Theor Appl Genet (2007) 115:981–991
DOI 10.1007/s00122-007-0625-y

ORIGINAL PAPER

Complex mutational patterns and size homoplasy at maize microsatellite loci

V. V. Lia · M. Bracco · A. M. Gottlieb · L. Poggio · V. A. Confalonieri

Received: 25 April 2006 / Accepted: 31 July 2007 / Published online: 22 August 2007

Abstract Microsatellite markers have become one of the most popular tools for germplasm characterization, population genetics and evolutionary studies. To investigate the mutational mechanisms of maize microsatellites, nucleotide sequence information was obtained for ten loci. In addition, Single-Strand Conformation Polymorphism (SSCP) analysis was conducted to assess the occurrence of size homoplasy. Sequence analysis of 54 alleles revealed a complex pattern of mutation at 8/10 loci, with only 2 loci showing allele variation strictly consistent with stepwise mutations. The overall allelic diversity resulted from changes in the number of repeat units, base substitutions, and indels within repetitive and non-repetitive segments. Thirty-one electromorphs sampled from six maize landraces were considered for SSCP analysis. The number of conformers *per* electromorph ranged from 1 to 7, with 74.2% of the electromorphs showing more than one conformer. Size homoplasy was apparent within landraces and populations. Variation in the amount of size homoplasy was observed within and between loci, although no differences were detected among populations. The results of the present study provide useful information on the interpretation of genetic data derived from microsatellite markers. Further efforts are still needed to determine the impact of these findings on the estimation of population parameters and on the inference of phylogenetic relationships in maize investigations.

Communicated by M. Bohn.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-007-0625-y) contains supplementary material, which is available to authorized users.

V. V. Lia (✉) · M. Bracco · A. M. Gottlieb · L. Poggio · V. A. Confalonieri
Laboratorio de Genética, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Intendente Güiraldes y Costanera Norte s/n, 4to. Piso – Pabellón II – Ciudad Universitaria, C1428EHA Ciudad Autónoma de Buenos Aires, Argentina
e-mail: verolia@ege.fcen.uba.ar; vlia@cnia.inta.gov.ar

V. V. Lia · M. Bracco · A. M. Gottlieb · L. Poggio · V. A. Confalonieri
Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

Introduction

Microsatellite markers have become the molecular tool of choice in a wide range of maize investigations, including germplasm characterization, population genetics and evolutionary studies (e.g., Laborda et al. 2005; Le Clerc et al. 2005; Matsuoka et al. 2002b; Pressoir and Berthaud 2004; Reif et al. 2004, 2005; Santa Cruz Varela et al. 2004).

Allelic variation at microsatellite loci is frequently assumed to conform to the Stepwise Mutation Model (SMM) (Kimura and Ohta 1978) or generalizations thereof (Two Phase Model, TPM; Generalized Stepwise Mutation Model, GSM) (Di Rienzo et al. 1994). According to the SMM, a mutation alters the length of a repetitive array through the addition or removal of one repeat unit at a fixed rate, a symmetric forward–backward random process that is independent of repeat length (Ellegren 2004). However, deviations from these assumptions have been reported in plants for both chloroplast (Doyle et al. 1998; Hale et al. 2004) and nuclear loci (Adams et al. 2004; England et al. 2002; Saltonstall 2003).

The mutational behaviour of maize microsatellites was first addressed by Matsuoka et al. (2002a) who examined the distribution of allele sizes for 46 loci, detecting deviations from the SMM at 44. Moreover, sequence analysis of

31 alleles from 6 loci revealed a high incidence of insertion/deletion events in the regions flanking the repeat motifs. A contrasting pattern was found in a series of mutation accumulation experiments conducted by Vigouroux et al. (2002), in which a total of 85 dinucleotide loci were shown to conform to the expectations of the SMM.

Size homoplasy (i.e. the occurrence of alleles that are identical in size but not identical by descent) is a phenomenon implicit to the SMM, TPM or GSM. For all these models alleles can mutate to allelic states already present in the populations, causing alleles from different lineages to converge in size and therefore affecting our ability to estimate population genetics parameters. Homoplasy also depends on evolutionary factors independent of the mutation model such as the mutation rate, the effective population size and the time of divergence between populations (Estoup and Cornuet 1999). Evidence on the occurrence and effects of size homoplasy in evolutionary studies has been obtained for several animal and plant species (Adams et al. 2004; Angers et al. 2000; Blankenship et al. 2002; Culver et al. 2001; Taylor et al. 1999; van Oppen et al. 2000; Viard et al. 1998; Xie et al. 2006; Yokoyama et al. 2004), but no references exist regarding its frequency in maize, or any other crop species populations.

