Universal Journal of Agricultural Research 11(4): 691-704, 2023 http://www.hrpub.org
DOI: 10.13189/ujar.2023.110403

Decoding the Repeat Landscape: Genomic Analysis of Phytopathogenic *Fusarium* Species

Mohammad Alhawatema

Department of Applied Biological Science, Faculty of Science, Tafila Technical University, Jordan

Received March 15, 2023; Revised June 9, 2023; Accepted July 19, 2023

Cite This Paper in the Following Citation Styles

(a): *[1] Mohammad Alhawatema , "Decoding the Repeat Landscape: Genomic Analysis of Phytopathogenic Fusarium Species," Universal Journal of Agricultural Research, Vol. 11, No. 4, pp. 691 - 704, 2023. DOI: 10.13189/ujar.2023.110403.*

(b): *Mohammad Alhawatema (2023). Decoding the Repeat Landscape: Genomic Analysis of Phytopathogenic Fusarium Species. Universal Journal of Agricultural Research, 11(4), 691 - 704. DOI: 10.13189/ujar.2023.110403.*

Abstract This study analyzes the genomes of ten *Fusarium* species using RepeatMasker software to identify complete Simple Sequence Repeats (SSRs) and interspersed repeats. Genome size and GC content among the ten species were clearly varied based on the comparative analysis of genomic features. The different levels of repetitive DNA among the species are partly responsible for the variations in genome size and the frequency of other genomic elements, such as retrotransposons and transposons, also differs dramatically among the species. The high frequency of simple repeats occurs often in *F. aywerte* and F. *falciforme*, indicating that their genomes are relatively unstable and susceptible to mutations and genetic variation. The findings show that there are more tri- and tetra-repeats than mono- and di-repeats, and that the number of repeats tends to rise with the size of the repeat unit. The tri-repeat is the most common repeat type, followed by hexa-repeats and penta-repeats. It is crucial to look at repeat number variation at both the species and repeat type levels because the number of repeats might vary greatly within a single species. The findings also show that some species clearly favor particular repeat forms. The understanding of the genetic variety and evolution of *Fusarium* species, as well as the making of successful management methods, are significantly impacted by these results. The presence of repeat type preferences may indicate to the potential functions that these repeats in the genome may play. The future research will be focused on these *Fusarium* species for controlling repeat expansion and contraction regarding of the variations in repeat numbers within each species. This study analysis sheds light on the relative density of SSRs in various *Fusarium* species.

Keywords *Fusarium*, RepeatMasker, Simple Sequence Repeats (SSRs), Repeat Type, Genetic Diversity

1. Introduction

Fusarium is a major genus of filamentous fungi. It is a member of the group of organisms that are often known as hyphomycetes. *Fusarium* is commonly found in soil and is connected to plants. The vast majority of species are non-pathogenic saprobes, and they are rather common members of the community of soil microbes. In cereal crops, certain species can release mycotoxins, which, if they make their way into the food chain, can have adverse effects on both human and animal health. Fumonisins and trichothecenes are the two primary carcinogens that are generated by these *Fusarium* species [1-3].

Certain *Fusarium* species have been identified as major pathogens of plants and animals, despite the fact that the majority of species appear to be harmless. The classification scheme of the genus is somewhat difficult to understand. There have been a variety of techniques taken, with schemes ranging from broad to narrow ideas of speciation, and there have been periods when as many as one thousand distinct species have been recognized [1-3].

The processes that underlie the development and evolution of pathogenesis in *Fusarium* species have

received the most attention from oomycete biologists as a research topic [1-3]. Genome structure is responsible for controlling virulence. Genome evolution and biodiversity in *Fusarium* species are driven by DNA repeats and in part by transposable elements (TEs), which also contribute to the instability of the genome [4, 5].

The TEs, also known as a transposon or jumping gene, is a sequence of nucleic acid in DNA that has the capability of altering its location within a genome. This can cause the formation or reversal of mutations, as well as changes to the genetic identity of the cell and the size of its genome [4]. Many are crucial to the functioning of the genome and the process of evolution [5]. One of the many kinds of mobile genetic elements is called a transposable element. TEs are placed into one of two classes based on the technique used by transposition, which can be either (Class II TEs) or (Class I TEs) [6]. Retrotransposons with long terminal repeats (LTRs), Retroposons, long interspersed nuclear elements (LINEs, LINE-1s, or L1s), and short interspersed nuclear elements (SINEs) are all examples of (Class I TEs) DNA transposons are included in Class II Transposable Elements [6].

In addition to TEs, simple sequence repeats (SSRs) or known as microsatellites, are a stretch of tandemly repeated (that is, next to each other) DNA sequences that are between one and ten nucleotides long and are usually repeated between 5 and 50 times and are known as mono, di, tri, tetra, penta, hexa, and heptanucleotide motifs refer to repeat units with one, two, three, four, five, six, or seven nucleotides [7, 8]. Most eukaryotic cells have microsatellites, which are found all over the genome [7, 8]. Mutations in microsatellites cause either the addition or deletion of a whole repeat unit, and often two or more repeats at the same time. A suggested reason for the occurrence of changes in length is the occurrence of replication slippage, which arises due to the mismatches between DNA strands during replication in meiosis [9]. The rate of mutation at microsatellite loci is likely to be higher than other rates of mutation [10]. No agreement about what makes mutations in microsatellite. However, the rate of mutation goes up as the number of repeats goes up, reaching a peak at eight repeats and then going down again [11]. Many SSRs are found in DNA that don't code for proteins and don't do anything biologically. Microsatellite mutations can cause phenotypic changes and diseases in places where they are found in regulatory regions and or perhaps even in coding DNA [12]. Microsatellites are used in many applications for population genetics, biodiversity, Genetic linkage analysis, Forensic and medical fingerprinting, plant breeding, and cancer diagnosis [12, 7, 8].

In this research study, we selected ten assembled *Fusarium species* genomes with complete sequences and annotations from Genbank. We then analyzed these *Fusarium* genomes for interspersed repeats and SSRs, focusing on their frequency, density, motifs, length, and conservation. Our hypothesis was that the genomic characteristics of these *Fusarium species*, including genome size, GC content, and frequency of genomic elements and SSRs, would exhibit variation among the species, and these variations which might be associated with their biology, evolution, and potential functional roles in their genomes.

2. Methodology

2.1. Selection Genomic Sequences from *Fusarium* Species

Genomes of ten *Fusarium* species were carefully selected from NCBI-Genome (https://www.ncbi.nlm.nih.gov/genome) (table 1). The genomes of all *Fusarium* species have been fully annotated, the full genome sequences of *Fusarium* species were downloaded in files as FASTA format. the genome sequences of ten *Fusarium* species analyzed were similar to the estimated genome sizes.

