

Article

## Development and Characterization of 18 Novel EST-SSRs from the Western Flower Thrips, *Frankliniella occidentalis* (Pergande)

Xian-Ming Yang, Jing-Tao Sun, Xiao-Feng Xue, Wen-Chao Zhu and Xiao-Yue Hong \*

Department of Entomology, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China;  
E-Mails: zqbxming@163.com (X.-M.Y.); 2009202023@njau.edu.cn (J.-T.S.);  
xfxue@njau.edu.cn (X.-F.X.); zzduke@126.com (W.-C.Z.)

\* Author to whom correspondence should be addressed; E-Mail: xyhong@njau.edu.cn;  
Tel.: +86-25-84395339; Fax: +86-25-84395339.

Received: 6 December 2011; in revised form: 27 February 2012 / Accepted: 28 February 2012 /  
Published: 5 March 2012

---

**Abstract:** The western flower thrips, *Frankliniella occidentalis* (Pergande), is an invasive species and the most economically important pest within the insect order Thysanoptera. For a better understanding of the genetic makeup and migration patterns of *F. occidentalis* throughout the world, we characterized 18 novel polymorphic EST-derived microsatellites. The mutational mechanism of these EST-SSRs was also investigated to facilitate the selection of appropriate combinations of markers for population genetic studies. Genetic diversity of these novel markers was assessed in 96 individuals from three populations in China (Harbin, Dali, and Guiyang). The results showed that all these 18 loci were highly polymorphic; the number of alleles ranged from 2 to 15, with an average of 5.50 alleles per locus. The observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosities ranged from 0.072 to 0.707 and 0.089 to 0.851, respectively. Furthermore, only two locus/population combinations (WFT144 in Dali and WFT50 in Guiyang) significantly deviated from Hardy–Weinberg equilibrium (HWE). Pairwise  $F_{ST}$  analysis showed a low but significant differentiation ( $0.026 < F_{ST} < 0.032$ ) among all three pairwise population comparisons. Sequence analysis of alleles per locus revealed a complex mutational pattern of these EST-SSRs. Thus, these EST-SSRs are useful markers but greater attention should be paid to the mutational characteristics of these microsatellites when they are used in population genetic studies.

**Keywords:** *Frankliniella occidentalis*; EST-SSRs; expressed sequence tags; microsatellite; population genetics

---

## 1. Introduction

The western flower thrips, *Frankliniella occidentalis* (Pergande), is the most economically important pest within the insect order Thysanoptera, which includes more than 5500 described species [1]. *F. occidentalis* causes enormous damage by directly feeding on greenhouse vegetable and ornamental crops and by transmitting plant-pathogenic tospoviruses [2]. *F. occidentalis* is endemic to North America in an area west of the Rocky Mountains from Mexico to Alaska [3]. Since the late 1970s, *F. occidentalis* has rapidly invaded most countries throughout the world where it not only causes severe economic losses but also threatens endemic invertebrates and associated ecosystems [4]. In order to control *F. occidentalis*, it is first necessary to know its genetic diversity, population structure and invasion history. Genetic tools, such as microsatellites markers, can reveal the origin of newly established populations, their genetic makeup and their routes of migration [5,6].

Microsatellites, or simple sequence repeats (SSRs), consist of tandemly repeated motifs that are 1–6 bp in length, and they are widely distributed throughout the eukaryotic genomes [7]. Conventionally, two models of mutations have been considered for microsatellites, the stepwise mutational model (SMM) and the infinite allele model (IAM). The SMM states that all mutational events involve a change in a single repeat only. The IAM assumes that every mutation results in the creation of a new allele [8]. The mutational mechanism of microsatellites is still under debate though it appears most likely to be slippage events during DNA replication [9]. Several other mechanisms may also be responsible for the generation of new alleles, e.g., insertions/deletions (indels) in the flanking region [10]. Matsuoka showed that the IAM model was appropriate for maize microsatellites mutated in the flanking regions [10]. Knowledge of the mutational pattern of one specific SSR could facilitate the selection of appropriate mutation model and combinations of markers in the population genetic studies. Currently, due to their codominant inheritance, highly polymorphic, easy detection by polymerase chain reaction (PCR) and broad distribution in the genome, microsatellites/SSRs are widely used for population genetic studies [11]. Ascunce *et al.* have used a large number of SSRs to investigate the global invasion route of the fire ant *Solenopsis invicta* [6]. However, population genetic studies of *F. occidentalis* have been hampered by a lack of polymorphic molecular markers. Presently, only 6 polymorphic microsatellites of *F. occidentalis* are known [12]. Recently, an enormous number of ESTs (expressed sequence tags) of *F. occidentalis* have become available in the public sequence database [13], and can be exploited to identify markers inexpensively. Hence, we isolated and characterized 18 novel EST-SSRs for *F. occidentalis*. These EST-SSRs will allow researchers to investigate the genetic diversity and population genetic structure of *F. occidentalis* in its native and invasive range and trace its global invasion history.

