Full Length Research Paper

Frequency and distribution of genome-based microsatellites in *Verticillium dahliae*

ZhiFang Li^{2#}, HeQin Zhu^{1,2#}, LiHong Zhao², ZiLi Feng², YongQiang Shi², JiaRong Yang^{1*}

¹State Key Laboratory of Crop Stress Biology in Arid Areas, and College of Plant Protection, Northwest A and F University, Yangling, Shanxi, 712100, China.

²State Key Laboratory of Cotton Biology, Cotton Research Institute of Chinese Academy of Agricultural Sciences, Anyang, Henan, 455000, China.

Accepted 27 November, 2012

A total of 5,418 simple sequence repeats (SSRs) were identified in the 33.8 Mb genomic DNA sequence of *Verticillium dahliae*. SSR loci were classified by repeat types and frequency in different genomic regions. The results show that the SSRs in different repeat units exhibited differential or non-random distribution in different genomic locations. Whole genome analyses showed that the tri-nucleotide (nt) repeat was the most abundant microsatellite type. The number of tri-nt SSRs was 1,677 comprising 31.0% of the total number of SSRs, followed by hexa-nt, mono-nt, di-nt SSRs, tetra-nt and penta-nt SSRs in that order. In the exonic regions of the genome, the tri-nt SSRs occurred more frequently than the other SSR types. A total of 1,037 (61.8%) tri-nt SSRs were distributed in the exonic regions, an approximately two-fold higher number than in the intergenic regions (66.1 per Mb versus 32.3 per Mb respectively). Nearly half the hexa-nt SSRs were also distributed in the coding region while most of the mono-nt, di-nt, tetra-nt and penta-nt SSRs were predominantly present in the intronic and intergenic regions. The biased distribution of the SSRs may reveal the functional significance of SSRs in the *V. dahliae* genome.

Key words: *Verticillium dahliae*, genome, simple sequence repeats, distribution.

INTRODUCTION

Verticillium dahliae Kleb. is the causal agent of vascular wilt diseases on plants worldwide, resulting in enormous economic losses. It attacks a wide range of plant hosts, including agriculturally important crops such as alfalfa, tomato, cotton, potato, eggplant, pepper, ornamental flowers, fruit trees, and shrubs (Pegg and Brady, 2002; Fradin and Thomma, 2006; Klosterman et al., 2009). In the *Verticillium* genus, it was usually only the isolates that produce microsclerotia that were classified as *V. dahliae* (Barbara and Clewes, 2003), however, sometimes the host specificity and its pathogenicity varies in different

mycelial-type *V. dahliae* isolates without microsclerotia were included (Song et al., 1997). *V. dahliae* exhibits hosts. Isolates from one host can cause a range of symptoms in other hosts but, generally, the symptoms are most severe on the host from which the isolates were obtained (Bhat and Subbarao, 1999). It is well documented that *V. dahliae* exhibits great phenotypic diversity and genetic plasticity.

Microsatellites or simple sequence repeats (SSR) refer to tandem repetitive DNA sequences with motifs that are from 1–6 nucleotides (nt) long. Microsatellites are defined by their ubiquitous abundance, short motif length, high variability and suitability for high-throughput applications (Ellegren, 2004; Selkoe and Toonen, 2006; Dettman and Taylor, 2004). Compared with the other molecular markers, SSRs are more widely used for genetic diversity and genetic mapping because they are largely co-dominant,

*Corresponding author. E-mail: yangjiaorong2012@163.com.
Fax: 86-29-87080156.