Table 1. List of Analyzed *Fusarium* species genomes in the study

Organism	GenBank ID	Genome Size (Mb)	GC Content (%)
Fusarium falciforme	WESZ00000000.1	59.29	48.9
Fusarium nygamai	WEYU00000000.1	45.77	46.6
Fusarium verticillioides	ASM1103527v1	42.67	48.3
Fusarium hostae	ASM1103582v1	54.21	46.5
Fusarium aywerte	ASM1103409v1	40.25	47.3
Fusarium solani	ASM1103338v1	55.02	50.7
Fusarium graminum	ASM1326616v1	34.98	48.9
Fusarium buxicola	ASM1489909v1	35.24	52.2
Fusarium sibiricum	ASM1489899v1	40.62	48.3
Fusarium goolgardi	ASM1489907v1	34.84	48.2

2.2. Identification of Repeats

To identify repeats in genomes of each *Fusarium* species, we used RepeatMasker (v 4.1.0) software (http://www.repeatmasker. org/ accessed on 15 October 2022) to search for repetitive sequences for interspersed and SSR motifs in each genome [13]. After the program was completed, results were sent out in four files for each analyzed genome. The results files include information about repeat types, repeat units, repeat locations, repeat sequences, and flanking regions of repeats. For counting and grouping each type repeat, output files were then downloaded into Microsoft Excel sheet.

The software results did not differentiate between the occurrence of repeats whether in coding and non-coding regions.

2.3. Analysis of Repeats

To make comparisons between genome sequences of varying lengths easy, the frequency and density of genomic elements were standardized as a value per 30 million base pairs (Mb) of sequence. Additionally, the total frequency and density values were adjusted to represent the number of SSRs per 10 million base pairs (Mb) of sequence. We calculated the estimated simple repeats density (bp/10Mb) for each genome by the number of base pairs of sequence contributed by each SSR. We searched for mono, di, tri, tetra, penta, hexa, and heptanucleotide (hepta) repeats. All SSRs types were searched for their (frequency) occurrence, relative abundance per 10Mb, and density per 10Mb and then were counted and calculated using Microsoft Excel sheet. Also, the most frequent SSRs, longest SSRs, and conserved SSRs in all genomes were searched. Also interspersed repeats were counted and the bases masked by each interspersed repeat and SSRs were calculated.

3. Results

Genomes of ten *Fusarium species* were analyzed for frequency of different genomic elements, total repeats of SSRs, relative density of SSR repeats, frequency of SSR motifs, the longest SSR repeats, and conserved SSRs through all genomes (Tables 1-7).

3.1. Genome Size and GC Content

The table 1 provides a comparative summary of genomes of ten different *Fusarium* species, including their specific GenBank IDs, genome sizes, and GC content. The genome sizes range from 34.84 Mb for *Fusarium goolgardi* to 59.29 Mb for *Fusarium falciforme*. The GC content also varies, with *Fusarium buxicola* having the highest GC content of 52.2% and *Fusarium hostae* having the lowest GC content of 46.5%. Interestingly, some of the *Fusarium* species with similar genome sizes have different GC content, such as *Fusarium aywerte* and *Fusarium sibiricum*, which have genome sizes of 40.25 Mb and 40.62 Mb, respectively, but have different GC contents of 47.3% and 48.3%.

3.2. Frequency of Different Genomic Elements

The analysis of all genomes for the frequency and density of various genomic elements per 30Mb offers insights into the differences and similarities between species at the genomic level (Table 2, Figure 1). For LTRs, *F. verticillioide* lacks LTRs (with a count of 0), while *F. solani* and *F. nygamai* exhibit the highest frequency (with a count of 8). In terms of LINEs, LINE1 and LINE2, *F. hostae* has the highest frequency of LINE1 (9), while *F. solani* has the highest frequency of LINE2 (41).

L3/CR1, another retrotransposon type, shows the highest frequency in *F. falciforme* (46) and the lowest in *F. nygamai* (6). For SINEs, *F. sibiricum* has the lowest frequency (6), whereas *F. nygamai*, *F. verticillioides*, and *F. solani* share the highest frequency (43). Regarding DNA transposons, *F. goolgardi* has the highest frequency (40), while *F. nygamai* has the lowest (6). Simple repeats analysis reveals that *F. aywerte* and *F. falciform* have the highest frequency (5814 and 5759, respectively), while *F. graminum* has the lowest (4423). *F. goolgardi* exhibits the highest frequency of low complexity regions (742), while *F. buxicola* has the lowest (504). Small RNA elements or short RNA molecules involved in gene regulation and other cellular processes, are most frequent in *F. nygamai* (276) and least frequent in *F. falciforme* (80). The highest bases masked by genomic element repeats were observed for *F. falciforme* (305,342 bp, or 1.04% of the total genome size) (Table 2, Figure 2, 3), while the lowest were observed for *F. graminum* (68,667 bp, or 0.76% of the total genome size) (Table 2, Figure 3).

Table 2. Shows the frequency of different types of genomic elements of repeats in the genomes of several species in the *Fusarium* genus.

Organism	Analyzed Genome	LTR Element	LINE1	LINE2	L3/CR1	SINE	DNA Transposon	Simple Repeats	Low Complexity	Small RNA	Unclassified	Bases Masked
F. falciforme	59.3	3 154	7 515	39 2946	46 3411	20 1232	28 2205	5759 253747	715 35662	80 7425	1 45	305342
F. nygamai	45.77	4 220	3 234	5 262	6 427	19 1218	6 316	4709 203678	592 30297	276 26203	0	266055
F. verticillioides	42.67	0	4 224	26 1693	39 2560	23 1476	32 2351	4904 201695	533 25422	159 11681	0	246602
F. hostae	54.21	3 168	9 591	35 2391	45 3050	23 1490	37 2506	4531 199010	581 29504	133 10537	0	242747
F. aywerte	40.25	2 101	8 617	40 2990	44 2935	22 1385	38 2470	5814 238354	680 33484	197 16645	1 75	309056
F. solani	55.02	4 260	6 481	41 3314	42 3111	23 1451	30 2233	5046 216062	628 30793	104 8509	0	263214
F. graminum	34.98	2 108	6 318	25 1670	32 1999	16 1161	34 2140	4423 17799	544 25254	174 15564	1 454	68667
F. buxicola	35.24	3 116	4 251	29 2191	41 2841	9 678	20 1505	5194 205825	504 24161	103 10226	0	245794
F. sibiricum	40.62	1 24	5 299	20 1351	21 1506	6 377	25 1533	4699 182759	553 25393	151 13459	0	223701
F. goolgardi	34.84	1 7808	5 398	39 2877	31 2076	11 817	40 2721	5660 217243	742 33756	167 18573	2 110	304479

The initial value in the sequence represents the frequency of genomic elements per 30Mb. The subsequent numbers in the table's cells indicate the total lengths in base pairs (bp) for every 30 Mb.