## 2. Results and Discussion

### 2.1. Characteristics of *F. occidentalis* EST-SSRs

We obtained 309 sequences containing SSRs by MicroSAteLLite (MISA) [14] analysis. Among these sequences, five contained two different SSRs and three of these were compound microsatellites (Table 1). The EST-SSR frequency (1SSR/24.1 kb) of *F. occidentalis* was smaller than that of brown planthopper (1SSR/13.0 kb; [15]), pea aphid (1SSR/3.0 kb; [16]) and several other insects (~1SSR/1 kb in fly, silkworm and mosquito; [17]) and was comparable to some crops (1SSR/23.80 kb in soybean and 1SSR /28.32 kb in maize; [18]). The most abundant repeat motif class was dinucleotide repeats (DNRs, 265/314). The AC/GT (41.5%) motif was the most common among DNRs, followed by AG/CT (31.7%), AT/AT (22.3%) and CG/CG (4.5%). The classification of repeats into classes was carried out according to the method of Jurka and Pethiyagoda [19]. For example, (AC)*n*, (CA)*n*, (TG)*n* and (GT)*n* were considered as the same class considering complementary sequences and/or different reading frames. Other repeat motifs, including trinucleotide, tetranucleotide, pentanucleotide, were also observed, albeit infrequently (49/314) (Table 1).

**Table 1.** Frequency and distribution of SSRs in the analyzed *Frankliniella occidentalis* ESTs.

Repeats	Number of repeat motif ( <i>n</i> )					Total
	4	5	6	7	8	
AC/GT	-	89	14	4	3	110
AG/CT	-	67	11	6		84
AT/AT	-	54	4	1		59
CG/CG	-	12				12
AAC/GTT	-	7	1			8
AAG/CTT	-	1				1
AAT/ATT	-	5	1			6
ACC/GGT	-	4	1			5
ACT/AGT	-	1				1
AGC/CTG	-	5				5
AGG/CCT	-	1	1			2
ATC/ATG	-	1				1
AAAC/GTTT	3					3
AAAT/ATTT	3					3
AACC/GGTT	2					2
AATC/ATTG	2					2
AATG/ATTC	3					3
ACAT/ATGT	1					1
ACTG/AGTC	1					1
AGAT/ATCT	3					3
ATGC/ATGC	1					1
ACGGC/CCGTG	1					1
NN(DNR)	-	222	29	11	3	265
NNN(TNR)	-	25	4			29
NNNN(TTNR)	19					19
NNNNN(PNR)	1					1

NN(DNR) is dinucleotidic repeats; NNN(TNR) is trinucleotidic repeats; NNNN(TTNR) is tetranucleotidic repeats; NNNNN(PNR) is pentanucleotidic repeats.

Of the primer pairs designed from 122 sequences suitable for primer design, 72 amplified the expected products, 50 yielded larger or no products. Finally, 18 primer pairs revealed polymorphism (Table 2), the remaining (54) were either monomorphic or amplified poorly. For these 18 selected ESTs, homology searches with the BLASTX against the NCBI nr database found three ESTs with significant hits to insect genes at an E-value cutoff level of  $1e-5$  (Table 2). No hit was found for any of the other 15 ESTs. Sequence length variation in coding sequences is rare. Examination of the three sequences with significant blast hits suggested that SSR sequences may be located on either 5'- or 3'- UTR (untranslated regions). The 15 remaining ESTs with unknown function may also originate from non-coding regions. These selected ESTs seem unlikely to come from non-insect sources because of the high amplification rates (approximately 99.5%) of these EST-SSRs across the 96 *F. occidentalis* samples. When analyzed all the published *F. occidentalis* ESTs, 17 of the 18 selected ESTs were singletons, the remaining one (GT306150) was part of a larger contig which contained only two ESTs. The low copy number of these ESTs suggested that they seem unlikely to be repetitive sequences in the nuclear genome. When considering all three populations (Table 3), 99 alleles were identified from 18 markers, the number of alleles ( $N_a$ ) ranged from 2 to 15, with an average of 5.50 alleles per locus. The observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities ranged from 0.072 to 0.707 and 0.089 to 0.851, respectively. The PIC values ranged from 0.088 to 0.860, with an average of 0.476 (Table 2).