[#]ZhiFang Li and HeQin Zhu contributed equally to this work.

multi-allelic, have good genome coverage and can be multiplexed on semi-automated systems (Varshney and Graner, 2005; Zwart et al., 2008). SSRs are not only useful tools for genetic biodiversity, comparison among relatives and phylogenetic analysis, but also for comparing isolates from different host plants (Giraud et al., 2002), strains or races (Barve et al., 2001). Therefore, microsatellite markers in phytopathogenic fungi can be used for population, evolutionary, biodiversity mapping, pathogenicity related genes, race, and ecological research. The development and use of SSR markers for *V. dahliae* lags behind the achievements in other major plant pathogens like, for example, *Fusarium oxysporum*, *Microbotryum violaceum*, *Magnaporthe grisea*, *Microcyclus ulei* and *Plasmopara viticola* (Barve et al., 2001; Giraud et al., 2002; Kaye et al., 2003; Le Guen et al., 2004; Delmotte et al., 2006).

With the rapid development of sequencing technology, the genome of the *V. dahliae* strain VdLs.17 on lettuce has been sequenced and a detailed physical map has been obtained (Klosterman et al., 2008). The current genome assembly of VdLs.17 comprises 52 sequence scaffolds with a total length of 33.8 Mb, and an N50 scaffold length of 1.27 Mb, indicating that 50% of all bases are contained in scaffolds of at least 1.27 Mb (Klosterman, et al., 2011). This high-quality draft *V. dahliae* genome sequence allowed high-throughput SSR markers to be successfully developed. These markers can be readily detected computationally and can be used for the screening of candidate genes, functional research and genetic diversity analysis. Thus, the mining of these resources provides an opportunity to greatly expand the database of molecular markers for *V. dahliae* at minimal cost.

In this study, we explore SSR markers based genome information of *V. dahliae*. The objectives of this research were: (1) to characterize the frequency and distribution of SSRs in the *V. dahliae* genome; and (2) to investigate the relative abundance of SSR types in different regions of the genome.

MATERIALS AND METHODS

The *V. dahliae* genome

The genome of *V. dahliae* VdLs.17 was sequenced using Sanger sequencing technology (Sanger et al., 1977). The sequences were assembled into 52 sequence scaffolds with a total length of 33.8 Mb (Klosterman et al., 2011). The *V. dahliae* VdLs.17 sequences can be downloaded from the Verticillium group Database web site (http://www.broadinstitute.org/annotation/genome/verticillium_dahliae/MultiDownloads.html).

SSR scanning

The genome was scanned for SSRs loci with program software

SciRoKo 3.4 (SSR Classification and Investigation by Robert Kofler) (Kofler et al., 2007). The parameters were set for detection of mono-nt, di-nt, tri-nt, tetra-nt, penta-nt and hexa-nt motifs with a minimum of 15, 7, 5, 3, 3 and 3 repeats, respectively (under the perfect MISA search mode) (Kofler et al., 2007). In our study, we defined two genomic location categories as genic (exon and intron) and intergenic regions. Initially, each SSR was considered to be unique and was subsequently classified according to theoretically possible combinations. For example, a poly-A repeat is equivalent to a poly-T repeat on a complementary strand, and an AAG is equivalent to AGA and GAA in different reading frames and to CTT, TCT and TTC on a complementary strand. Thus, there are two possible combinations for mono-nt repeats, four for di-nt repeats, ten for tri-nt repeats, 33 for tetra-nt repeats, 102 for penta-nt repeats, and 350 for hexa-nt repeats. To localize the distribution of SSRs in different genomic regions, the position of SSRs were compared with the genome annotation by Perl scripts. To describe the abundance of SSRs in different genomic regions, we calculated the "relative abundance" (RA) by dividing the number of SSRs by the mega base-pair (MB) of sequences in our analyses.

RESULTS

Genome-wide distribution of microsatellites in *V. dahliae*

A total of 5,418 microsatellites were identified in the 33.8 Mb genomic DNA sequence of *V. dahliae* using the SciRoKo programs. The results show that SSRs were abundant in the *V. dahliae* genome with about one SSR every 6.24 kb. The SSR loci were classified by repeat type and frequency of repeats per motif (Table 1). The most commonly occurring number of repeats per motif ranged from 3 to 7. The most abundant microsatellite was the tri-nt repeats of which 1,677 (31.0% of the SSRs) were identified, followed by the hexa-nt repeats (930, 17.2%) and mono-nt repeats (811, 15%). The numbers of di-nt, tetra-nt and penta-nt repeats were similar, 659, 633 and 678, respectively (Table 2).