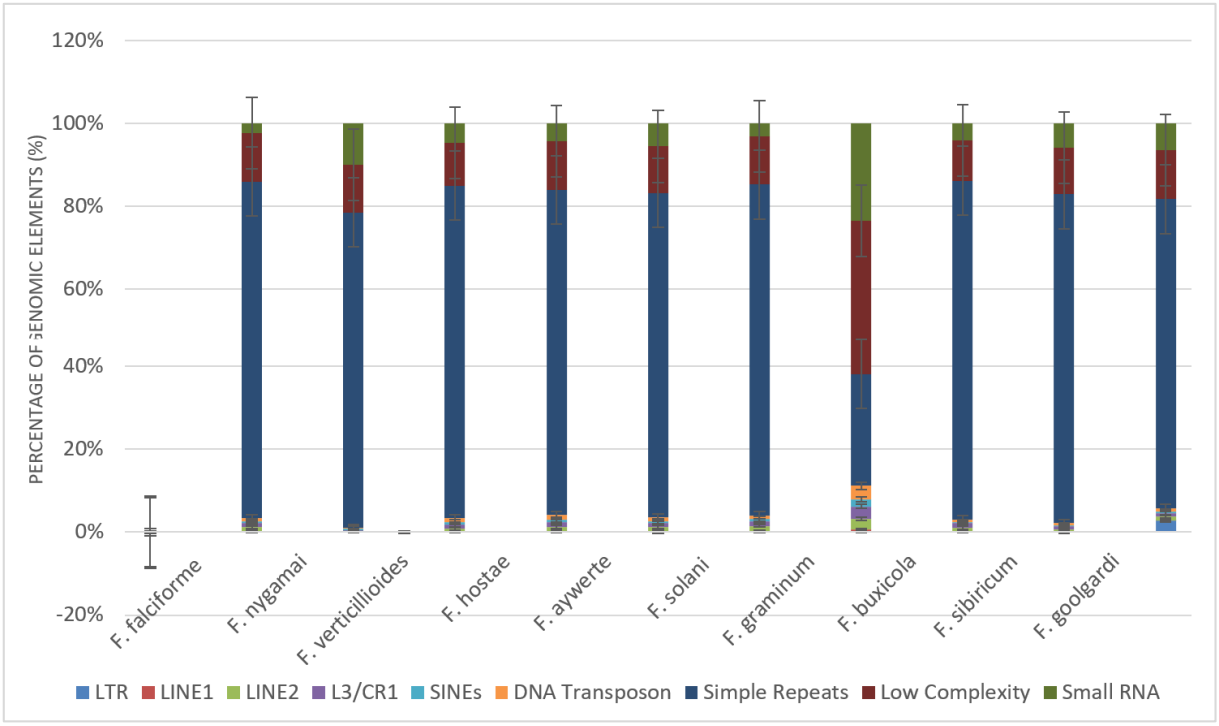

Figure 1. The percentage distribution of various genomic elements in the genomes of several *Fusarium* species

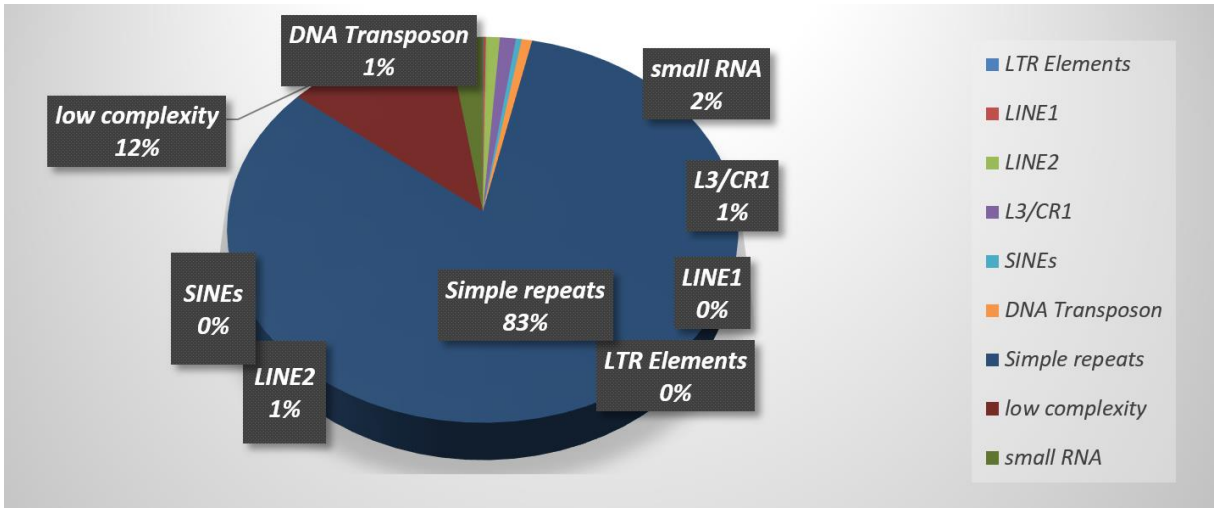

Figure 2. The percentage of different types of genomic elements in the genomes of *F. falciforme*

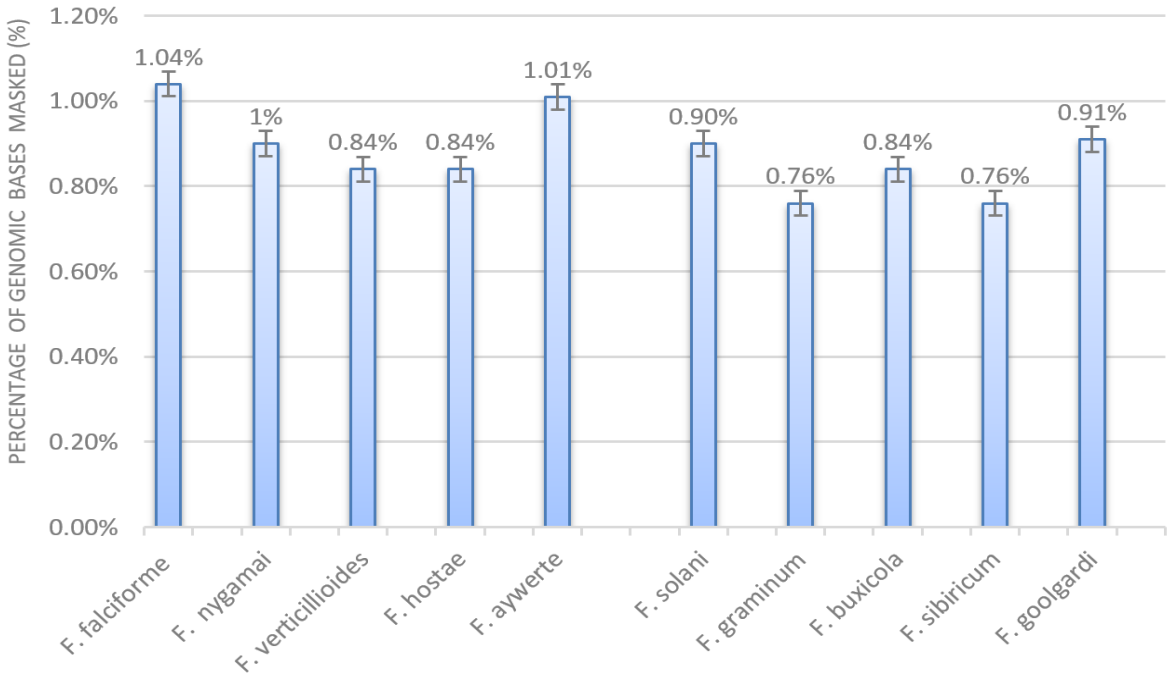

Figure 3. Comparison of the percentage (%) of total genomic bases masked by all genomic elements in various *Fusarium* species. The error bars shown in each column