**Table 2.** Characteristics of 18 new EST-SSRs in *Frankliniella occidentalis*.

Locus	Genbank number (Sequence length)	Putative gene function <sup>a</sup>	Repeat motif	Primer sequence (5'-3')	Size range (bp)	T <sub>a</sub> (°C)	H <sub>O</sub>	H <sub>E</sub>	N	N <sub>a</sub>	PIC
<b>WFT20</b>	GT306150 (292bp)	XP_001945214.1 PREDICTED: uncharacterized protein C14orf138-like [ <i>Acyrtosiphon pisum</i> ] 4e-08	(AC)6	F: CGTAGCCGCCAAACTGTT R: CCTTCCAATTCAAATTCCT	144–153	52	0.654	0.657	96	5	0.609
<b>WFT24</b>	GT303793 (745bp)	Unknown	(TG)6	F: ACGAAGTTTGGTTTGGGTGG R: AAGTTTCTCCGCTCATTTC	208–214	52	0.257	0.290	96	2	0.258
<b>WFT25</b>	GT303588 (608bp)	Unknown	(GA)7	F: CACCAGTCGCGTTCATTGA R: GCCTCCAGCAGCACAAGTA	96–149	52	0.707	0.851	96	14	0.860
<b>WFT28</b>	GT303349 (784bp)	Unknown	(TA)6	F: GGGCTTGAAATAATGTTCTG R: GTAAATAAATCAGTGGAGGGT	91–95	52	0.241	0.224	96	3	0.230
<b>WFT37</b>	GT302424 (759 bp)	Unknown	(TTA)5	F: GCATACCCTGTGAACGAGTG R: ACAGAAGCAAATGTCTACCTGA	213–222	52	0.372	0.410	94	4	0.384
<b>WFT50</b>	GT310239 (785 bp)	Unknown	(TG)7	F: CGGAGTGAGCAGGAGTTGT R: TTGCCCTACCAAAAATATGA	171–181	52	0.313	0.458	95	6	0.427
<b>WFT51</b>	GT310133 (741bp)	XP_002425957.1  abrupt protein, putative [ <i>Pediculus humanus corporis</i> ] 7e-76	(TG)8	F: GTACGCAGGAGAAGTAAATG R: ACAAATCCAGATGGCAACC	297–305	52	0.623	0.590	96	5	0.555
<b>WFT64</b>	GT311293 (518 bp)	Unknown	(TCC)6	F: CTTTTCGGATTCTCCTTCG R: GGAGACCTGATTCACCGTATG	243–250	52	0.590	0.539	95	4	0.443
<b>WFT66</b>	GT305093 (175 bp)	Unknown	(ACT)5	F: AACTTAGGAAGAAAGACTGTAGA R: TGTTTACGCACGCACGCAT	113–116	52	0.505	0.476	96	3	0.432
<b>WFT83</b>	GT304065 (616 bp)	Unknown	(TA)5	F: GGAGGTACTGACTAAAGCATG R: GGGACAGACAAAACAGGAAA	172–176	52	0.462	0.451	96	4	0.394
<b>WFT87</b>	GT303951 (830 bp)	Unknown	(TG)5	F: GGTCTGAACTGTATGGGATG R: CAGGACCCTAGTATGTAAGAAA	259–277	52	0.346	0.391	94	7	0.394
<b>WFT98</b>	GT299180 (436 bp)	Unknown	(TG)5	F: GGGCAGTTTGCTCTTGT R: CTGTTTCATGGTCACTTTGG	145–153	52	0.645	0.636	96	5	0.587

Table 2. Cont.

Locus	Genbank number (Sequence length)	Putative gene function <sup>a</sup>	Repeat motif	Primer sequence (5'-3')	Size range (bp)	T <sub>a</sub> (°C)	H <sub>O</sub>	H <sub>E</sub>	N	N <sub>a</sub>	PIC
WFT104	GT298583 (602 bp)	XP_002063687.1  GK15779 [ <i>Drosophila willistoni</i> ] 4e-17	(GT)5	F: TCACGCAAGCTAACAGCCCT R: ACAAAGTTGCCTGCCTGAAT	150–159	52	0.478	0.473	96	4	0.427
WFT108	GT300460 (686 bp)	Unknown	(AT)5	F: AGGATAGCTTGTGTTTGTGG R: CCATTTGTAAGTAGCGTAGGA	135–140	52	0.355	0.350	96	4	0.328
WFT124	GT311015 (1114 bp)	Unknown	(TG)5	F: CATTATGTGCCTCACCTCCG R: GCCTCAATTCTTCCTTGCG	278–281	52	0.521	0.664	95	4	0.626
WFT139	GT301579 (439 bp)	Unknown	(TC)5	F: CATGGGTCCTCCAGTGAG R: GCGAAACCTATCCCCTTATC	138–142	52	0.072	0.089	96	3	0.088
WFT141	GT310355 (612 bp)	Unknown	(GT)5	F: GCTTTTGCATACCTTGCTTC R: GGTAAGGGCCGGTTTTGTT	174–183	52	0.588	0.678	96	7	0.651
WFT144	GT310315 (766 bp)	Unknown	(GT)5	F: TCGCAGAAGTTTGTGGTGAG R: GAGCCGATAAAAGTAGTGGAG	94–137	52	0.598	0.844	94	15	0.875
<b>Mean</b>							0.463	0.504		5.500	0.476

T<sub>a</sub>, annealing temperature; N: number of analyzed individuals; N<sub>a</sub>: number of alleles detected; H<sub>O</sub>: observed heterozygosity; H<sub>E</sub>: expected heterozygosity; PIC: polymorphism information content; <sup>a</sup> Based on BLASTX analysis. The species source and Accession No. of the best hit(s) is indicated, together with the E-value for the match.

**Table 3.** Collection information for samples used in this study.

Location	Nb samples	Coordinates	Sampling dates	Host
Harbin	31	45°44'51.13"N, 126°38'2.54"E	23 July 2010	<i>Tagetes erecta</i> L.; <i>Hosta ventricosa</i> (Salisb.) Stearn
Dali	35	25°36'17.49"N, 100°14'49.75"E	30 July 2011	<i>Petunia hybrida</i> Vilm; <i>Nicandra</i> <i>physalodes</i> ; <i>Canna indica</i> L.
Guiyang	30	26°37'40.97"N, 106°49'23.06"E	26 July 2011	<i>Solanum melongena</i> L.; <i>Cucurbita pepo</i> L.

Nb samples, number of samples.

After sequential Bonferroni correction for multiple tests, only WFT144 in Dali and WFT50 in Guiyang significantly deviated from Hardy-Weinberg equilibrium (HWE), possibly due to the presence of null alleles, which was further confirmed by the MICRO-CHECKER [20] analysis (Table 4). In addition, no band-stuttering, large allele dropouts or significant genotypic linkage disequilibrium was detected. Genetic diversity analysis indicated that Dali displayed the highest number of alleles ( $N_a = 4.944$ ) and expected heterozygosity ( $H_E = 0.522$ ) and Harbin the lowest ( $N_a = 4.389$ ;  $H_E = 0.468$ ) (Table 4).