In the *V. dahliae* genome, more SSRs were found in the intergenic regions (64.1%) than in the genic regions (35.9%) (Figure 1 and Table 2). The different SSR repeat units showed obviously differential or non-random distributions in the different genomic locations. The tri-nt repeats were the most abundant SSR type in the genic region, whereas, mono-nt repeats were the most abundant SSR type in the intergenic region (Figure 1 and Table 2). In an attempt to analyze the differential distribution of SSRs more clearly, we characterized the distribution of the SSR types in each repeat unit across the different genomic locations.

Mono-nt SSRs

The mono-nt SSRs were the third largest class of SSRs, representing 15.0% of the SSRs in the *V. dahliae*

Table 1. Frequency and distribution of the 5,418 SSRs identified in the *V. dahliae* genome.

SSR motifs	Number of repeats																	Total
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	>18	
A	0	0	0	0	0	0	0	0	0	0	0	0	36	33	23	14	151	257
C	0	0	0	0	0	0	0	0	0	0	0	0	87	73	60	51	283	554
AG	0	0	0	0	41	46	34	36	36	18	22	17	10	22	16	5	60	363
AC	0	0	0	0	23	32	27	21	32	14	15	13	7	6	11	12	21	234
AT	0	0	0	0	4	7	3	3	1	2	2	1	0	1	2	0	24	50
CG	0	0	0	0	8	3	0	1	0	0	0	0	0	0	0	0	0	12
AGC	0	0	101	68	77	45	45	26	24	18	10	2	13	9	3	5	10	456
CCG	0	0	157	59	66	25	15	6	7	3	0	0	1	2	0	0	0	341
ACG	0	0	71	40	45	21	11	6	11	2	3	1	0	2	0	0	3	216
AGG	0	0	61	35	31	15	11	12	5	3	1	3	4	0	1	0	2	184
AAG	0	0	61	27	41	14	14	7	5	3	2	3	0	1	1	1	1	181
ACC	0	0	55	25	21	18	8	8	6	2	3	4	0	0	0	0	0	150
AAC	0	0	34	13	14	6	7	4	0	0	2	0	0	0	0	0	0	80
ATC	0	0	19	6	11	4	2	1	2	0	0	0	0	0	1	0	1	47
Other Tri-*	0	0	10	1	6	3	0	1	0	0	0	0	0	0	0	0	1	22
ACCT	35	84	47	19	22	9	4	2	1	0	0	1	0	0	0	0	0	224
AGGC	14	17	14	17	4	7	1	2	1	0	1	0	1	0	1	0	1	81
AGCC	8	11	6	11	0	2	3	2	0	0	0	0	0	0	0	0	0	43
AAGC	2	14	7	1	4	0	1	1	1	1	0	0	0	0	0	0	2	34
ATCC	9	11	3	4	1	0	0	0	1	1	0	0	0	0	0	0	0	30
ACAG	5	5	3	10	2	2	0	1	1	0	0	0	0	0	0	0	0	29
AAAG	5	5	6	6	2	1	0	0	0	0	0	0	0	0	1	0	0	26
AGGG	8	7	2	0	1	2	0	0	0	0	0	0	0	0	0	0	0	20
Other Tetra-*	58	58	30	14	5	4	1	1	0	1	0	0	0	0	1	0	3	176
AAGGT	42	10	4	2	1	0	0	0	0	0	0	0	0	0	0	0	0	59
AGGGG	23	6	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	32
AAGGC	19	6	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	27
AGCCC	16	6	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24
ACCTC	8	7	4	0	2	0	0	0	0	0	0	0	0	0	0	0	0	21
AGGCG	16	1	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	21
Other Penta-*	358	77	26	17	9	3	1	2	0	0	0	0	0	0	0	0	1	494
AGCGGC	17	14	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	35
ACCAGC	9	12	4	2	2	0	1	0	0	0	0	0	0	0	0	0	1	31
ACGAGG	12	3	3	3	0	1	1	0	0	0	0	0	0	0	0	0	0	23
AGGGCG	12	4	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	23
AGCAGG	12	5	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	22
ACGGCG	13	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21
Other Hexa-*	423	172	86	33	17	16	6	3	4	2	2	2	1	1	1	0	6	775