3.3. Total Repeats of SSRs

The analysis shows that the number of repeats varies widely between *Fusarium* species and repeat types per 10Mb (Table 3). In general, the number of repeats tends to increase with the size of the repeat unit (Table 3). For example, tri- and tetra-repeats tend to be more common than mono- and di-repeats, while penta-, hexa-, and hepta-repeats tend to be even more common. Overall, the most common repeat type is the tri-repeat, followed by the hexa-repeat and penta-repeat, while the least common repeat type is the mono-repeat. In general, the number of repeats tends to increase with the size of the repeat unit. For example, tri- and tetra-repeats tend to be more common than mono- and di-repeats, while penta-, hexa-, and hepta-repeats tend to be even more common. The number of repeats can also vary widely within a single species. For example, *F. aywerte* has the highest total number of repeats overall, but the number of repeats within each repeat type varies from 109 (hepta) to 497 (hexa). Some species show a clear preference for certain repeat types. For example, *F. nygamai* and *F. graminum* both have relatively low numbers of mono- and di-repeats, but relatively high numbers of hexa-repeats. *F. buxicola*, on the other hand, has an unusually high number of hexa-repeats.

3.4. Relative Density of SSR Repeats

The relative density of SSR repeats in different *Fusarium* species was analyzed (Table 4):

Mono SSR repeats were found to be the abundant in *Fusarium* species, with the highest density observed in *F. aywerte* (6289) and the lowest in *F. sibiricum* (1407).

Di SSR repeats were also found to be common in all organisms, with the highest density observed in *F. nygamai* (4770) and the lowest in *F. buxicola* (2144). Tri SSR repeats were found to be the most abundant in *F. aywerte* (12652) and *F. solani* (15788) but were relatively less abundant in *F. graminum* (10754). Tetra SSR repeats were found to be relatively less abundant in all organisms, with the highest density observed in *F. aywerte* (11339) and the lowest in *F. graminum* (4843).

Penta SSR repeats were found to be more abundant in *F. aywerte* (17315) and *F. buxicola* (11200) but less abundant in *F. graminum* (10221). Hexa SSR repeats were found to be the most abundant in F. *buxicola* (26164) and relatively less abundant in *F. nygamai* (19484) and *F. hostae* (19762).

Hepta SSR repeats were found to be the most abundant in *F.* nygamai (7408) and relatively less abundant in *F. goolgardi* (4422) and *F. sibiricum* (3700).

Table 3. Total repeats of SSRs in Fungal Genome

Organisms	Mono	Di	Tri	Tetra	Penta	Hexa	Hepta
F. falciforme	125	115	436	323	290	467	142
F. nygamai	83	89	273	237	338	409	132
F. verticillioides	89	95	293	255	362	438	142
F. hostae	87	84	277	219	297	409	130
F. aywerte	160	109	324	292	446	497	109
F. solani	91	88	381	272	289	437	118
F. graminum	74	59	276	135	270	525	112
F. buxicola	18	64	410	201	298	575	144
F. sibiricum	53	61	304	253	354	459	81
F. goolgardi	103	102	359	270	419	527	94

SSR relative abundance is the total repeats per 10Mb of sequence analyzed.

Table 4. Relative Density of SSR Repeats

Organisms	Mono	Di	Tri	Tetra	Penta	Hexa	Hepta
F. falciforme	4223	4770	19198	14119	12135	22843	7408
F. nygamai	3196	3882	10969	9844	14328	19484	6854
F. verticillioides	3428	4165	11766	10560	15370	20900	7352
F. hostae	3255	3596	11744	9433	12570	19762	6633
F. aywerte	6289	4147	12652	11339	17315	23379	5424
F. solani	3062	3524	15788	11464	11797	21067	5949
F. graminum	1993	1889	10754	4843	10221	24114	5237
F. buxicola	417	2144	15365	6967	11200	26164	6588
F. sibiricum	1407	1971	11220	8994	13395	21434	3700
F. goolgardi	2928	3307	13202	9146	15817	24497	4422

Repeat lengths per 10 Mb of sequence analyzed.

3.5. Frequency of SSR Motifs

The most grouping SSRs motifs were calculated through excel sheet (Table 5). It was noted that the frequency of SSR motifs varies among different fungal species. For example, *F. falciforme* has a higher frequency of AG/GA and CT/TC motifs, with 170 and 130 occurrences, this suggests that different fungal species have unique patterns of SSR motif frequencies, which may be related to their evolutionary history and biology. One interesting finding from this table is the high frequency of the TA/AT motif in all the fungal genomes analyzed. This motif is the most common two-base-pair repeat unit, with 327 occurrences in *F. falciforme*, 240 in *F. nygamai*, *F. verticillioides*, and 317 in *F. hostae*. This suggests that this motif may be evolutionarily conserved and functionally important in fungi. The most frequent motif identified in the table is TAA, which was found in the genome of *F. falciforme*, with a total of 651 trinucleotide repeats. The second most common motif is TTAA, which is found in the genomes of four different organisms, including *F. falciforme*, *F. nygamai*, *F. verticillioides*, *F. hostae*, *F. aywerte*, *F. solani* with repeat counts ranging from 55 to 584. Another interesting observation in Table 5 is the presence of heptanucleotide repeats. These repeats have been generally considered rare in fungal genomes, with only a few reports in the literature. However, the table shows that *F. falciforme* and *F. solani* both have two heptanucleotide repeats, AAGAGAT and TGCACCT, respectively.

3.6. The Longest SSR Repeats

The longest SSR motifs present in different *Fusarium* species were detected using an excel sheet (table 6), among the species listed, *F. falciforme* has the longest SSR motifs, with a hexanucleotide repeat of AGGCTG, comprising 318 nucleotides. The other species also have varying lengths of SSR motifs, ranging from mono- to hexanucleotide repeats. *F. falciforme* has the longest di-nucleotide repeat of TT, with a length of 195 nucleotides. F. *hostae* has the longest tri-nucleotide repeat of TGC, consisting of 360 nucleotides. *F. hostae* (216) has the longest penta-nucleotide repeat of ATTAT comprising 619 nucleotides. *F. falciforme* has the longest tetra-nucleotide repeat of TAAT, with a length of 293 nucleotides. *F. solani* has the longest mono-nucleotide repeat of T, with a length of 309 nucleotides.

3.7. Conserved SSRs through All Genomes

The number of conserved SSRs of different lengths (Di, Tri, Tetra) was detected in all *Fusarium* species (table 7).

These SSRs (TA, TAA, CTT, TTAA, and TAAT) are known to be highly conserved among the organisms listed, *Fusarium falciforme* has the highest number of conserved SSRs, with 129 Di, 252 Tri, 129 Tetra, and 293 TAAT repeats. Interestingly, *F. aywerte* has a high number of CTT Tetra repeats (235), which is considerably more than any other organism in the table 7. Overall, there is a general trend of having more Tri repeats or Tetra repeats than Di repeats among *Fusarium* species. It is also worth noting that *F. sibiricum* has the lowest number of conserved SSRs among the listed organisms, with only 34 Di, 59 Tri, 63 Tetra, and 28 TAAT repeats. This could be indicative of differences in genetic variability and gene regulation among these species.