**Table 4.** Population genetic parameters at 18 microsatellite loci in 3 *Frankliniella occidentalis* populations.

Locus	Harbin					Dali					Guiyang				
	<i>N</i>	<i>N<sub>a</sub></i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>r</i>	<i>N</i>	<i>N<sub>a</sub></i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>r</i>	<i>N</i>	<i>N<sub>a</sub></i>	<i>H<sub>O</sub></i>	<i>H<sub>E</sub></i>	<i>r</i>
<b>WFT20</b>	31	4	0.613	0.618	−0.003	35	5	0.714	0.671	−0.039	30	5	0.633	0.684	0.035
<b>WFT24</b>	31	2	0.452	0.437	−0.017	35	2	0.086	0.133	0.110	30	2	0.233	0.299	0.095
<b>WFT25</b>	31	10	0.613	0.811	0.119	35	11	0.743	0.871	0.069	30	12	0.767	0.873	0.064
<b>WFT28</b>	31	3	0.290	0.252	−0.155	35	3	0.400	0.355	−0.062	30	3	0.033	0.065	0.149
<b>WFT37</b>	30	3	0.200	0.238	0.062	35	3	0.400	0.483	0.078	29	4	0.517	0.510	0.007
<b>WFT50</b>	31	6	0.290	0.409	0.112	34	5	0.382	0.476	0.076	30	5	0.267	0.491	<b>0.193</b>
<b>WFT51</b>	31	4	0.613	0.558	−0.069	35	5	0.657	0.616	−0.040	30	5	0.600	0.596	−0.012
<b>WFT64</b>	31	3	0.548	0.529	−0.019	34	3	0.588	0.543	−0.056	30	3	0.633	0.546	−0.084
<b>WFT66</b>	31	3	0.387	0.395	0.013	35	3	0.629	0.568	−0.049	30	3	0.500	0.466	−0.023
<b>WFT83</b>	31	3	0.548	0.505	−0.047	35	4	0.371	0.348	−0.049	30	3	0.467	0.499	0.036
<b>WFT87</b>	30	5	0.200	0.187	−0.102	34	6	0.471	0.501	0.009	30	6	0.367	0.484	0.106
<b>WFT98</b>	31	4	0.516	0.546	0.027	35	4	0.686	0.644	−0.043	30	5	0.733	0.718	−0.011
<b>WFT104</b>	31	3	0.419	0.370	−0.064	35	4	0.514	0.573	0.051	30	4	0.500	0.474	−0.027
<b>WFT108</b>	31	3	0.355	0.297	−0.193	35	4	0.343	0.394	0.093	30	3	0.367	0.359	0.026
<b>WFT124</b>	31	4	0.452	0.618	0.129	34	4	0.412	0.654	0.176	30	4	0.700	0.722	0.019
<b>WFT139</b>	31	3	0.097	0.151	0.114	35	3	0.086	0.083	−0.043	30	2	0.033	0.033	−0.017
<b>WFT141</b>	31	5	0.516	0.668	0.103	35	7	0.514	0.601	0.069	30	5	0.733	0.766	0.020
<b>WFT144</b>	31	11	0.645	0.838	0.116	34	13	0.529	0.881	<b>0.196</b>	29	11	0.621	0.812	0.124
<b>Mean</b>		4.389	0.431	0.468	0.007		4.944	0.474	0.522	0.030		4.722	0.484	0.522	0.039

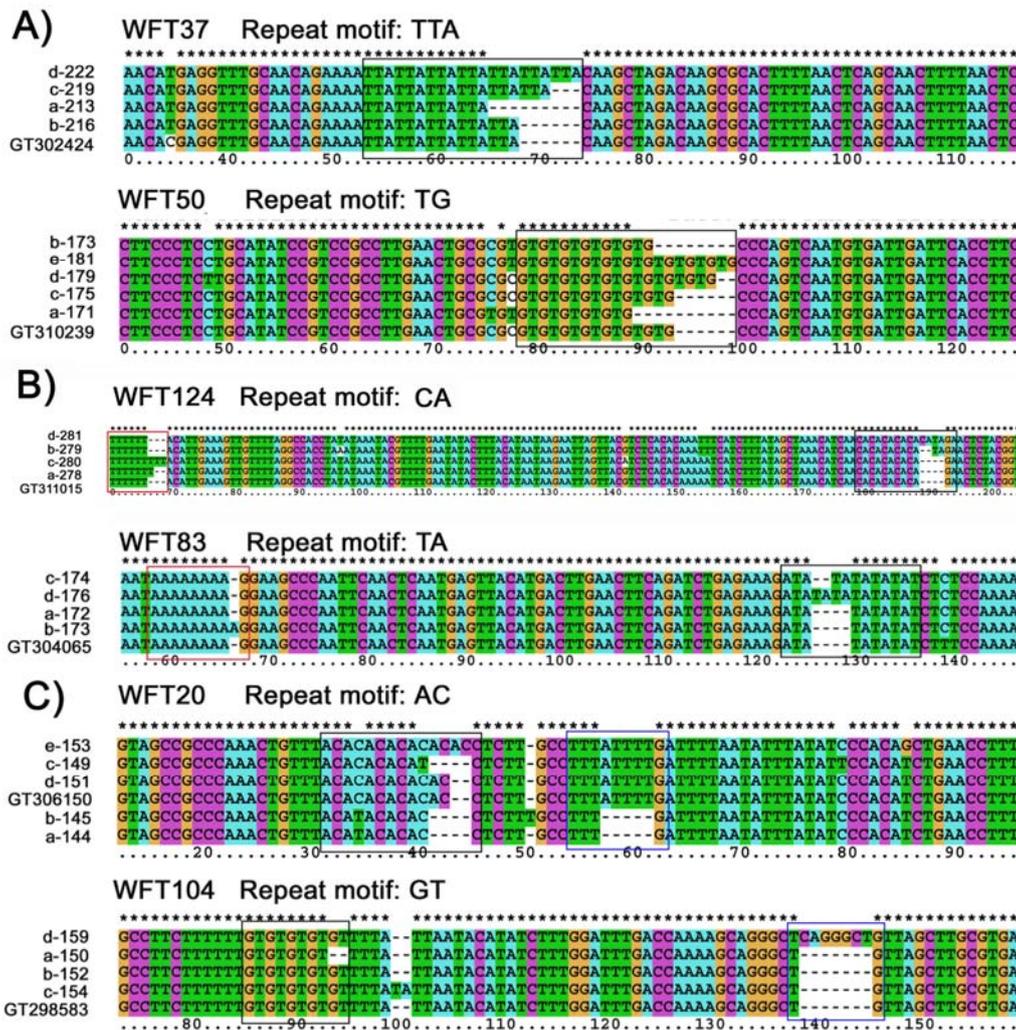
*N*: number of analyzed individuals; *N<sub>a</sub>*: number of alleles detected; *H<sub>O</sub>*: observed heterozygosity; *H<sub>E</sub>*: expected heterozygosity; *r*: null allele frequency; Bold indicates deviations from Hardy-Weinberg equilibrium after sequential Bonferroni correction for multiple tests ( $P < 0.05$ ).