The Single-Strand Conformation Polymorphism (SSCP) technique (Orita et al. 1989) has repeatedly been proposed as a cost-effective tool to empirically evaluate Molecularly Accessible Size Homoplasy (MASH, *sensu* Estoup et al. 2002) in natural populations, as single-nucleotide variation can be detected without extensive sequencing (Angers et al. 2000; Estoup and Cornuet 1999; Sunnucks et al. 2000). Its underlying principle is that the electrophoretic mobility of a single-stranded DNA molecule in a non-denaturing gel is determined by its size and structure, the latter being highly dependent on primary sequence.

To further explore the mutational dynamics of maize microsatellite loci we implemented the SSCP technique to assess the occurrence of size homoplasy at ten loci in a sample of individuals from eight populations representing six maize landraces. Allele sequence information was also obtained as a preliminary step to characterize the loci included in this study.

Materials and methods

A set of 146 individuals representing six maize landraces (eight populations) from northwestern Argentina were included in the present study (Table 1). Samples were collected from isolated open-pollinated populations maintained by native farmers.

DNA extracts were obtained from 2- to 3-day-old seedlings germinated from individual grains, using the methods described in Dellaporta et al. (1983).

Microsatellite typing

Individual genotypes were assessed for ten randomly selected loci (one from each maize chromosome). The list of loci evaluated, including chromosomal location (bin number), previously reported repeat motifs and associated primers, is shown in Table 2. Additional information on these loci can be found at the MAIZE DATA BASE (http://www.maizegdb.org/ssr.php).

Amplifications were performed in 25 µl, using 125 µM of each dNTP, 30 ng of each primer, 10% of 10× PCR buffer (Promega), 1.5 mM $MgCl_2$, 0.5 units of *Taq* DNA polymerase (Promega), 1 µl of genomic DNA template (*ca.* 100 ng), and sterile double-distilled water under a touchdown cycling profile (annealing temperature 65–55°C).

Table 1 Maize landraces included in the present study

Population	Voucher[a]	Collection site	Landrace[b]	Sample size
6473	VAV 6473	Susques, Dpto. Susques, Jujuy, Argentina. 3,620 m.a.s.l.	Altiplano	20
6167	VAV 6167	El Puesto, Dpto. Santa Victoria, Salta, Argentina. 3,000 m.a.s.l.	Altiplano	14
6484	VAV 6484	Tumbaya, Dpto. Tumbaya, Jujuy, Argentina. 2,010 m.a.s.l.	Amarillo Chico	20
6476	VAV 6476	Termas de Reyes, Dpto. Capital, Jujuy, Argentina. 1,690 m.a.s.l.	Amarillo Chico	14
6480	VAV 6480	La Ciénaga de Pumamarca, Dpto. Tumbaya, Jujuy, Argentina	Amarillo Grande	25
6485	VAV 6485	Colonia San José, Dpto. Tilcara, Jujuy, Argentina. 2,670 m.a.s.l.	Blanco	13
6313	VAV 6313	Los Toldillos, Piedras Blancas, Dpto. Ambato, Catamarca, Argentina. 1,600 m.a.s.l.	Pisingallo	16
6482	VAV 6482	La Candelaria, Dpto. Candelaria, Salta, Argentina. 910 m.a.s.l.	Orgullo Cuarentón	24

M.a.s.l meters above sea level

[a] Voucher specimens are deposited at the "Laboratorio de Recursos Genéticos Vegetales N.I. Vavilov" (VAV), Facultad de Agronomía, Universidad de Buenos Aires

[b] Taxonomic identification based on morphological criteria

Table 2 Microsatellite loci used for this study

Locus	Chromosome location (bin)	Repeat motif[a]	Forward primer (5′–3′)	Reverse primer (5′–3′)
phi037	1.08	CT	CCCAGCTCCTGTTGTCGGCTCAGAC	TCCAGATCCGCCGCACCTCACGTCA
phi127	2.08	GTCT	ATATGCATTGCCTGGAACTGGAAGGA	AATTCAAACACGCCTCCCGAGTGT
phi029	3.04	CCCT-CT	TTGTCTTTCTTCCTCCACAAGCAGCGAA	ATTTCCAGTTGCCACCGACGAAGAACTT
nc135	4.01	AG	CACAAAGAGCAGCCCACTTT	AAGTTGCTGACATCGATCCA
bnlg1700	5.03	AG	GTCACATCCATGTAGTGCACG	GGCACCCTTTTGAAACCTTT
bnlg1165	6.01	AG	CGCTTGCATCATCTCAAGAA	TTCAAGTTTAGCCACCCACC
phi069	7.05	GAC	AGACACCGCCGTGGTCGTC	AGTCCGGCTCCACCTCCTTC
Phi121	8.03	CCG	AGGAAAATGGAGCCGGTGAACCA	TTGGTCTGGACCAAGCACATACAC
bnlg1209	9.04	AG	GTCCCGGGCAGAATAATACC	TTCCTCCTTGAAGTGCTCGT
Phi059	10.02	ACC	AAGCTAATTAAGGCCGGTCATCCC	TCCGTGTACTCGGCGGACTC

[a] As reported in maizeDB

PCR products were separated on 6% denaturing polyacrylamide gel (8 M urea) following standard procedures. Gels were silver-stained with Silver Sequence DNA Staining Reagents (Promega). Alleles were identified by comparison with products of known size using GelPro Analyzer 4.0 (Media Cybernetics).