*Motifs with less than 20 SSRs in the genome are not listed.

genome. The mono-nt SSRs were distributed preferentially in the intergenic (690 loci) and intronic

regions (112 loci) but were rare in the exonic regions (only 9 loci). The relative abundances of the mono-nt SSRs in

Table 2. Relative abundance of SSR types in different regions of the *V. dahliae* genome

Parameter	Intergenic region(17.1Mb)		Genic region (16.7Mb)				Total(33.8Mb)	
	Count	RA*	Exon(15.7Mb)		Intron(1.0Mb)		Count	RA
SSR	Count	RA*	Count	RA	Count	RA	Count	RA
Mono-nt	690	40.4	9	0.6	112	112	811	24.0
Di-nt	571	33.4	7	0.4	81	81	659	19.5
Tri-nt	552	32.3	1037	66.1	88	88	1677	49.6
Tetra-nt	611	35.7	14	0.9	38	38	663	19.6
Pentra-nt	588	34.4	39	2.5	51	51	678	20.1
Hexa-nt	459	26.8	443	28.2	28	28	930	27.5

*Relative abundance = number of SSR type/region size in mega bases (Mb).

Figure 1. Genome-wide distribution and relative abundance of SSR types by their unit size. Each bar represents the relative abundance of the SSR types in different genome locations.

the intergenic, exonic and intronic regions were 40.4, 0.6 and 112 per Mb, respectively (Table 2 and Figure 1). Of the two possible types of mono-nt SSRs, poly-G/C was the predominant form with 554 loci and a relative abundance of 16.3 poly-G /C repeats per Mb in the genome. The poly-A/T SSR type had 257 loci with a relative abundance of 7.6 per Mb (Table 3).

Di-nt SSRs

The total number of di-nt SSRs in the *V. dahliae* genome was 659. Most of di-nt SSRs (571) were distributed in the intergenic regions, 81 were in the intronic regions, and only seven were in the exonic regions. Overall, 98% of the di-nt SSRs were found in the noncoding regions (intergenic and intronic) (Table 2). Among them, the AG/TC SSRs (363) were the most abundant with a

relative abundance of 10.7 SSRs per Mb. The AC/TGs (234) also had a high relative abundance with about 6.9 SSRs per Mb in the genome.

Tri-nt SSRs

The tri-nt SSRs were the most abundant in terms of unit numbers being 31.0% of the total number of SSRs in the genome (1,677 of the 5,418) (Table 2). 1,037 (61.8%) of the tri-nt SSRs were distributed in the exonic regions, 552 were in the intergenic regions, and only 88 were found in the intronic regions. Thus, the relative abundance of tri-nt SSRs in the exonic regions was approximately two fold higher than in the intergenic regions (66.1 per Mb versus 32.3 per Mb) (Table 2 and Figure 1). The distribution of some of the tri-nt SSR types in the genome was non-random. For example, the AGC/TCG, CCG/GGC,

ACG/TGC, AAC/TTG and AGG/TCC SSRs were preferentially located in the exonic regions, the ACC/TGG and ATC/TAG SSRs were exclusively located in the intergenic regions, and the AAG/TTC SSRs were almost equally distributed in the exonic and intergenic regions (Table 3).