## 2.2. Mutations of EST-SSRs

Ninety-five different alleles, whose allele frequency was approximately 98%, were successfully sequenced. All the sequences obtained corresponded exactly to the expected EST sequences. The other 4 rare alleles with a low frequency were not sequenced. Sequence analysis of these alleles revealed that three types of mutational events are responsible for the generation of new alleles (Six loci exhibiting all three mutation patterns are listed in Figure 1, the other 12 are shown in Figures S1 and S2). First, size variation of sequenced alleles was explained by the differences in the numbers of repeat motifs for 7 microsatellites (Figure 1A, Figure S1). Second, at the WFT51, WFT83, WFT108 and WFT124 loci, two different repeat motifs, including one in the flanking region, were found, both contributing to the allele-size variation (Figure 1B, Figure S2A). Third, indels in the flanking region were observed in 7 loci (WFT20, WFT66, WFT87, WFT104, WFT139, WFT141 and WFT144; (Figure 1C, Figure S2B)), but the frequencies of these alleles were very low in four loci (WFT20: 0.088; WFT66: 0.088; WFT87: 0.144 and WFT139: 0.021). Besides the three mutation patterns mentioned above, base substitutions in the repeat or the flanking region were also observed in 9 and 9 loci respectively. They did not contribute to the length changes of the microsatellites. In addition, several loci had multiple mutation types mentioned above, e.g., WFT37 contained both base substitutions in the flanking region and step-wise mutation in the repeat region; WFT104 had both indels in the flanking regions and step-wise mutation in the repeat motif. A minimum number of contiguous repeats might be necessary for slippage to occur. These have been suggested to be four in di-nucleotide repeats and two in tri- and tetra-nucleotide repeats [21,22]. Using these criteria, we calculated the frequency of slippage consistent and inconsistent mutations. The allele size variation mainly came from the slippage at the repeat motifs. Sixty-six alleles with a frequency of 75.5% from the 18 loci possessed this mutation mechanism. Twenty-two alleles from 7 loci showed slippage in the flanking region, with the frequency of 19.9%. In addition, mutation mechanisms other than slippage also occurred in our microsatellites, with the frequencies of 12.6% (20 alleles from 6 loci) in the repeat motifs and 8.4% (11 alleles from 7 loci) in the flanking region. Generally speaking, slippage in the repeat motif and flanking region was the main mutation mechanism for the newly developed microsatellites.

Length changes in microsatellite DNA are generally thought to arise from replication slippage [9]. However, a complex mutational pattern of *F. occidentalis* EST-SSRs was observed in this study. These mutational patterns (changes in the number of microsatellite repeat units, base substitutions and indels within flanking region) were also found in microsatellites of insects [15,23] and other organisms, including the maize [24] and birds [25]. It seems that the complex mutational pattern is common in the eukaryotic genomes. Zhu *et al.* showed that indel slippage or length independent slippage tended to duplicate short sequences [26]. The number of repeat motifs of *F. occidentalis* ESTs was low ( $n < 9$ ; Table 1) suggesting that indel slippage may be responsible for the complex mutational pattern of EST-SSRs in *F. occidentalis*.

**Figure 1.** Mutational patterns of EST-SSRs in *Frankliniella occidentalis*. (A) microsatellites mutated in the repeat motif; (B) microsatellites which have one SSR in the flanking region; (C) microsatellites which have indels in the flanking region. Each base is indicated by a different color. The repeat motifs are shown in black box. The red box indicates the repeat motifs in the flanking region and the blue one indicates the indels in the flanking region.