Table 5. Most Frequent SSR Motifs in genomes of *Fusarium* species

Organisms	Mono	Di	Tri	Tetra	Penta	Hexa
F. falciforme	TA/AT (327) AG/GA(177) CT/TC (130)	TAA (651) TTA (537) ATA (420) CTT(243) ATT (240) CCT(216)	TTAA (584) TAAT (468) TATT (248)	TTATA (150) TATAA (140) TTATT (125)	TCATCG (90) TCCTCG (78) TTAATT (78)	AAGTTTA (21) TAATTAA (6) ACTTTAA (6)
F.nygamai	TA/AT (240) CT/TC (65) AC/CA(44)	TAA(91) ATA (72) TAA(64) TAT(65) CTT(55) TCT(54)	TTAA (55) TAAT (46) TATT (32) TTTA(29)	TATAA(33) TATTA(28) TAATA(20) TTATA(16)	TTAATA(16) TATTAA(14) TATTTA(12) TCCCTCA(11)	ATAATTA (7) ATAATAA (5) TATATAT (5) TATTATA(4)
F. verticillioides	TA/AT (240) CT/TC (65) AC/CA(44)	TAA(91) ATA (72) TAA(64) TAT(56) CTT(55) TCT(54)	TTAA (55) TAAT (46) TATT (32) TTTA(29)	TATAA(33) TATTA(28) TAATA(20) TTATA(16)	TTAATA(16) TATTAA(14) TATTTA(12) TCCTCA(12)	ATAATTA (7) ATAATAA (5) TATATAT (5) TATTATA(4)
F. hostae	TA/AT (317) CT/TC (52) AC/CA(31)	TTA(118) TAA(102) ATA(82) CTT(76) TCC(73) TGA(53)	TTAA (41) TAAT (57) ATAA(56) TTTA(42)	TATAA(26) TATTA(23) ATTTG(22) TTATA(21)	TTAATA(16) CCTCAT(11) GAGGAT(10) TATAAT(10)	TTATTTC (7) TTATATA (6) TATTTTA (6) ATAAGAA (6)
F. aywerte	TA/AT (247) CT/TC (68) AG/GA(46)	TCC(73) TTA(102) TCT(50) TAA(50) TTC(44) TCA(40)	TAAT (57) TACC(30) TTAA (55) TAGG(28)	TACAG (19) TTATA(17) TAATA(13) CCAGT(11)	TAACT(37) TTAATA(12) TCTTCC(10) TCCTCA(9)	TTAACTT(8) ATATAAG(5) ATAATTA(4) ATATAAT(3)
F. solani	TA/AT (193) CT/TC (92) AC/CA(83)	TAA(105) CTT(89) TCC(86) CCT(79) TTA(77) CAG (57)	TTAA(74) TAAT(65) ATTA(40) AATT(38)	TAATA(17) AATAA(15) CCCAT(15) CCATC(13)	TCCTCC(11) TCCTCG(11) CTCTTC(11) GACGAG(11)	AAGTTTA(17) ACTTTAA(6) TAATATA(5) TAACTCA(4)
F. graminum	CT/TC (59) TA/AT (48) AC/CA(35)	CTT(51) TTC(43) CAG(38) CAT(36) CAA(360 TCT(35)	TACC (13) TTCT(12) TTTC(12)	TCTTT(5) TGTCT(9) CTCTT(9)	CATCGT(10) TCCTCT(9) TGATGG(9)	TTCTCTC(3) TTGTTGG(3) TTCCATA(2) TTCCTCC(2)
F. buxicola	TC/CT(65) GT/TG(48) TA/AT(41)	CTT(72) CTC (57) CCT(57) TCC(56) TGC(55)TCT(48)	TACC(25) TAGG(20) AGGT(19) CTAC(16)	GTACT(13) TACTG(10) TCTCT(9) CAGCA(9)	TCGTCC(13) TCATCC(10) TCCTCT(9) GACGAC(9)	TACTCTA(3) TTTTATA(2) ATGAGAG(2) ATAATAA(2)
F. sibiricum	TC/CT(85) AC/CA(45) TA/AT(35)	TTC(76) CTT(63) TCT(46) CAG(40) TCA(39) TCG (39)	TAGG(33) TTTC(26) CTAC(21) TTCT(21)	CTTTT(13) AGACA(12) ATTGA(12) CAAGA(11)	GACGAG(14) ATGTAG(11) GACGAA(8) CTCTTC(8)	CTTCTCT(3) TGTTATG(3) TATGGAT(3) GCCGACC(2)
F. goolgardi	TC/CT(101) TA/AT(90) GT/TG(54)	CTT(67) TTC(63) TCT(57) GAT(42) GTT(40) TCA(38)	TACC(28) CTAC(25) TACA(20) TAGG(18)	TTTCT(15) TTTTC(14) ACAGC(13) CAGCA(11)	TCATCC(13) TCGTCA(11) GAAGAT(9) TCATCG(7)	TTCTCTC(5) GATAAGA(3) TATATAA(3) TCTTTTC(2)

Numbers within parentheses indicate the total count (occurrence) of each motif.

Table 6. Longest SSR Motifs in Fungal Genomes

Organism	Mono	Di	Tri	Tetra	Penta	Hexa
F. falciforme	G (148)	TT(195)	CTG(268)	TTAA (280)	ATCAG (161)	AGGCTG (318)
	A (129)	CT(152)	TAA (252)	TAAT (293)	CACAAA (157)	CTGGCT (290)
	A (126)	TA(129)	ATA (250)	TATT (207)	ATATA (152)	CTCCGA (277)
F. nygamai	A (176)	TC (96)	CGG(99)	TTAA(53)	TAAGT(207)	TATTAC (99)
	A (126)	TA (94)	TCA(98)	TTAA(48)	AATTA(196)	AGCCGA (98)
	T (106)	AT (90)	GTC(98)	AGTT(26)	AATTA(178)	GATGAG (97)
F. verticillioides	A (176)	TC (194)	ATT(187)	TTTA(260)	TAAGT(207)	TTACTC(393)
	A (126)	AT (113)	GCC(151)	TTTA (254)	AATTA(196)	AATTTA(318)
	T (106)	TA(112)	AGC(149)	ATTA (210)	AATTA(178)	TTAACC (262)
F. hostae	A (149)	TA(151)	TGC (366)	CTTC (260)	ATTAT (216)	CTATTA (276)
	A (139)	TA(141)	TTC (259)	ATCG (215)	AATTA (191)	CTTCTT (242)
	A (116)	TA(140)	TTA (234)	ATCG (203)	TATAA (169)	AGGAAG (226)
F. aywerte	A (162)	AC145	CTT(235)	AATT(282)	TAATA(161)	GCTGTT(619)
	A (148)	TA128	GCT(208)	AATT(262)	TTATA(124)	ACGACG(241)
	A (139)	TA108	CAG(152)	AATT(227)	AGGTA(123)	CCAGCA(206)
F. solani	T(306)	TA(117)	TAA(105)	TTAA(74)	TAATA(17)	TCCTCC(11)
	A(156)	AT(76)	CTT(89)	TAAT(65)	AATAA (15)	TCCTCG(11)
	C(19)	TC(50)	TCC(86)	ATTA(40)	CCCAT(15)	CTCTTC(11)
F. graminum	T(134)	TC30	CTT(51)	TACC(13)	TCTTT(9)	CATCGT(10)
	A(117)	CT29	TTC(43)	TTCT(12)	TGTCT(9)	TCCTCT(90)
	C(7)	AT28	CAG(38)	TTTC(12)	CTCTT(9)	TGATGG(9)
F. buxicola	T(31)	CT37	CTT(72)	TACC(25)	GTACT(13)	TCGTCC(13)
	A(21)	TC28	CTC(57)	TAGG(20)	TACTG(10)	TCATCC(10)
	C(6)	GT25	CCT(57)	AGGT(19)	TCTCT(9)	TCCTCT(9)
F. sibiricum	T(131)	CT(44)	TTC(76)	TAGG(33)	CTTTT(13)	GACGAG(14)
	A(76)	TC(41)	CTT(63)	TTTC(26)	AGACA(12)	ATGTAG(11)
	C(4)	TG(26)	TCT(46)	CTAC(21)	ATTGA(12)	GACGAA(8)
F. goolgardi	T(191)	CT(55)	CTT(67)	TACC(28)	TTTCT(15)	TCATCC(13)
	A(151)	TA(51)	TTC(63)	CTAC(25)	TTTTC(14)	TCGTCA(11)
	C(12)	TC(46)	TCT(57)	TACA(20)	ACAGC(13)	GAAGAT(9)

