma, P., Sharma, A., Jasrotia, P., and Singh, G.P. 2020. Identification of Novel Microsatellite Markers to Assess the Population Structure and Genetic Differentiation of Ustilago hordei Causing Covered Smut of Barley. Front. Microbiol. 10. Frontiers Media S.A. doi:10.3389/fmicb.2019.02929.

Katti, M. V, Ranjekar, P.K., and Gupta, V.S. 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. 18: 1161–7. Society for Molecular Biology and Evolution. doi:10.1093/oxfordjournals.molbev.a003903.

Kofler, R., Schlötterer, C., Luschützky, E., and Lelley, T. 2008. Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. BMC Genomics 9: 1–14. doi:10.1186/1471-2164-9-612.

Li, C.Y., Liu, L., Yang, J., Li, J.B., Su, Y., Zhang, Y., Wang, Y.Y., and Zhu, Y.Y. 2009. Genome-wide analysis of microsatellite sequence in seven filamentous fungi. Interdiscip. Sci. Comput. Life Sci. 1: 141-150.

Liban, S.H., Cross, D.J., Kutcher, H.R., Peng, G.,

and Fernando, W.G.D. 2016. Race structure and frequency of avirulence genes in the western Canadian *Leptosphaeria maculans* pathogen population, the causal agent of blackleg in brassica species. Plant Pathol. 65: 1161–1169. Wiley.

Luo, H., Wang, X., Zhan, G., Wei, G., Zhou, X., Zhao, J., Huang, L., and Kang, Z. 2015. Genome-wide analysis of simple sequence repeats and efficient development of polymorphic SSR markers based on whole genome resequencing of multiple isolates of the wheat stripe rust fungus. PLoS One 10.

Manawasinghe, I.S., Zhang, W., Li, X., Zhao, W., Chethana, K.W.T., Xu, J., Chen, Z., Dissanayaka, A.J., Mugnai, L., Úrbez-Torres, J.R., Savocchia, S., Hyde, K.D., and Yan, J. 2018. Novel microsatellite markers reveal multiple origins of *Botryosphaeria dothidea* causing the Chinese grapevine trunk disease. Fungal Ecol. 33: 134–142. Elsevier Ltd.

Marulanda, M.L., López, A.M., Isaza, L., and López, P. 2014. Microsatellite isolation and characterisation for *Colletotrichum* spp, causal agent of anthracnose in Andean blackberry. Genet. Mol. Res. 13: 7673–7685. Genetics and Molecular Research.

Metzgar, D., Bytof, J., and Wills, C. 2000. Selection against frameshift mutations limits microsatellite expansion in coding DNA. Genome Res. 10: 72–80.

Moges, A.D., Admassu, B., Belew, D., Yesuf, M., Njuguna, J., Kyalo, M., and Ghimire, S.R. 2016. Development of microsatellite markers and analysis of genetic diversity and population structure of *colletotrichum gloeosporioides* from Ethiopia. PLoS One 11: 1–18. Public Library of Science. doi:10.1371/journal.pone.0151257.

Murat, C., Riccioni, C., Belfiori, B., Cichocki, N., Labbé, J., Morin, E., Tisserant, E., Paolocci, F., Rubini, A., and Martin, F. 2011. Distribution and localisation of microsatellites in the Perigord black truffle genome and identification of new molecular markers. Fungal Genet. Biol. 48: 592–601.

Neik, T.X., Barbetti, M.J., Batley, J., Fernandez-Aparicio, M., An, H., Peng, G., Bruce, C., Gossen, D., Batley, J., Neik, T.X., Barbetti, M.J., Batley, J., Fernandez-Aparicio, M., An, H., Peng, G., Bruce, C., Gossen, D., Batley, J., Neik, T.X., Barbetti, M.J., and Batley, J. 2017. Current status and challenges in identifying disease resistance genes in *Brassica napus*. Front. Plant Sci. 8: 1788. Frontiers Media S.A.

Owati, A., Agindotan, B., Burrows, M., Biology, M.B.-F., and 2019, U. 2019. First microsatellite markers developed and applied for the genetic diversity study and population structure of *Didymella pisi* associated with ascochyta blight of dry pea in Montana. Fungal Biol. 123: 384–392. Elsevier Ltd.

Raman, R., Taylor, B., Marcroft, S., Stiller, J., Eckermann, P., Coombes, N., Rehman, A., Lindbeck, K., Luckett, D., Wratten, N., Batley, J., Edwards, D., Wang, X., and Raman, H. 2012. Molecular mapping of qualitative and quantitative loci for resistance to *Leptosphaeria maculans* causing blackleg disease in canola (*Brassica napus* L.). Theor. Appl. Genet. 125: 405–418. Springer Science and Business Media {LLC}.