Global and pairwise  $F_{ST}$  and  $R_{ST}$  among three populations were then calculated.  $F_{ST}$  assumes an infinite allele model and  $R_{ST}$  assumes a stepwise mutation model [27]. Global  $F_{ST}$  and  $R_{ST}$  considering all 18 loci showed a low but significant differentiation (global  $F_{ST} = 0.029$ ,  $P < 0.001$ ; global  $R_{ST} = 0.023$ ,  $P < 0.001$ ) among all three populations (Table 5). Moreover, including the loci which have one SSR or (and) indels in the flanking region did not significantly change the global  $F_{ST}$  and  $R_{ST}$  values with overlapping 95% confidence intervals (Table 5). When considering the same loci combinations, the global and pairwise  $F_{ST}$  and  $R_{ST}$  values did not differ significantly from each other with overlapping 95% confidence intervals (Table 5). However, no clear correlation was found between pairwise estimate of  $F_{ST}$  and  $R_{ST}$  (Spearman  $r = -0.202$ ;  $P = 0.264$ ). Pairwise  $F_{ST}$  results are quite consistent in all cases, Dali/Guiyang exhibited the lowest differentiation estimates and Harbin/Guiyang exhibited the highest. However, they were not reflected in the pairwise  $R_{ST}$  results.

This might be due to the fact that the microsatellites mutated in the flanking region did not strictly conform to IAM and/or SMM model(s). Eleven out of 18 loci showed multiple sources of length variation which cannot be explained solely by gain or loss of one or two repeats as in the case of SMM based models. Thus, methods based on the IAM might be appropriate for many loci in our study, although they were not supported by our analysis. Anderson also suggested that IAM was more appropriate for one parasite's (*Plasmodium falciparum*) microsatellites which have complex mutation patterns [28]. Furthermore, the precision of global differentiation estimates improves (the confidence intervals narrows) with increasing numbers of loci analyzed (Table 5). Thus, if users of the described microsatellites want precision in their estimates, more loci should be used.

**Table 5.** Pairwise  $F_{ST}$  (below the diagonal) and  $R_{ST}$  (above the diagonal) matrix using different combinations of EST-SSRs of *Frankliniella occidentalis*.

Number of loci	Populations	Populations			Global $F_{ST}$	Global $R_{ST}$
		Harbin	Dali	Guiyang		
7 Loci <sup>a</sup>	Harbin		<b>0.025</b> (0.008–0.074)	0.006 (–0.006–0.066)	<b>0.022</b> (0.007–0.044)	<b>0.020</b> (0.011–0.067)
	Dali	<b>0.033</b> (0.007–0.074)		<b>0.021</b> (0.007–0.071)		
	Guiyang	<b>0.034</b> (0.016–0.049)	0.005 (–0.010–0.026)			
11 Loci <sup>b</sup>	Harbin		<b>0.024</b> (0.013–0.068)	0.020 (0.010–0.069)	<b>0.030</b> (0.011–0.056)	<b>0.030</b> (0.023–0.069)
	Dali	<b>0.031</b> (0.008–0.057)		<b>0.040</b> (0.026–0.087)		
	Guiyang	<b>0.032</b> (0.021–0.041)	<b>0.029</b> (0.002–0.085)			
14 Loci <sup>c</sup>	Harbin		<b>0.023</b> (0.015–0.061)	0.008 (0.002–0.051)	<b>0.025</b> (0.015–0.036)	<b>0.016</b> (0.014–0.051)
	Dali	<b>0.031</b> (0.014–0.055)		0.011 (0.007–0.052)		
	Guiyang	<b>0.033</b> (0.019–0.048)	<b>0.012</b> (–0.001–0.027)			
18 Loci	Harbin		<b>0.022</b> (0.017–0.056)	<b>0.016</b> (0.011–0.056)	<b>0.029</b> (0.016–0.046)	<b>0.023</b> (0.022–0.057)
	Dali	<b>0.030</b> (0.014–0.047)		<b>0.025</b> (0.019–0.063)		
	Guiyang	<b>0.032</b> (0.020–0.045)	<b>0.026</b> (0.004–0.062)			

<sup>a</sup> including the 7 loci mutated in the repeat motif; <sup>b</sup> including the 7 loci mutated in the repeat motif and 4 loci which have one SSR in the flanking region; <sup>c</sup> including the 7 loci mutated in the repeat motif and 7 loci which have indels in the flanking region; Bold indicates significant after Bonferroni correction ( $P = 0.05$ ); Values in parentheses indicate 95% confidence intervals.

### 3. Experimental Section

#### 3.1. EST Database Mining

13,839 *F. occidentalis* EST sequences were obtained from GenBank [29]. EST-trimmer [30] was then used to remove poly (A/T) stretches from the 5' or 3' ends until there were no (A)5 or (T)5 within the range of 50 bp. EST sequences shorter than 100 bp were excluded and those longer than 700 bp were clipped at their 5' end to preclude the inclusion of low-quality sequences [31]. Those obtained sequences were screened for microsatellites containing at least five di-, five tri-, four tetra-, four penta- and four hexa-nucleotide repeats using the software MISA [14]. PCR primers flanking the microsatellite repeats were designed using Primer Premier 5.0 [32]. The selected ESTs were compared to the NCBI nr protein database using the BLASTX program. A suggested cut-off value of  $1e-5$  was chosen to assign a potential homologue for each EST sequence [13].