Values within parentheses indicate the overall base pair lengths for each motif.

Table 7. Conserved SSR Motifs in all genomes of *Fusarium* species

Organism	Di	Tri	Tetra
F. falciforme	TA(129)	TAA (252) CTT(129)	TTAA (280) TAAT (293)
F. nygamai	TA (94)	TAA(115)CTT(112)	TTAA(48) TAAT (133)
F. verticillioides	TA(112)	TAA(115) CTT(112)	TTAA(140) TAAT (133)
F. hostae	TA(151)	TAA(123) CTT(96)	TTAA(122) TAAT (133)
F. aywerte	TA(128)	TAA(111) CTT(235)	TAAT (139) TTAA(107)
F. solani	TA(117)	TAA(105) CTT(89)	TAAT (276) TTAA(74)
F. graminum	TA(62)	TAA (69) CTT(51)	TTAA(60) TAAT (36)
F. buxicola	TA(66)	TAA (57) CTT(72)	TAAT (76) TTAA(52)
F. sibiricum	TA(34)	TAA(59) CTT(63)	TTAA(37) TAAT (28)
F. goolgardi	TA(34)	TAA(69) CTT(67)	TTAA(70) TAAT(51)

Values within parentheses indicate the overall base pair lengths for each motif.

4. Discussion

Examining the genomes of ten *Fusarium* species for complete simple sequence repeats (SSRs) and interspersed repeats using the Repeat Masker tool offers valuable understanding of the genomic features of these fungi.

4.1. Genome Size and GC Content

The comparison summary in Table 1 underlines the differences in genome size and GC content among the ten species, which could hold significant consequences for their biology and evolution. The genome size of *Fusarium* species analyzed in this research ranges from 34.84 Mb to 59.29 Mb. These differences in genome size can be ascribed to variations in genomic repeats elements, gene content, and chromosome number across the species [14-16]. For example, *Fusarium oxysporum* possesses a relatively small genome size of 36.39 Mb, which can be ascribed to its compact genome structure with genomic repeats elements compared to other *Fusarium* species. On other hand, *Fusarium falciforme* has an expanded genome size of 59.29 Mb, which is probably due to the presence of numerous DNA repeats elements. In addition to genome size, the GC content of *Fusarium* species also exhibits substantial variation. Interestingly, some *Fusarium* species with similar genome sizes show different GC contents, as it is noted in *Fusarium aywerte* and *Fusarium sibiricum*, emphasizing the complexity of evolution of DNA within this genus.

4.2. Frequency of Various Genomic Elements

The information in Table 2 shows considerable differences in the frequency of various genomic elements among the species, indicating differences and similarities at the genomic level. LTRs are a type of retrotransposon, and the absence of LTRs in *F. verticillioides* and the highest frequency of LTRs in *F. nygamai* and *F. solani* are intriguing findings. The occurrence or lack of LTRs can influence genome stability and gene expression. The prevalence of LINE1 and LINE2 retrotransposons in the analyzed *Fusarium* genomes aligns with earlier studies, where LINEs are recognized as the most prevalent transposable elements in eukaryotic genomes. The differences in frequency of LINE1 and LINE2 among the *Fusarium* species signify the ever-changing nature of genome evolution in this genus [16-18].

The frequency of SINEs and DNA transposons also exhibits considerable variation among the *Fusarium* species. SINEs and DNA transposons can trigger genomic instability and result in mutations and genetic diversity. The variation in frequency of these elements among the species implies that they might have distinct evolutionary

histories and selective pressures that have shaped their genome composition [16-18]. Simple repeats are short nucleotide motifs susceptible to mutations, which can generate genetic diversity and morphological characteristics variation. The high frequency of simple repeats in *F. aywerte* and *F. falciforme* suggests that this species has a comparatively unstable genome that is prone to mutations and genetic diversity [15]. Low complexity regions are parts of the genome with a high content of a single nucleotide or simple nucleotide repeats, which can affect gene expression and genome stability. The variation in frequency of low complexity regions among the *Fusarium* species indicates that they might have distinct genome stability and gene expression control mechanisms [16,19].

4.3. Total repeats of SSRs

The findings presented in table 3 reveal that the number of repeats tends to increase with the size of the repeat unit, with tri- and tetra-repeats being more prevalent than mono- and di-repeats, and penta-, hexa-, and hepta-repeats being even more frequent. The most widespread repeat type is the tri-repeat, followed by the hexa-repeat and penta-repeat, while the mono-repeat is the least common. These observations imply that the evolutionary processes of repeat expansion and contraction vary based on the repeat unit size [19,20].

Interestingly, the number of repeats can differ significantly within a single species, emphasizing the significance of investigating repeat number variation at both the species and repeat type levels. For instance, *F. aywerte* has the highest total number of repeats overall, but the number of repeats within each repeat type varies significantly [15]. The results also show that some species exhibit a clear preference for certain repeat types. For example, *F. nygamai* and *F. graminum* have relatively low numbers of mono- and di-repeats, but relatively high numbers of hexa-repeats, while *F. buxicola* has an unusually high number of hexa-repeats [15, 21]. These observations are crucial for understanding the genetic diversity and evolution of *Fusarium* species, as well as for developing effective strategies for their control. For example, identifying repeat type preferences could offer insights about the potential functional roles of these repeats in the genome [15]. Moreover, the variation in repeat number within species suggests that these species may have unique mechanisms for regulating repeat expansion and contraction, which could be targeted for future research [19].