Sequencing of microsatellite alleles

Microsatellite alleles found to have overall frequencies higher than 1% were used for sequence analysis.

Nucleotide sequences of those alleles found in homozygosis in denaturing PAGE were obtained via amplification and direct sequencing. Alleles only found in heterozygosis were isolated from 6% (w/v) denaturing polyacrilamide gels following the procedure of Stumm et al. (1997) and subsequently amplified for automated sequencing. Each allele class was represented by a single individual selected at random from the populations listed in Table 1.

Amplifications were performed as described above. For the amplicon purification, PCR products were separated by electrophoresis on 2% agarose gels with 1× TAE buffer containing 0.5 mg/ml of ethidium bromide. Bands were excised from the gel under UV light, and the DNA was purified with QIAquik Gel Extraction Kit (QIAGEN Inc.). When needed, samples were cloned using pGEM-T Easy Vector System (Promega); at least three clones were sequenced *per* amplicon. Nucleotide sequences of both strands were obtained *via* direct sequencing of PCR products using the corresponding microsatellite primers, or with primers T7 and SP6 for cloned DNA fragments. Sequences were produced on an Applied Biosystems automated 377 DNA sequencer (Perkin Elmer).

Microsatellite loci showing allele size variation due exclusively to the differences in the number of repeat units were considered to be consistent with the SMM. Allele size variation originating from indels either within or outside the repetitive tract was considered to be inconsistent with the SMM.

All nucleotide sequences have been deposited in GenBank (*phi059*: AY965934, 38–40, 45; *phi127*: AY965947-51; *phi029*: AY965917-23; *phi121*: DQ350843; *nc135*: DQ350844-48; *phi037*: DQ350849-54; *phi069*: DQ350855-58; *bnlg1700*: DQ350859-64; *bnlg1165*: DQ350865-74; *bnlg1209*: DQ350875-79).

SSCP analysis

The SSCP analysis was conducted using only homozygous individuals from each population studied. Heterozygous individuals were not considered for the analysis due to the complexity of their SSCP profiles.

The PCR amplification products were mixed with an equal volume of 98% (v/v) formamide loading buffer. Five microlitres aliquots of heat-denatured samples were separated by electrophoresis in a Model S2 apparatus (Life Technologies-Gibco BRL Sequencing System) through 6% (w/v) native polyacrylamide gels. Electrophoresis was carried out in 1× TBE buffer at constant power (25W) and room temperature for 3–4 h. Possible differences on native PAGE performance due to temperature variation were accounted for by using internal standards in every run. Gels were stained with silver nitrate as described above. Air-dried gels were scanned and banding patterns were registered.

Hereafter, SSCP variants will be referred to as conformers (C), whereas microsatellite alleles detected by denaturing PAGE will be referred to as electromorphs (E). The reliability of the SSCP experiments was checked by sequence analysis of nine conformers from locus *phi059*. The complete conformer data matrix is available upon request.

Data analysis

Allelic frequencies of the electromorphs detected by denaturing PAGE were calculated by the direct count method. Sequence alignments were performed using ClustalW (Higgins et al. 1994) with minor manual modifications.

Following a similar approach to that implemented in Viard et al. (1998), we explored whether the number of conformers revealed by SSCP analysis was dependent on the sample sizes examined. For this purpose, Spearman's coefficient of rank correlation (R_s) was used to test the relationship between the number of conformers detected for a given electromorph and: (1) the number of individuals examined; and (2) the number of populations from which each electromorph was sampled.

The length of the repetitive tract has been shown to substantially influence microsatellite mutation rates, microsatellite loci with higher repeat counts being associated with higher mutation rates (Estoup and Cornuet 1999). To examine whether this parameter also has an impact on the number of conformers detected at each locus the length of the repetitive tract was tested against the number of conformers. Additionally, to evaluate if there is an influence of the non-repetitive regions in the number of conformers, the relationship between the latter and the length of the amplified region was also examined.

Two groups of loci were defined according to the mutational patterns inferred here from sequence analysis. The first group is composed of those loci consistent with the SMM, TPM or GSM mutation models (i.e.: *bnlg1700, bnlg1165, bnlg1209*), and the second group is composed of those loci not consistent with these models (i.e.: *phi121*, *phi069*, *nc135*, *phi037*, *phi127*, *phi029*, *phi059*) (see Results). Comparison of the number of conformers between these two groups was accomplished using Mann–Whitney *U* test, as implemented in STATISTICA software package (StatSoft 1999). Comparison of the number of conformers *per* electromorph (C/E) among populations was conducted using the Kruskal–Wallis ANOVA test provided by the same software package.