Riju, A. 2009. SSR - Identification from EST. Nat. Preced. doi:10.1038/npre.2009.3908.1.

Rizatto, S., de Araújo Batista, C.E., Bajay, M.M., Sigrist, M.S., Ito, M.F., Monteiro, M., Cavallari, M.M., Pinheiro, J.B., and Zucchi, M.I. 2009. A new set of microsatellite markers for the genetic characterisation of Ceratocystis fimbriata, an economically important plant pathogen. Conserv. Genet. Resour. 2: 55–58. Springer Science and Business Media {LLC}.

Robinson, A.J., Love, C.G., Batley, J., Barker, G., and Edwards, D. 2004. Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics 20: 1475–1476.

Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., Van De Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., Linglin, J., Kema, G. H. J., Lapalu, N., Lawrence, C. B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C. L., Simon, A., Spatafora, J. W., Stachowiak, A., Turgeon, B.G., Tyler, B. M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R. P., Wincker, P., Balesdent, M.-H., and Howlett, B.J. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun.

Rouxel, T., Grandaubert, J., Hane, J.K., Hoede, C., Van De Wouw, A.P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A.J., Ciuffetti, L.M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S.B., Gout, L., Glaser, N., Linglin, J., Kema, G.H.J., Lapalu, N., Lawrence, C.B., May, K., Meyer, M., Ollivier, B., Poulain, J., Schoch, C.L., Simon, A., Spatafora, J.W., Stachowiak, A., Turgeon, B.G., Tyler, B.M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R.P., Wincker, P., Balesdent, M.H., and Howlett, B.J. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. Nat. Commun. 2: 202. Nature Publishing Group.

Singh, A.K., Singh, P.K., Arya, M., Singh, N.K., and Singh, U.S. 2015. Molecular screening of blast

resistance genes in Rice using SSR markers. Plant Pathol. J. 31: 12–24. Korean Society of Plant Pathology.

Srivastava, S., Avvaru, A.K., Sowpati, D.T., and Mishra, R.K. 2019. Patterns of microsatellite distribution across eukaryotic genomes. BMC Genomics 20. BioMed Central Ltd.

Stachowiak, A., Olechnowicz, J., Jedryczka, M., Rouxel, T., Balesdent, M.-H.H., Happstadius, I., Gladders, P., Latunde-Dada, A., and Evans, N. 2006. Frequency of avirulence alleles in field populations of *Leptosphaeria maculans* in Europe. Eur. J. Plant Pathol. 114: 67–75. Springer Science and Business Media {LLC}.

Szabo, L.J., and Kolmer, J.A. 2007. Development of simple sequence repeat markers for the plant pathogenic rust fungus *Puccinia triticina*. Mol. Ecol. Notes 7: 708–710. Wiley.

Taheri, S., Abdullah, T.L., Yusop, M.R., Hanafi, M.M., Sahebi, M., Azizi, P., and Shamshiri, R.R. 2018. Mining and development of novel SSR markers using Next Generation Sequencing (NGS) data in plants. Molecules 23.

Tautz, D. 1989. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. 17: 6463–6471.

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cart inhour, S., and McCouch, S. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res. 11: 1441-1452.

Wang, X., Yang, S., Chen, Y., Zhang, S., Zhao, Q., Li, M., Gao, Y., Yang, L., and Bennetzen, J.L. 2018. Comparative genome-wide characterisation leading to simple sequence repeat marker development for Nicotiana. BMC Genomics 19: 1–12.

Weber, J.L. 1990. Informativeness of human (dC-dA)n · (dG-dT)n polymorphisms. Genomics 7: 524–530. Academic Press.

Winton, L.M., Krohn, A.L., and Leiner, R.H. 2007. Microsatellite markers for *Sclerotinia subarctica* nom. prov., a new vegetable pathogen of the High North. Mol. Ecol. Notes 7: 1077–1079. Wiley.

Zamanmirabadi, A., Hemmati, R., Dolatabadian, A., and Batley, J. 2021. Current progress in studying blackleg disease (*Leptosphaeria maculans* and *L . biglobosa*) of canola in Iran : Where do we stand now ? Plant Pathol.: 1–12.