4.4. Relative Density of SSR Repeats

The analysis conducted in this study provides insights into the relative density of SSRs in different *Fusarium* species. The results in Table 4 show that mono SSR repeats are the most abundant in all organisms listed, with the highest density observed in *F. aywerte* and the lowest in *F. sibiricum* [15, 21]. Additionally, di SSR repeats were found to be prevalent in all species, with the highest density observed in *F. nygamai* and the lowest in *F. buxicola*. These results imply that mono and di SSR repeats may play a significant role in the genetic diversity and evolution of *Fusarium* species [15, 21]. Furthermore, the study revealed that the relative abundance of tri, tetra, penta, hexa, and hepta SSR repeats varied among different Fusarium species. For example, tri SSR repeats were the most abundant in *F. aywerte* and *F. solani* but relatively less abundant in *F. graminum*. Similarly, tetra SSR repeats were relatively less abundant in all organisms, while penta SSR repeats were more abundant in *F. aywerte* and *F. buxicola* but less abundant in *F. graminum* [19, 22]. Moreover, hexa SSR repeats were the most abundant in *F. buxicola*, while hepta SSR repeats were the most abundant in *F. nygamai*. These results propose that different SSR repeat types may have distinct functional roles in the genome of *Fusarium* species, helping to better understand the molecular mechanisms underlying their pathogenicity [14, 19, 22].

4.5. Frequency of SSR Motifs

Table 5 presents evidence that suggests that SSR motifs are non-randomly distributed across fungal genomes, and instead display selectivity towards particular repeat units. The study revealed that the AG/GA and CT/TC motifs exhibited a comparatively high frequency in *F. falciforme*, whereas the TA/AT motif was observed to be significantly prevalent in all the fungal species that were examined. The aforementioned observation is consistent with prior studies that have demonstrated the higher frequency of specific repetitive patterns in certain genomic regions, including gene coding regions and promoter regions [14, 15, 21]. Table 5 presents a noteworthy observation regarding the variability of the number of repetitions within a specific motif. The motif AG/GA exhibited a spectrum of repeat counts ranging from 2 to 17 in *F. falciforme*, whereas the motif TAA displayed a range of repeat counts from 3 to 651. The aforementioned observation implies that distinct fungi may possess diverse mechanisms for producing and preserving SSR motifs, which could potentially influence their evolutionary dynamics [17, 20]. It is noteworthy that certain motifs that have been identified in Table 5 have previously been linked to the regulation and expression of genes. The motif TCT has been identified to play a role in transcriptional regulation, whereas the motif AG has been hypothesized to be involved in splicing regulation. The prevalence of SSR motifs throughout fungal genomes may hold significant functional implications for the regulation of gene expression. The potential impact of specific fungal species exhibiting a high frequency of certain motifs on molecular identification and diagnostics is a matter of consideration [14]. The identification of species and the monitoring of fungal pathogen dissemination could potentially be facilitated through the utilization of specific

SSR motifs as molecular markers. The findings presented in Table 5 underscore the significance of taking into account SSR motifs as a genetic determinant for variability in fungal genomes. Subsequent research endeavors may endeavor to explore the functional significance of particular motifs and assess their relevance in the realm of fungal biology and evolution, as suggested by previous literature [15].

4.6. The Longest SSR Repeats

Table 6 presents evidence of the existence of long SSR motifs across diverse *Fusarium* species. It is noteworthy that *F. hostae* exhibits the longest hexanucleotide repeat motif of AGGCTG, which extends over a length of 318 nucleotides. The length of the hexanucleotide motifs identified in this species is comparatively greater than those observed in other species, such as *F. aywerte* (139) and *F. graminum* (121). The variability in the length of simple sequence repeat (SSR) motifs across different species may serve as an indicator of evolutionary modifications and can be utilized for the construction of DNA markers, as reported in previous studies [15,23]. The species *F. falciforme* exhibited the longest di-nucleotide repeat motif of TT, measuring 195 nucleotides in length. The length of the di-nucleotide motifs identified in this species is notably greater than those observed in other species, such as *F. verticillioides* (47) and *F. sibiricum* (68). The existence of long SSR motifs in fungal genomes may have functional implications, including the potential to influence gene expression and regulation [15]. *F. hostae* possesses the longest tri-nucleotide repeat motif of TGC, encompassing 360 nucleotides. The tri-nucleotide motif identified in the table is the most extensive among the listed species. *F. falciforme* was found to have the longest tetra-nucleotide repeat motif of TAAT, which spans 293 nucleotides. It is noteworthy that *F. solani* possesses the lengthiest mono-nucleotide repeat motif of T, measuring 309 nucleotides [15, 23]. It is noteworthy to mention that the length of Simple Sequence Repeats (SSR) motifs can exhibit variability even among conspecific individuals. *F. falciforme* exhibits *hexanucleotide* repeat motifs that span a range of 12 to 318 nucleotides in length. The aforementioned suggests that the conservation of SSR motifs in terms of length within a genome is not a certainty, and that factors such as genetic selection or other processes may have an impact on their length [24].

4.7. Conserved SSRs through All Genomes

The identification of conserved Simple Sequence Repeats (SSRs) in *Fusarium* species, as presented in Table 7, suggests that these repetitive DNA sequences serve a vital functional purpose in these fungal species. The considerable amount of preserved Simple Sequence Repeats (SSRs) in *F. falciforme* implies that this particular organism possesses a higher capacity for genetic diversity and adaptability in comparison to other members of the *Fusarium* genus. The observed phenomenon may provide an explanation for the comparatively elevated count of conserved simple sequence repeats (SSRs) of varying lengths in *F. falciforme*, in relation to the other organisms listed. The observation that *F. aywerte* exhibits a substantial count of CTT Tetra repeats implies that this particular repetition may hold importance for the biology and evolutionary processes of this species. The repetition in question may potentially serve a function in gene regulation or in maintaining the stability of the genome in *F. aywerte*, as indicated by previous research [15]. The observed pattern of *Fusarium* species exhibiting a higher frequency of Tri or Tetra repeats in comparison to Di repeats is consistent with prior research conducted on other organisms. The observed phenomenon could potentially be attributed to the comparatively heightened stability and diminished mutation rates of Tri and Tetra repeats in contrast to Di repeats. The previously mentioned observation may indicate the functional significance of Tri and Tetra nucleotide repeats in the regulation of genes and the structure of proteins [15]. The low number of conserved SSRs in *F. sibiricum* implies that this organism may possess restricted genetic variability and adaptability in comparison to other Fusarium species. The observed phenomenon may be attributed to the biological and ecological characteristics *of F. sibiricum*, which exhibit a higher degree of specialization and adaptation to a narrower range of environmental conditions. The identification of conserved SSR motifs (TA, TAA, CTT) [25] in *Fusarium* species represents a valuable asset for forthcoming investigations on the functional implications and evolutionary dynamics of these repetitions. The indication of conserved Simple Sequence Repeats (SSRs) within these organisms implies that these repetitive DNA sequences may hold significance in their biology and ability to adapt, as per previous research [15]. Additional research is required to investigate the functional significance of conserved simple sequence repeats (SSRs) in *Fusarium species*, as well as their potential utility in the management of plant diseases in the fields of agriculture and biotechnology.