Tetra-nt and penta-nt SSRs

The tetra-nt and penta-nt SSRs, with 33 and 102 SSR types respectively were predominantly distributed in the intergenic and intronic regions (Table 2 and Figure 1). Among them, the most frequent tetra-nt SSR type was ACCT/TGGA with 6.6 per Mb in the genome. 97% of the ACCT/TGGA SSRs were located in the intergenic regions, followed by AGGC/TCCG, 88.9% of which were distributed in the intergenic regions and 11.1% were in the intronic regions. No AGGC/TCCG SSRs were found in the exonic regions (Table 3). The distribution of the penta-nt SSRs was similar to that of the tetra-nt SSRs. The most abundant penta-nt SSR type was AAGGT/TTCCA, 98.3% of which were in the intergenic regions. The relative abundance of the penta-nt SSRs in the intergenic regions was close to 14-fold higher than in the exonic regions (Table 2).

Hexa-nt SSRs

Hexa-nt SSRs were the second largest class by repeat units representing 17.2% of the total SSRs in the *V. dahliae* genome with a relative abundance of 27.5 per Mb. The hexa-nt SSRs were found preferentially in the intergenic (459) and exonic regions (443), and were rare in the intronic regions (28) (Table 2). The hexa-nt SSRs can combine to form 350 SSR types in all. The AGCGGC/TCGCCG SSRs with a total number of 35 were the most common type, 77.1% of them were found in the exonic regions of the genome (Table 3).

DISCUSSION

As the primary destructive phytopathogenic species in the genus, population and genetic diversity of *V. dahliae* have been areas of intensive study (Klosterman et al., 2009). Although random amplified polymorphic DNA (RAPD) (Perez-Artes et al. 2000) and amplified fragment length polymorphic (AFLP) markers (Collins et al., 2003) have been used to characterize molecular diversity in the populations of *V. dahliae*, the dominance of the RAPD and AFLP markers prevent the detection of heterozygotes in diploid species. The development of co-dominant, multi-allelic and sequence-specific microsatellite markers

may enable the molecular genotypes at a number of loci to be determined, and markers of this type could be more widely used for the genetic diversity and genetic mapping of *V. dahliae*. An array of 22 simple sequence repeat markers of *V. dahliae* has been developed to detect recombination, transcontinental gene flow and genetic drift (Almany et al., 2009; Atallah et al., 2010). But there must be more potential SSR loci linked with other genetic characteristics across the whole genome to be explored.

The completed genome sequences of *V. dahliae* has greatly assisted the understanding of SSRs at the genome-wide level. We identified 5,418 SSR loci in the *V. dahliae* genome. The relative abundance was 160 SSRs per Mb (Table 1) and the distribution of the SSRs was 811, 659, 1,677, 633, 678 and 930 for the mono-nt, di-nt, tri-nt, tetra-nt, penta-nt and hexa-nt SSRs, respectively (Table 2). The tri-nt SSRs were the major class, accounting for 31.0% of the total. This result was consistent with the findings for *Neurospora crassa* in which 39.4% of the SSRs were tri-nt repeats (Kim et al., 2008). However, the *Aspergillus nidulans*, *Cryptococcus neoformans*, *Magnaporthe grisea*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Ustilago maydis* genomes contained more mono-nt SSR loci than the other types while di-nt SSRs were predominant in the genomes of *Encephalitozoon cuniculi* and *Fusarium graminearum* (Karaoglu et al., 2005).

In this study, we found that the distribution of the SSRs in the *V. dahliae* genome is not random. This finding is consistent with the results from previous studies (Li et al., 2004; Kim et al., 2008). The abundance of SSRs was differentially distributed across exonic, intergenic, and intronic regions of the genome. Mono-nt repeats were found at a relatively high frequency in the *V. dahliae* genome, most (690 out of 811) were distributed in the intergenic regions and only nine were in the exonic regions (Table 2). We found a strong over-representation of poly-G/C compared with poly-A/T SSRs in the *V. dahliae* genome, unlike the findings reported in other eukaryotes (Karaoglu et al., 2005; Lawson and Zhang, 2006; Chistiakov et al., 2006). Di-nt SSRs were preferentially distributed in the intergenic regions and rare in genic regions as reported in other organisms (Li et al., 2002). Among the di-nt SSR motifs, AG/TC and AC/TG were the most abundant in agreement with the reported differences in SSR composition between fungi and other organisms (Ellegren, 2004). The GC/CG SSRs were the least abundant di-nt motif with only nine repeats, as described in many other fungi (Karaoglu et al., 2005).