Estimation of the rate of indels and point mutations relative to stepwise-like mutation events

At mutation-drift equilibrium, the heterozygosity (H) is a known function of the parameter $M = 4N_e\mu$ under both the Infinite Allele Model (IAM: $H = M/(1 + M)$) and the Stepwise Mutation Model (SMM: $H = 1-(1 + 2M)^{-0.5}$) (Estoup and Cornuet 1999). To estimate the rate of indels and point mutations relative to stepwise-like mutation events we computed heterozygosity as follows: (1) considering the set of alleles generated by indels and point mutations (IP) as detected by sequence and SSCP analysis, and (2) considering the set of alleles generated by differences in the number of repeat units (NR), as revealed by sequence analysis (see Results for further details). For the first group, M was deduced using the equation given by the IAM (M_{IP}), and for the second group M was deduced following the assumptions of the SMM (M_{NR}). The ratio between these M values (M_{IP}/M_{NR}) constitutes an estimate of the rate of indels and point mutations relative to stepwise-like mutation events, provided that effective population sizes are the same. To that end, only those individuals having data on both electromorphs and conformers were considered for this analysis. Due to the small sample sizes of the SSCP dataset on a *per* population basis (Electronic Supplementary Material, Table E1) data from several populations had to be pooled for the calculation of heterozygosities. Previous studies have shown that populations 6473, 6167, 6485, 6484 and 6480 can be regarded as a single gene pool (Lia et al. 2007). Following from these results, populations 6476, 6313 and 6482 were excluded from the analysis of mutation rates, and the data from populations 6473, 6167, 6485, 6484 and 6480 were pooled into a single dataset for subsequent analysis.

Statistical significance of the difference between M_{IP} and M_{NR} was assessed indirectly by comparing the average heterozygosities from which these values were obtained using Mann–Whitney *U* test.

Results

A total of 68 electromorphs were detected by denaturing PAGE. Allele frequency distributions *per* locus and population are given as Electronic Supplementary Material (ESM, Table E1).

Sequence analysis

Microsatellite alleles found to have overall frequencies higher than 1% were sequenced (54 out of a total of 68). In spite of meeting this criterion, alleles 208, 210 and 218 from locus *bnlg1700* and allele 166 from locus *bnlg1209* could not be sequenced due to the lack of suitable plant material.

Sequence analysis of the 54 electromorphs revealed a complex pattern of mutation at 8 out of 10 loci examined, with only two loci showing size variation strictly consistent with stepwise mutations (i.e. *bnlg1700* and *bnlg1165*) (Fig. 1). The overall allelic diversity resulted from a combination of various types of mutational events: changes in the number of microsatellite repeat units, base substitutions and indels within repetitive and flanking segments (Fig. 1). The most extreme deviations from the expectations of the SMM are exemplified by loci *phi069* and *phi059*. These loci possess very short and degraded repeat arrays, with most of the