5. Conclusions

This study noticed differences in various genomic features, including genome size, GC content, and repeat frequency after examining the genomes of ten *Fusarium* species using the RepeatMasker software. A high frequency of simple repeats in *F. aywerte* and *F. falciforme* indicates a less stable genome prone to mutations and genetic diversity. Our hypothesis, which suggested that these genomic variations might be linked in this *Fusarium* genus to biology, evolution, and potential functional roles,

was tested and proven based on the observed differences among the species. The study presents valuable insights into the functional roles of these repeats in the genome and the unique mechanisms regulating repeat expansion and contraction by identifying repeat type preferences and variations among species. Our findings are very important for understanding the genetic diversity and evolution of *Fusarium* species and also for developing effective control strategies.

REFERENCES

[1] Nelson, P. E., Dignani, M. C., & Anaissie, E. J., "Taxonomy, biology, and clinical aspects of *Fusarium* species," Clinical Microbiology Reviews, vol. 7, no. 4, pp. 479-504, Oct. 1994, doi: 10.1128/cmr.7.4.479, PMCID: PMC358338, PMID: 7834602.

[2] Moretti, A., "Taxonomy of *Fusarium* genus: A continuous fight between lumpers and splitters," Zbornik Matice Srpske Za Prirodne Nauke, no. 117, pp. 7-13, 2009, doi: 10.2298/ZMSPN0917007M.

[3] Watanabe, M., Yonezawa, T., Lee, K., Kumagai, S., Sugita-Konishi, Y., Goto, K., & Hara-Kudo,Y., "Molecular phylogeny of the higher and lower taxonomy of the *Fusarium* genus and differences in the evolutionary histories of multiple genes," BMC Evolutionary Biology, vol. 11, no. 1, p. 322, Nov. 2011, doi: 10.1186/1471-2148-11-322, PMCID: PMC3270093, PMID: 22047111.

[4] Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., et al.,"Ten things you should know about transposable elements," Genome Biology, vol. 19, no. 1, p. 199, Nov. 2018, doi: 10.1186/s13059-018-1577-z.

[5] Bucher, E., Reinders, J., & Mirouze, M., "Epigenetic control of transposon transcription and mobility in Arabidopsis," Current Opinion in Plant Biology, vol. 15, no. 5, pp. 503-510, Nov. 2012, doi: 10.1016/j.pbi.2012.08.006, PMID: 22940592.

[6] Kapitonov, V. V., & Jurka, J., "A universal classification of eukaryotic transposable elements implemented in Repbase," Nature Reviews. Genetics, vol. 9, no. 5, pp. 411-2, author reply 414, May 2008, doi: 10.1038/nrg2165-c1, PMID: 18421312

[7] Richard, G. F., Kerrest, A., & Dujon, B., "Comparative genomics and molecular dynamics of DNA repeats in eukaryotes," Microbiology and Molecular Biology Reviews, vol. 72, no. 4, pp. 686-727, Dec. 2008, doi: 10.1128/MMBR.00011-08, PMCID: PMC2593564, PMID: 19052325.

[8] Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M., "Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics," Aquaculture, vol. 255, no. 1-4, pp. 1-29, May 2006, doi: 10.1016/j.aquaculture.2005.11.031.

[9] Jeffreys, A. J., "Hypervariable 'minisatellite' regions in human DNA," Nature, vol. 314, pp. 67-73, 1985.

[10] Andreassen, R., "Mutation rate in the hypervariable VNTR g3 (D7S22) is affected by allele length and a flanking DNA sequence polymorphism near the repeat array," Am. J. Hum. Genet, vol. 59, pp. 360-367, 1996.

[11] Amos, W., "Mutation biases and mutation rate variation around very short human microsatellites revealed by human-chimpanzee-orangutan genomic sequence alignments," Journal of Molecular Evolution, vol. 71, no. 3, pp. 192-201, Sep. 2010, doi: 10.1007/s00239-010-9377-4, PMID:20700734, S2CID: 1424625.

[12] Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., et al., "Abundant contribution of short tandem repeats to gene expression variation in humans," Nature Genetics, vol. 48, no. 1, pp. 22-29, Jan. 2016, doi: 10.1038/ng.3461, PMCID: PMC4909355, PMID: 26642241.

[13] Smit, S. A., Hubley, R., & Green, P., "RepeatMasker Open-4.0 (Version 4.1.1)," http://www.repeatmasker.org, 2013-2015.

[14] Wöstemeyer, J., and Kreibich, A., "Repetitive DNA elements in fungi (Mycota): impact on genomic architecture and evolution," Current genetics, vol. 41, no. 4, pp. 189-198, 2002, doi: 10.1007/s00294-002-0306-y.

[15] Karaoglu, H., Lee, C. M., Meyer, W., "Survey of simple sequence repeats in completed fungal genomes," Molecular biology and evolution, vol. 22, no. 3, pp. 639-649, 2005, doi: 10.1093/molbev/msi057.

[16] Dodsworth, S., Chase, M. W., Kelly, L. J., Leitch, I. J., Macas, J., Nov ák, P., Piednoël, M., Weiss-Schneeweiss, H., and Leitch, A. R., "Genomic Repeat Abundances Contain Phylogenetic Signal," Systematic Biology, vol. 64, no. 1, pp. 112-126, Jan. 2015, doi: 10.1093/sysbio/syu080.

[17] Fedoroff, N. V., "Transposable elements, epigenetics, and genome evolution," Science, vol. 338, pp. 758-767, 2012.

[18] Kelly, L. J., and Leitch, I. J., "Why size really matters when sequencing plant genomes," Plant Ecol. Div., vol. 5, pp. 415-425, 2011.

[19] Harr, B., Todorova, J., & Schlötterer, C., "Mismatch repair-driven mutational bias in D. melanogaster," Mol. Cell, vol. 10, pp. 199-205, 2002.

[20] Kruglyak, S., Durrett, R. T., Schug, M. D., & Aquadro, C. F., "Distribution and abundance of microsatellites in the yeast genome can be explained by a balance between slippage events and point mutations," Mol. Biol. Evol., vol. 17, pp. 1210-1219, 2000.

[21] Mathur, M., "A Comparative Study of Various SSRs Identification Tools Using Aspergillus Fumigatus Chromosome Sequences," J. Bioinfo. Comp. Genom., vol. 3, pp. 1-13, 2020.

[22] Toth, G., Gaspari, Z., & Jurka, J., "Microsatellites in different eukaryotic genome: survey and analysis," Genome Res., vol. 10, pp. 1967-1981, 2000.

[23] Zarlenga, D. S., Aschenbrenner, R. A., & Lichtenfels, J. R., "Variations in Microsatellite Sequences Provide Evidence for Population Differences and Multiple Ribosomal Gene Repeats within Trichinella pseudospiralis," The Journal of Parasitology, vol. 82, no. 4, pp. 534-538, 1996, doi: 10.2307/3283777.

[24] Harr, B., and Schlötterer, C., “Long microsatellite alleles in Drosophila melanogaster have a downward mutation bias and short persistence times, which cause their genome-wide under-representation,” Genetics, vol. 155, pp. 1213-1220, 2000.

[25] Zamanmirabadi, A., Hemmati, R., Dolatabadian, A., & Batley, J., "Status of SSR, cSSR, iSSR, and VNTR motifs in Leptosphaeria maculans based on high throughput sequencing data," Mycologia Iranica, vol. 8, no. 2, pp. 95-107, 2021, doi: 10.22043/MI.2022.359536.