In the exonic regions, tri-nt and hexa-nt SSRs were the dominant types. The tri-nt repeats were approximately two-fold more frequent in the exonic regions than in the intronic and intergenic regions combined (Table 2, Figure 1). The enrichment of the tri-nt SSRs in the exonic regions was observed in a variety of other eukaryotic organisms

Table 3. Distribution of SSRs types in different regions of the *V. dahliae* genome.

SSR motifs	Genic region repeat number		Intergenic region repeat number	Total
	Exon	Intron		
A	4	50	203	257
C	5	62	487	554
AG	1	53	309	363
AC	2	17	215	234
AT	0	9	41	50
CG	4	2	6	12
AGC	307	20	129	456
CCG	288	9	44	341
ACG	149	12	55	216
AGG	100	15	69	184
AAG	81	18	82	181
ACC	52	7	91	150
AAC	49	3	28	80
ATC	11	2	34	47
Other Tri-	0	2	20	22
ACCT	2	4	218	224
AGGC	0	9	72	81
AGCC	1	2	40	43
AAGC	1	3	30	34
ATCC	1	5	24	30
ACAG	2	1	26	29
AAAG	0	4	22	26
AGGG	0	3	17	20
Other Tetra-	7	7	162	176
AAGGT	0	1	58	59
AGGGG	1	6	25	32
AAGGC	1	1	25	27
AGCCC	3	1	20	24
ACCTC	0	2	19	21
AGGCG	1	1	19	21
Other Penta-	33	39	422	494
AGCGGC	27	0	8	35
ACCAGC	14	0	17	31
ACGAGG	18	1	4	23
AGGGCG	16	0	7	23
AGCAGG	8	3	11	22
ACGGCG	14	0	7	21
Other Hexa-	346	24	405	775
Total	1549	398	3471	5418

(Li et al., 2004; Kim et al., 2008). The tri-nt and hexa-nt SSRs in the exonic regions are translated into amino acid repeats in the proteins that they encode. These amino acid repeats may contribute to the biological functions of the proteins. The dominance of SSR triplets over the other repeats in coding regions may be explained by specific

selection operating against frameshift mutations and a tight negative selection on the other SSRs that would perturb the reading frames in coding regions (Metzgar et al., 2000; Young et al., 2000). Of the tri-nt SSRs, the AGC/TCG repeat that encodes an Asp amino acid repeat was the most abundant. The most common hexa-nt SSR

was the AGCGGC/TCGCCG repeat that encodes an Asp-Gly dipeptide repeat (Table 3). This finding suggested that the differential appearance of particular tri-nt and hexa-nt SSRs within the exonic regions may be the result of functional selection on amino acid repeats in encoded proteins (Li et al., 2002, 2004; Katti et al., 2001). Furthermore, the tri-nt and hexa-nt triplet repeats in exonic regions are variable across the *V. dahliae* genome (data not shown), perhaps indicating that these SSRs play important roles in the evolution of gene functions that may help promote adaptation to new environments (Li et al., 2004; Kashi and King, 2006).