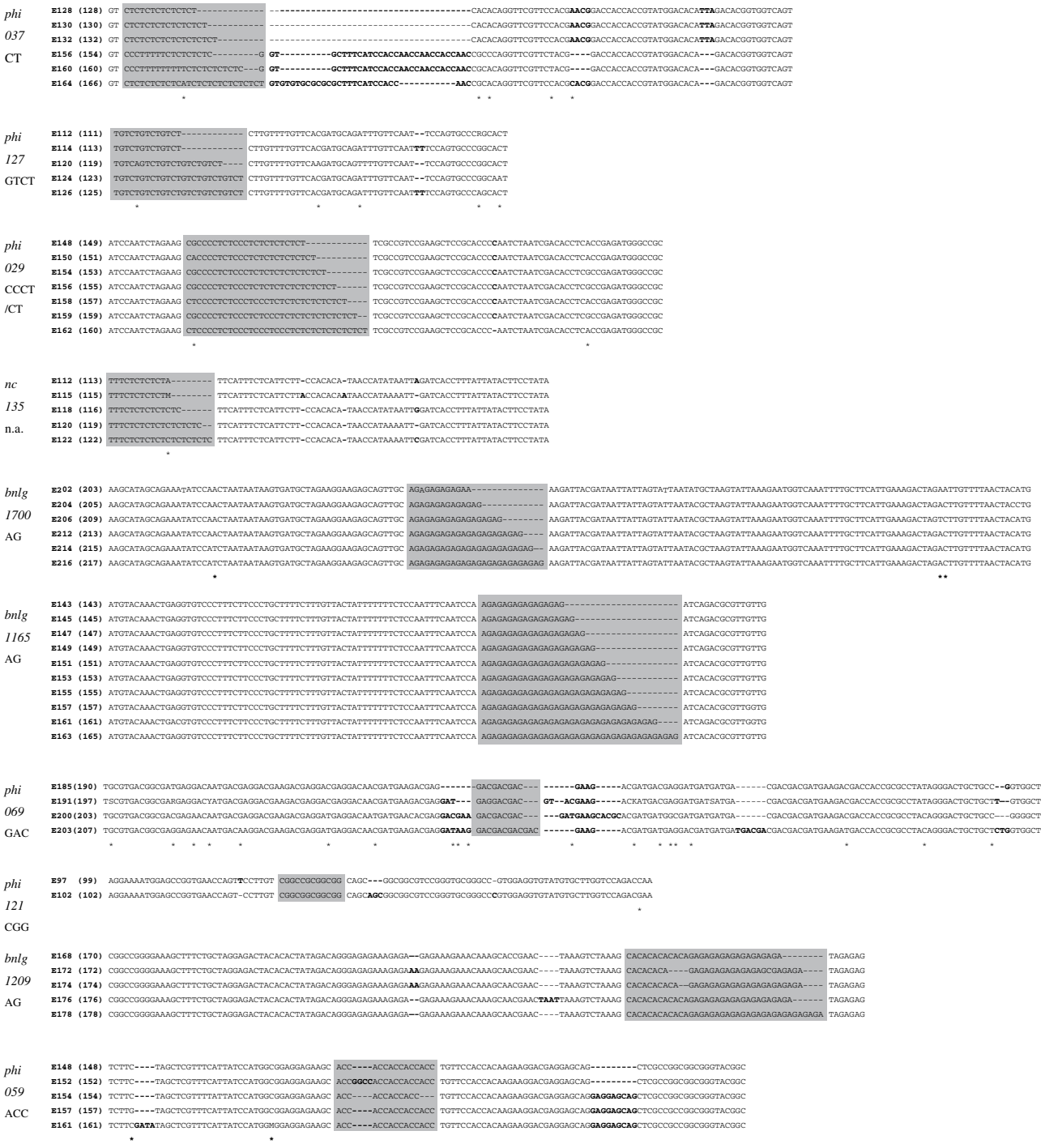

Fig. 1 Nucleotide sequences of electromorphs (E) corresponding to ten maize microsatellite loci. Loci are arranged according to chromosome location. Expected microsatellite motifs previously reported for each locus in MaizeGDB are given below locus designation; *n.a.* not available. Electromorphs are named according to their deduced size in PAGE. *Numbers* in parenthesis correspond to electromorph sizes as determined by sequence analysis. Repetitive arrays are indicated by *shaded boxes*. Indels are highlighted in *bold*. Nucleotide substitutions are indicated by *asterisks*. Amplification primers are not included. The sequence corresponding to electromorph 102 from locus *phi121* was taken from Matsuoka et al. (2002a, GenBank Accession No AY033463)

variation being confined to those regions flanking the putative repeat motif.

Loci *nc135*, *phi127* and *phi037* can be considered an intermediate class of markers because they exhibit a mixture of different mutation events, inside and outside the repeat motif. For instance, length variation at locus *phi037* is characterized by a series of indels (spanning 3 to 27 bp) occurring outside the microsatellite motif (Fig. 1). Even in

the absence of large differences in the number of repeat units, two distinct allele size categories can be distinguished at locus *phi037*: a group of alleles ranging in size from 128 to 132 bp, and another one composed by 156–164 bp alleles. In addition, conversion of the expected CT array to a mononucleotide T-tract appears to have taken place at alleles 156 and 160, further contributing to the diversity of factors involved in the observed mutation pattern.

According to MaizeGDB, locus *phi029* possesses a compound motif CCCT-CT; however, only CT dinucleotide repeats seem to differ between electromorphs 148, 150, 154, and 156, whereas differences in the number of CCCT arrays are exclusively apparent for larger electromorphs (158, 159, 162). A similar situation was observed at locus *bnlg1209*. Originally described in MaizeGDB as a pure AG microsatellite, it was found to contain a compound repeat motif mainly comprising two contiguous stretches of CA and AG dinucleotides. Both regions exhibited comparable levels of variation among alleles, with two additional insertion/deletion events, occurring outside the repeat motif, also contributing to the differences in allele sizes.

In summary, the patterns of mutation found at loci *phi121*, *phi069*, *nc135*, *phi037*, *phi127*, *phi029*, *phi059* and *bnlg1209* do not conform to the assumptions of the SMM or any of its generalizations. Notwithstanding, as size homoplasy can arise either by convergence on repeat numbers or by convergence on fragment size due to non-repeat indels at flanking regions, we decided to further explore this issue by an SSCP analysis of electromorphs.