#### 3.2. Sample Collection and DNA Extraction

In total, 96 *F. occidentalis* female adults were sampled representative of 3 sites in China during July 2010 to July 2011 (Table 3). Total genomic DNA was extracted by homogenizing a single female adult

in a 50  $\mu$ L mixture of STE buffer (100 mM NaCl, 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) in a 1.5 mL Eppendorf tube. The mixture was incubated with 2  $\mu$ L proteinase K (10 mg/mL) at 37 °C for 30 min, followed by 5 min at 95 °C. The samples were centrifuged briefly, and used immediately or stored at –20 °C for the PCR reactions.

### 3.3. Primer Testing

The forward primer of each set was tailed with U19 (GGTTTTCCCAGTCACGACG) to facilitate labeling. PCR amplifications were performed on an Applied Biosystems Veriti™ Thermal Cycler (Applied Biosystems). Each 10  $\mu$ L amplification mixture contained 1  $\times$  PCR buffer, 0.2 mM of each dNTP, ~50 ng of DNA, 0.25 units of Maxima Hot Start Taq DNA polymerase (Fermentas, Canada), 0.04  $\mu$ M of each forward primer, 0.2  $\mu$ M of each of the reverse primer and the dye-labeled U19 primer (FAM, VIC, NED or PET). These cycling conditions were an initial denaturing for 4 min at 95 °C; 10 cycles of 95 °C for 30 s, 51 °C for 30 s, 72 °C for 30 s; 25 cycles of 95 °C for 30 s, 54 °C for 30 s, 72 °C for 30 s, and a final extension at 72 °C for 10 min. PCR products were run on the ABI 3130 capillary sequencer along with the GeneScan-500 LIZ size standard and allele sizes were determined using GENEMAPPER version 4.0 (Applied Biosystems).

### 3.4. Allele Sequencing

Different alleles per locus detected by the capillary sequencer were amplified using a 50  $\mu$ L PCR reaction with non-fluorescent labeling primers (conditions as above with a specific anneal temperature at 52 °C). The purified PCR products (purified using Axygen cleanup kit) were subsequently ligated into the pGEM-T vector (Promega) and introduced into *Escherichia coli* DH5 $\alpha$  cells. Six positive clones for each allele were sequenced to exclude PCR artefacts. Alignments of the sequenced alleles were generated using the Clustal X 2.0.11 program [33]. Several loci were then manually aligned using BioEdit 7.0.4 [34].

### 3.5. Data Analysis

All genetic statistics were carried out based on the genotyping data from three populations. MICRO-CHECKER 2.2.3 was used to detect genotyping errors due to null alleles, stuttering, or allele dropout using 1000 randomizations [20]. The program Genepop 4.0.10 [35] was used to test for linkage disequilibrium between pairs of loci in each population (100 batches, 1000 iterations per batch) and for deviations from Hardy–Weinberg equilibrium (HWE) at each locus/population combination using Fisher's exact tests. The population genetic diversity indices such as total alleles per locus ( $N_A$ ), observed heterozygosity ( $H_O$ ), expected heterozygosity ( $H_E$ ) and mean number of alleles ( $N_a$ ) was assessed using GenAIEx 6.41 [36]. We also calculated the polymorphism information content (PIC) using CERVUS version 3.0 [37]. Pairwise  $F_{ST}$  and  $R_{ST}$  value and their significance for each population comparison were calculated with 10,000 permutations in Arlequin 3.0 [38] and RST CALC 2.2 [27], respectively.

#### 4. Conclusions

In summary, 18 highly polymorphic EST-SSRs have been specifically developed for *F. occidentalis* in this study. Sequence analysis of alleles per locus revealed a complex mutational pattern of these EST-SSRs. Thus, these EST-SSRs are useful markers for the invasive species *F. occidentalis* but greater attention should be paid to the mutational characteristics of these markers when they are used in population genetic studies.

#### Acknowledgments

We thank Da-Song Chen and Yan-Kai Zhang of the Department of Entomology, Nanjing Agricultural University (NJAU) for help with the collection of western flower thrips. We are also grateful to Kai-Jun Zhang, Ming-Zhi Yu and Jin-Bo Li of the Department of Entomology, NJAU for their kind help with experiments.

This work was supported by grants from the National Key Basic Research Program (973 Program, No. 2009CB119200) from the Ministry of Science and Technology of China and the Science and Technology Research Program of the National Agricultural Public Welfare Fund (No. 200803025) from the Ministry of Agriculture of China.

#### References

1. Mound, L.A.; Morris, D.C. The insect order Thysanoptera: classification *versus* systematics. *Zootaxa* **2007**, *1668*, 395–411.
2. Reitz, S.R. Biology and ecology of the western flower thrips (Thysanoptera: Thripidae): The making of a pest. *Fla. Entomol.* **2009**, *92*, 7–13.
3. Bryan, D.E.; Smith, R.F. The *Frankliniella occidentalis* (Pergande) complex in California. *Univ. Calif. Publ. Entomol.* **1956**, *10*, 359–410.
4. Kirk, W.D.J.; Terry, L.I. The spread of the western flower thrips *Frankliniella occidentalis* (Pergande). *Agric. For. Entomol.* **2003**, *5*, 301–310.
5. Behura, S.K. Molecular marker systems in insects: current trends and future avenues. *Mol. Ecol.* **2006**, *15*, 3087–3113.
6. Ascunce, M.S.; Yang, C.C.; Oakey, J.; Calcaterra, L.; Wu, W.J.; Shih, C.J.; Goudet, J.; Ross, K.G.; Shoemaker, D. Global invasion history of the fire ant *Solenopsis invicta*. *Science* **2011**, *331*, 1066–1068.
7. Toth, G.; Gaspari, Z.; Jurka, J. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Res.* **2000**, *10*, 967–981.
8. Bhargava, A.; Fuentes, F.F. Mutational dynamics of microsatellites. *Mol. Biotechnol.* **2010**, *44*, 250–266.
9. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445.
10. Matsuoka, Y.; Mitchell, S.E.; Kresovich, S.; Goodman, M.; Doebley, J. Microsatellites in *Zea*-variability, patterns of mutations, and use for evolutionary studies. *Theor. Appl. Genet.* **2002**, *104*, 436–450.