In the *V. dahliae* genome, the occurrence of tetra-nt and penta-nt SSRs was similar, and most of these SSRs were distributed in the intergenic regions. The most frequent tetra-nt SSRs were the ACCT/TGGA repeats, in concurrence with the findings in *N. crassa* and *M. grisea* (Karaoglu et al., 2005). The most common penta-nt SSRs were found much less frequently than the tetra-nt SSRs; for example, 59 AAGGT/TTCCA repeats were found in the *V. dahliae* genome and 58 of them were in the intergenic regions, the remaining one was in an intronic region. In conclusion, this study of SSRs in the completely sequenced plant pathogenic fungus *V. dahliae*, is a small step towards a better understanding of the important role of these sequences. The microsatellite analysis showed that the distribution of SSRs in exonic, intronic and intergenic regions of the genome were non-random and strongly biased, probably reflecting the functional significance of SSRs. The enhanced frequency of tri-nt SSRs in the coding regions might indicate the effects of selection against possible frameshift mutations. The SSRs in *V. dahliae* could be explored as evolutionary neutral DNA markers, and the data that we have obtained could be used to select highly polymorphic SSR markers to study the population genetic diversity and evolution of the *Verticillium dahliae*. Furthermore, the SSRs loci mentioned in this study also have the potential to explore useful markers linked with significant genetic characteristics, such as pathogenicity related loci, key loci for differentiation between defoliating and nondefoliating pathotypes, and microscleotia producing loci.

ACKNOWLEDGEMENTS

We greatly acknowledge Dr. Changsong Zou for providing Perl scripts to analyze the SSRs position in the genome. We also sincerely appreciate Dr. Muhammad Ibrahim Khaskheli for his comments and critical reading of the manuscript. The authors also extend our gratitude to the senior editor and reviewers of the manuscript. This work was supported by the Central Public-interest Scientific Institution Basal Research Fund, China (SJB1203) and National Science and Technology

Supporting Planed Project (2011BAD08A07).

REFERENCES

- Almany GR, De Arruda MP, Arthofer W, et al. (2009). Permanent genetic resources added to molecular ecology resources database 1. Mol. Ecol. Resour. 9:1460-1466.
- Atallah ZK, Maruthachalam K, Toit L, Koike ST, Davis RM, Klosterman SJ, Hayes RJ, Subbarao KV (2010). Population analyses of the vascular plant pathogen *Verticillium dahliae* detect recombination and transcontinental gene flow. Fungal Genet. Biol. 47:416-422.
- Barbara DJ, Clewes E (2003). Plant pathogenic *Verticillium* species: How many of them are there? Mol. Plant Pathol. 4:297-305.
- Barve MP, Haware MP, Sainani MN, Ranjekar PK, Gupta VS (2001). Potential of microsatellites to distinguish four races of *Fusarium oxysporum* f sp. ciceri prevalent in India. Theor. Appl. Genet. 102:138-147.
- Bhat RG, Subbarao KV (1999). Host range specificity in *Verticillium dahliae*. Phytopathology 89:1218-1225.
- Chistiakov DA, Hellemans B, Volckaert FAM (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture 255:1-29.
- Collins A, Okoli CAN, Morton A, Parry D, Edwards SG, Barbara DJ (2003). Isolates of *Verticillium dahliae* pathogenic to crucifers are of at least three distinct molecular types. Phytopathology 93:364-376.
- Delmotte F, Chen WJ, Richard-Cervera S, Greif C, Papura D, Giresse X, Mondor-Genson G, Corio-Costet MF (2006). Microsatellite loci from the grape downy mildew (*Plasmopara viticola*). Mol. Ecol. Notes 6:379-381.
- Dettman JR, Taylor JW (2004). Mutation and evolution of microsatellite loci in *Neurospora*. Genetics 168:1231-1248.
- Ellegren H (2004). Microsatellites: simple sequences with complex evolution. Nat Rev Genet. 5(6):435-445.
- Fradin EF, Thomma BPHJ (2006). Physiology and molecular aspects of *Verticillium* wilt diseases caused by *V. dahliae* and *V. albo-atrum*. Mol. Plant Pathol. 7: 71-86.
- Giraud FE, Vautrin D, Solignac M, Shykoff A (2002). Isolation of 44 polymorphic microsatellite loci in three host races of the phytopathogenic fungus *Microbotryum violaceum*. Mol. Ecol. Notes 2:142-146.
- Karaoglu H, Lee CM, Meyer W (2005). Survey of simple sequence repeats in completed fungal genomes. Mol. Biol. Evol. 22(3):639-649.
- Kashi Y, King DG (2006). Simple sequence repeats as advantageous mutators in evolution. Trends Genet. 22(5):253-259.
- Katti MV, Ranjekar PK, Gupta VS (2001). Differential distribution of simple sequence repeats in eukaryotic genome sequences. Mol. Biol. Evol. 18(7):1161-1167.
- Kaye C, Milazzo J, Rozenfeld S, Lebrun M, Tharreau D (2003). The development of simple sequence repeat markers for *Magnaporthe grisea* and their integration into an established genetic linkage map. Fung. Genet. Biol. 40:207-214.
- Kim TS, Booth JG, Gauch HG, Sun Q, Park JS, Lee YW, Lee K (2008). Simple sequence repeats in *Neurospora crassa*: distribution, polymorphism and evolutionary inference. BMC Genomics 9:31-49.
- Klosterman SJ, Atallah ZK, Vallad GE, Subbarao KV (2009). Diversity, Pathogenicity, and Management of *Verticillium* Species. Annu. Rev. Phytopathol. 47:39-62.
- Klosterman SJ, Veronese P, Dobinson KF, Subbarao K, Heiman DI (2008). *Verticillium* comparative genomics: understanding pathogenicity and diversity. Phytopathology 98: S185 (Abstr.).
- Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, Zehua Chen ZH, Henrissat B, Lee YH, Park JS, Garcia-Pedrajas MD, Barbara DJ, Anchieta1 A, Jonge RD, Santhanam P, Maruthachalam K, Atallah Z, Amyotte SG, Paz Z, Inderbitzin P, Hayes RJ, Heiman DI, Young S, Zeng QD, Engels R, Galagan J, Cuomo CA, Dobinson KF, Ma LJ (2011). Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog. 7: e1002137.