SSCP analysis

The SSCP analysis was restricted to those electromorphs found in homozygotes in denaturing PAGE. Following this criterion, a total of 31 electromorphs (E) were considered for analysis. The number of individuals analyzed *per* population is provided in the ESM (Table E1). Locus descriptions, sample sizes and distribution of conformers (C) *per* locus are given in Table 3. A total of 85 conformers were detected, with 23/31 (74.2%) electromorphs showing more than one SSCP variant. The number of conformers ranged from 1 to 7 depending on the electromorph considered (mean = 2.74 ± 1.51), with only 8 electromorphs exhibiting a single conformer. Variation in the number of conformers was found within and among loci. Loci *nc135* and *phi059* displayed the highest number of SSCP variants (C/E = 3.67 and 5, respectively), whereas for locus *phi037* only one electrophoretic allele showed multiple conformers.

Multiple conformers were observed for the same electromorph not only when electromorphs were sampled from different populations, but also when they were sampled from the same population. Moreover, all populations exhibited similar C/E ratios in spite of belonging to different landraces (Kruskal–Wallis $H = 4.9458$, $P = 0.6666$) (Table 4).

Considering the whole dataset, no significant association was found between the number of conformers detected for a given electromorph and the number of individuals examined ($R_s = 0.3539$, $P = 0.0551$) or the number of populations from which electromorphs were sampled ($R_s = 0.2613$, $P = 0.1630$), indicating that sample sizes were large

Table 3 SSCP analysis of 10 maize microsatellite loci. Locus characteristics, sample sizes and distribution of conformers *per* locus

Locus	Total no. of electromorphs[a]	Range of electromorph sizes[b]	Length of the repetitive tract (bp)[c]	*N*	No. of electromorphs evaluated (E)	No. of conformers detected (C)	Average C/E *per* locus	Percentage of E with a single conformer
phi037	7	128–164	14–27	64 (87)	4	5	1.25	75
phi127	5	112–126	13–25	60 (93)	2	4	2	0
phi029	8	148–162	23–35	50 (67)	3	9	3	33.3
nc135	6	112–130	12–20	50 (79)	3	11	3.67	0
bnlg1700	10	202–220	12–28	59 (64)	6	17	2.83	16.6
bnlg1165	13	141–165	14–38	19 (23)	3	5	1.67	66.6
phi069	5	178–203	6–12	35 (61)	2	5	2.5	0
phi121	1	99	6	81 (147)	1	2	2	0
bnlg1209	8	166–184	30–38	52 (71)	5	17	3.4	20
phi059	5	148–161	12–19	60 (74)	2	10	5	0

N Number of homozygous individuals analysed by SSCP; total number of homozygous individuals observed are given in parenthesis

[a] Total number of electromorphs detected by denaturing PAGE

[b] Electromorphs sized in denaturing PAGE

[c] Calculated including point insertions and compound repetitive sequences

Table 4 Distribution of conformers *per* population

Population	Number of electromorphs (E) evaluated[a]	Number of conformers (C) found	Average C/E
6473	14	25	1.79
6167	17	21	1.24
6484	18	28	1.56
6476	18	27	1.5
6480	22	32	1.45
6485	15	21	1.40
6313	13	18	1.38
6482	19	32	1.68

[a] A detailed description of the loci considered for each population is given in ESM

enough to produce an accurate depiction of conformer occurrence.

The electromorphs studied here, ranged in size from 99 bp (locus *phi121*) to 220 bp (locus *bnlg1700*) (Table 3). The length of the amplified region (electromorph size) showed no correlation with the number of conformers detected ($R_s = 0.0987$, $P = 0.6037$), nor did the length of the repetitive tract ($R_s = 0.0405$, $P = 0.8317$). The latter being concordant with the non-stepwise nature detected here for the majority of loci included in the study.

As previously mentioned, size homoplasy is expected under the SMM, TPM or GSM; however, its frequency for loci showing allelic variation not compatible with these models (hereafter referred to as non-SMM loci) may be difficult to ascertain. In order to compare the relative amounts of size homoplasy among loci, the microsatellite markers included in this study were assigned to either of the two groups according to the mutational mechanisms inferred here by sequence analysis. The first group includes those loci exhibiting mutational patterns consistent with the above-mentioned mutation models (i.e.: *bnlg1700, bnlg1165, bnlg1209*), and the second group is composed of non-SMM loci (i.e.: *phi121*, *phi069*, *nc135*, *phi037*, *phi127*, *phi029*, *phi059*). Locus *bnlg1209* presented the only case of dubious placement, because the compound structure of the repetitive region identified by sequence analysis does not entirely agree with the strict SMM pattern previously reported for this locus by Vigouroux et al. (2002). However, given the mutational behaviour described by these authors on the basis of pedigree analysis, and considering that most of the variation was restricted to the gain or loss of repeats, the electromorphs from this locus were included within the first group. No significant differences were found between the number of conformers observed at each group (Mann–Whitney *U* test, $U = 105.5$, $P = 0.78$).