11. Bruford, M.W.; Wayne, R.K. Microsatellites and their application to population genetic studies. *Curr. Opin. Genet. Dev.* **1993**, *3*, 939–943.
12. Brunner, P.C.; Frey, J.E. Isolation and characterization of six polymorphic microsatellite loci in the western flower thrips *Frankliniella occidentalis* (Insecta, Thysanoptera). *Mol. Ecol. Notes* **2004**, *4*, 599–601.
13. Rotenberg, D.; Whitfield, A.E. Analysis of expressed sequence tags for *Frankliniella occidentalis*, the western flower thrips. *Insect Mol. Biol.* **2010**, *19*, 537–551.
14. Thiel, T. MISA—Microsatellite identification tool. Available online: <http://pgrc.ipk-gatersleben.de/misa/misa.html> (accessed on 15 August 2011).
15. Sun, J.T.; Zhang, Y.K.; Ge, C.; Hong, X.Y. Mining and characterization of sequence tagged microsatellites from the brown planthopper *Nilaparvata lugens*. *J. Insect Sci.* **2011**, *11*, 134:1–134:11.
16. Weng, Y.; Azhaguvel, P.; Michels, G.J.; Rudd, J.C. Cross-species transferability of microsatellite markers from six aphid (Hemiptera: Aphididae) species and their use for evaluating biotypic diversity in two cereal aphids. *Insect Mol. Biol.* **2007**, *16*, 613–622.
17. Li, B.; Xia, Q.; Lu, C.; Zhou, Z.; Xiang, Z. Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes. *Genomics Proteomics Bioinforma.* **2004**, *2*, 24–31.
18. Gao, L.; Tang, J.; Li, H.; Jia, J. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.* **2003**, *12*, 245–261.
19. Jurka, J.; Pethiyagoda, C. Simple repetitive DNA sequences from primates: Compilation and analysis. *J. Mol. Evol.* **1995**, *40*, 120–126.
20. van Oosterhout, C.; Hutchinson, W.F.; Wills, D.P.M.; Shipley, P. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **2004**, *4*, 535–538.
21. Messier, W.; Li, S.H.; Stewart, C.B. The birth of microsatellites. *Nature* **1996**, *381*, 483.
22. Zhu, Y.; Queller, D.C.; Strassmann, J.E. A phylogenetic perspective on sequence evolution in microsatellite loci. *J. Mol. Evol.* **2000**, *50*, 324–338.
23. Colson, I.; Goldstein, D.B. Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* **1999**, *152*, 617–627.
24. Lia, V.V.; Bracco, M.; Gottlieb, A.M.; Poggio, L.; Confalonieri, V.A. Complex mutational patterns and size homoplasy at maize microsatellite loci. *Theor. Appl. Genet.* **2007**, *115*, 981–991.
25. Primmer, C.R.; Ellegren, H. Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol.* **1998**, *15*, 997–1008.
26. Zhu, Y.; Strassmann, J.E.; Queller, D.C. Insertions, substitutions, and the origin of microsatellites. *Genet. Res.* **2000**, *76*, 227–236.
27. Goodman, S.J. RST CALC: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Mol. Ecol.* **1997**, *6*, 881–885.

28. Anderson, T.J.C.; Su, X.Z.; Roddam, A.W.; Day, K.P. Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. *Mol. Ecol.* **2000**, *9*, 1599–1608.
29. National Center for Biotechnology Information Expressed Sequence Tags database. Available online: <http://www.ncbi.nlm.nih.gov/dbEST/> (accessed on 15 August 2011).
30. EST trimmer. Available online: [http://pgrc.ipk-gatersleben.de/misa/download/est\\_trimmer.pl](http://pgrc.ipk-gatersleben.de/misa/download/est_trimmer.pl) (accessed on 15 August 2011).
31. Thiel, T.; Michalek, W.; Varshney, R.K.; Graner, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* **2003**, *106*, 411–422.
32. Lalitha, S. Primer Premier 5.0. *Biotech Software Internet Report* **2000**, *1*, 270–272.
33. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
34. Hall, T.A. BioEdit: A user friendly biological sequence alignment editor and analyses program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
35. Rousset, F. GenePop'007: A complete re-implementation of the GenePop software for Windows and Linux. *Mol. Ecol. Resour.* **2008**, *8*, 103–106.
36. Peakall, R.; Smouse, P.E. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **2006**, *6*, 288–295.
37. Marshall, T.C.; Slate, J.; Kruuk, L.; Pemberton, J.M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **1998**, *7*, 639–655.
38. Excoffier, L.; Laval, G.; Schneider, S. Arlequin, ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **2005**, *1*, 47–50.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).