- Kofler R, Schlötterer C, Lelley T (2007). SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics* 23:1683-1685.
- Lawson MJ, Zhang L (2006). Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol.* 7:R14.
- Le Guen V, Rodier-Goud M, Troispoux V, Xiong TC, Brottier P, Billot C, Seguin M (2004). Characterization of polymorphic microsatellite markers for *Microcyclus ulei*, causal agent of South America leaf blight of rubber trees. *Mol. Ecol. Notes* 4:122-124.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 11(12):2453-2465.
- Li YC, Korol AB, Fahima T, Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21(6):991-1007.
- Metzgar D, Bytof J, Wills C (2000). Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.* 10:72-80.
- Pegg GF, Brady BL (2002). *Verticillium wilts*. CABI Publishing, New York, USA pp. 1-432.
- Perez-Artes E, Garcia-Pedrajas MD, Bejarano-Alcazar J, Jimenez-Diaz RM (2000). Differentiation of cotton-defoliating and nondefoliating pathotypes of *Verticillium dahliae* by RAPD and specific PCR analyses. *Eur. J. Plant Pathol.* 106: 507-517.
- Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci.* 74(12):5463-5467.
- Selkoe KA, Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* 9:615-629.
- Song XX, Zhu HQ, Guo JC (1997). Study on pathogenicity differentiation for *Verticillium dahliae* Kleb. Anyang strains in cotton. *Sci. Agr. Sin.* 1:13-18.
- Varshney RK, Graner A, Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23:48-55.
- Young ET, Sloan JS, Riper KV (2000). Trinucleotide repeats are clustered in regulatory genes in *Saccharomyces cerevisiae*. *Genetics* 154:1053-1068.
- Zwart RS, Muylle H, Bockstaele EV, Roldán-Ruiz I (2008). Evaluation of genetic diversity of *Fusarium* head blight resistance in European winter wheat. *Theor. Appl. Genet.* 117:813-828.