Estimation of the rate of indels and point mutations relative to stepwise-like mutation events

Average and single-locus heterozygosities calculated having into account either the alleles derived from indels and point substitutions (H_{IP}) or the alleles generated by differences in the number of repeats (H_{NR}) are given in Table 5. Single-locus calculations were conducted considering only the corresponding mutation events. For instance, in the reduced dataset used for this analysis, locus *bnlg1965* exhibits two alleles at frequencies 0.167 and 0.833 when taking into account the differences in the number of repeats, while it shows three alleles at frequencies 0.7, 0.2 and 0.1 when these are scored contemplating indels and point mutations (data not shown).

Using the equations described in Materials and methods, *M* values were estimated from average heterozygosities under each mutation model ($M_{IP} = 1.455$ and $M_{NR} = 0.58828$), with the ratio M_{IP}/M_{NR} being 2.473. Differences between *M* values were considered significant because average H_{IP} was significantly higher than average H_{NR} (Table 5) (Mann–Whitney *U* test, $U = 12$, $P = 0.01$).

Table 5 Number of alleles and heterozygosities used for the estimation of the rate of indels and point mutations relative to stepwise-like mutation events

Locus	Number of alleles IP[a]	H_{IP}[a]	Number of alleles NR[b]	H_{NR}[b]
phi037	3	0.258	3	0.385
phi029	6	0.654	2	0.224
nc035	8	0.687	2	0.495
bnlg1700	7	0.568	6	0.633
bnlg1965	3	0.460	2	0.278
phi069	5	0.747	2	0.484
phi121	2	0.488	1	0
bnlg1209	13	0.728	5	0.400
phi059	7	0.742	1	0
Average	6.000	0.593	2.667	0.322

[a] Number of alleles and heterozygosity based on the alleles determined by indels and point mutations

[b] Number of alleles and heterozygosity based on the alleles determined by differences in the number of repeats. Locus *phi127* was monomorphic under both criteria and was not included in the analysis

Discussion

Microsatellite markers have been shown to deviate from the expectations of SMMs in a wide range of organisms (Anderson et al. 2000; Angers and Bernatchez 1997; Colson and Goldstein 1999; Makova et al. 2000; Orti et al. 1997). The results of the present study constitute a new example of the different factors that may be involved in maize microsatellite variation. Out of ten loci examined here, only two showed allele sequence differences strictly compatible with the SMM, while a diverse array of indels and substitution events, inside and outside the repeat motif, were responsible for allele size variation at the remaining loci. These observations are in agreement with the high levels of polymorphism detected by Mogg et al. (2002) in the characterization of the flanking regions of maize microsatellites, and are mostly concordant with the mutational behaviour previously reported by Matsuoka et al. (2002a) on the basis of allele size distributions. The argentinean landraces studied here and the inbreds used by Matsuoka et al. (2002) showed a similar pattern of variation at locus *phi059*. Although one microsatellite repeat change was observed among electromorphs, indels in the flanking regions were more frequent and thus responsible for allele size variation. On the other hand, in spite of having been described as strictly stepwise by Matsuoka et al. (2002), sequence analysis of locus *phi121* revealed a degraded repeat array, with differences in allele sizes being generated by insertion/deletion events outside the repeat motif. Similarly, locus *bnlg1209*, which was previously deduced to evolve in a stepwise fashion according to mutation accumulation experiments (Vigouroux et al. 2002), was shown here to conceal a compound repetitive structure that would have gone undetected without sequence analysis of existing alleles (Fig. 1).

Interruptions of microsatellite repetitive tracts are known to reduce the propensity of slippage and, therefore, the generation of polymorphism (Estoup and Cornuet 1999). Although the existence of selective pressures cannot be rigorously ruled out, the pattern of variability detected here at locus *phi121*, where only two alleles are observed and only two consecutive repeats are present at allele 99, seems to be concordant with these observations. In contrast, those loci that were found here to contain perfect repetitive arrays showed the highest number of alleles (Fig. 1; Table 3).

Microsatellite loci with higher repeat counts appear to be associated with higher mutation rates as suggested by direct analysis of mutation events (Beck et al. 2003; Brohede et al. 2002; Primmer et al. 1998). Furthermore, a recent analysis of microsatellite evolution in *Arabidopsis thaliana* revealed that, as predicted, low diversity loci were frequently associated with interruptions within the repeat region and that they showed higher rates of rejection to the SMM than high diversity loci (Symonds and Lloyd 2003). In agreement with these findings, the majority of the loci shown here to deviate from SMM expectations possess short or interrupted repeats, while the only two loci compatible with stepwise mutations exhibit the highest number of repeat units (Fig. 1) and the highest number of alleles (Table 3).