Zamanmirabadi, A., Hemmati, R., Dolatabadian, A., and Batley, J. 2022. Genetic structure and phylogenetic relationships of *Leptosphaeria maculans* and *L. biglobosa* in Northern regions of Iran. Arch. Phytopathol. Plant Prot.: 1-20.

Zamanmirabadi, A., Rahnama, K., Sadravi, M., and Mehdi Alamdarlou, R.M. 2008. First report of *Leptosphaeria maculans* teleomorph on canola stem in the north of Iran (short article). Rostaniha 9: 128–130. [Online] Available: http://en.journals.sid.ir/ViewPaper.aspx?ID=122787 [2018 Jan. 12].

تعیین وضعیت توالی‌های SSR، cSSR، iSSR و VNTR در *Leptosphaeria maculans* بر پایه داده‌های توالی یابی کامل ژنوم

علی زمان میرآبادی[1]، رقیه همتی [1]✉ ،آریا دولت آبادیان[2]، ژاکلین بتلی[3]

۱-گروه گیاهپزشکی دانشگاه زنجان، زنجان،ایران

۲-گروه علوم گیاهی دانشگاه مانیتوبا، وینیپگ،کانادا

۳-مدرسه علوم زیستی دانشگاه UWA، استرالیای غربی، استرالیا

چکیده: گونه *Leptosphaeria maulans* متعلق به شاخه آسکومیکوتا است که باعث بیمای ساق سیاه در کلزا (*Brassica napus*) می شود. به دلیل تنوع بالا و پراکنش جهانی *L. maculans*، این قارچ به طور وسیع در مطالعات بیماری زایی گیاهی به عنوان یک مدل استفاده می شود. توالی های ساده تکراری (SSRs)، از نشانگرهای قوی مولکولی هستند که در تحقیقات مربوط به تنوع جمعیتی استفاده می گردند. در این مطالعه با بهره گیری از داده های توالی یابی ژنوم، در چهار جدایه ایرانی از این قارچ (Pk4,Ar3,Ar5, Alam10) و همچنین ژنوم مرجع JN3، انواع مختلف توالی های ساده تکراری شامل SSR، cSSR، iSSR و VNTR شناسایی و بررسی گردیدند. بر این اساس میانگین طول SSRs در حدود 155kb و در حدود ۰,۳۶٪ کل ژنوم محاسبه گردید. به طور میانگین حدود ۷۱۳۸ موتیف SSR با یک فروانی هر SSR در 169.5bp با یک فراوانی ۳۳,۸۶٪ برای تری ، ۲۵,۶۹٪ برای دی ، ۱۴,۴۸٪ برای مونو، ۱۰,۸۷٪ برای تترا ، ۸,۵۲٪ برای هگزا و ۶,۵۸٪ برای پنتا نوکلئوتید شناسایی شد. از مجموع SSRs شناسایی شده در جدایه Pk4، ۴۵۹ موتیف در ناحیه CDS شناسایی گردید. تقریبا ۱۳ ٪ از کل توالیهای تکراری مربوط به cSSRs بودند. میانگین تراکم cSSR در چهار جدایه 487.32bp/Mb محاسبه گردید و موتیف C، AG و AC بیشترین فراوانی را در cSSRs داشتند. طول cSSRs بین ۲۴ تا 295bp مشاهده گردید. بزرگترین موتیفcSSR در چهار جدایه $(GA)_{26}$-$(CAGAGA)_{15}$ با 142bp شناسایی شد. AAT فراوانترین موتیف iSSRs و در ادامه دی، تترا، مونو، هگزا و پنتا نوکلئوتید بیشترین فراوانی را داشتند. در حدود ۳۰٪ iSSRs محتوی AAT، AT و AAG بودند. در میان موتیف های ۷ تا ۳۰ نوکلئوتیدی، موتیف های هفت، هشت، نه و ۱۰ نوکلئوتیدی بیشترین فراوانی را داشتند. همچنین ۱۱ موتیف با بیش از ۱۰۰ نوکلئوتید در جدایه های مطالعه شده و ژنوم مرجع شناسایی شد. نتایج این تحقیق نشان می دهد که این داده ها را می توان برای توصیف جدایه های *L. maculans* از مکانهای مختلف جغرافیایی و قابل تعمیم به دیگر جدایه ها استفاده نمود.

کلمات کلیدی: ساق سیاه، SSR، *Leptosphaeria maculans* و NGS

مکاتبه کننده: رقیه همتی Email: rhemati@znu.ac.ir

تاریخ دریافت: ۱۴۰۰/۵/۲۴تاریخ پذیرش: ۱۴۰۰/۹/۲۷