Size homoplasy

To the best of our knowledge, this is the first report on maize microsatellite size homoplasy at the population level. Earlier evidences of allele size convergence in maize were limited to 11 temperate inbreds (Mogg et al. 2002). According to these authors, approximately 10% of maize microsatellite loci exhibited allele size homoplasy. However, as shown here, this phenomenon is even more widespread than previously deduced because the coexistence of different sequence variants under the same size category was apparent at all ten loci included in this study.

As previously mentioned, the occurrence of size homoplasy is tightly linked to the way mutations produce new alleles and hence to the mutation model. Theoretically, ignoring microsatellite size homoplasy in evolutionary and population genetic studies can lead to the underestimation of genetic diversity and may have a profound impact on gene flow estimates and phylogenetic inferences. Although computer simulations have shown allele size convergence to be of concern only in those instances where high mutation rates, large populations sizes or strong allele size constraints are encountered (Estoup et al. 2002), empirical evidence has yielded somewhat contradictory results. Sequence analysis of microsatellite electromorphs from five interrupted and/or compound loci in three invertebrate species showed that uncovering size homoplasy altered phylogenetic reconstructions and produced higher estimates of population structure (Viard et al. 1998). Conversely, comparison of SSCP conformers and electromorphs for one microsatellite locus in a freshwater snail suggested that size homoplasy does not invariably produce higher estimates of population structure (Angers et al. 2000). In plants, sequencing of electromorphs from 12 microsatellite loci in the tropical tree *Corythophora alta* revealed that size homoplasy had little to no impact on the estimates of population differentiation, and that the effects of size homoplasy were comparable to those derived from other sources of data error (Adams et al. 2004). However, simulation studies conducted on chloroplast microsatellites have shown size homoplasy to have a considerable influence on several measures of genetic distance and diversity (Navascués and Emerson 2005).

It is generally assumed that if two microsatellite alleles of the same size differ at one or more flanking nucleotides,

these alleles are homoplastic, having converged by stepwise mutation to a common repeat number along separate flanking sequence lineages. This rationale derives from consideration of the relative contribution of point mutations and replication slippage to microsatellite variability. A recent estimation of maize mutation rates have shown point mutations to be of the order of 3×10^{-8} events *per* locus *per* generation (Clark et al. 2005), while mutation rates proposed for microsatellites lie between 5.2×10^{-4} and 1.1×10^{-3} for dinucleotide repeat motifs and are less than 5.1×10^{-5} for repeat motifs of more than 2 bp (Vigouroux et al. 2002). Under a non-stepwise scenario, with replication slippage having only a minor impact, if any, on the generation of polymorphism, the balance between the mutational forces operating on microsatellite diversity differs from that presented above. Therefore, the coexistence of more than one sequence variant under the same size category may either be a consequence of convergence or the product of divergence within the same size lineage (i.e. the accumulation of point substitutions for a given electromorph). Considering that the majority of the loci studied here evolve in a non-stepwise fashion, and that no statistical differences were observed between the number of conformers at non-SMM and SMM-compatible loci, both convergence and divergence phenomena may be responsible for the high levels of homoplasy detected in maize landraces. Because convergence and divergence lead to opposite errors in the computation of allele frequencies, determining which one is acting on the generation of size homoplasy may have a significant impact on the analysis of population data. Consideration of the estimate obtained here for the mutation rate of indels and point substitutions relative to stepwise-like mutations may help discriminate between both phenomena. Relying on our estimate of the M_{IP}/M_{NR} ratio, and given that significant differences were found between the average heterozygosities used to calculate each of these M values, it seems that the mutation rate of indels and point mutations is approximately two-and-a-half times that of stepwise-like mutation events. In this context, the accumulation of substitutions on a lineage determined by a given number of repeats becomes more plausible than the change in the number of repeats in a given flanking sequence lineage. Therefore, it appears that divergence, rather than convergence, is driving size homoplasy at the loci included in this study.

In agreement with the non-stepwise nature of most of the loci studied the ratio of mutation rates obtained here differs from the ratio generally assumed for microsatellite evolution, in which indels and point mutations are less frequent than changes in the number of repeat units. It has been proposed that an indel-like, length-independent slippage process (indel slippage) governs the evolution of short microsatellite loci (Dieringer and Schlotterer 2003). Moreover, short microsatellites are known to be more stable than large microsatellites (for a review, see Ellegren 2004) suggesting that the rate of indel slippage events is probably lower than that of replication slippage. Interestingly, the majority of the loci studied in this work fall within the short category.

It is interesting to point out that heterozygosity estimates are substantially increased when the alleles generated by indels and point mutations are considered (Table 5). Therefore, the present data show that, at least for a fraction of microsatellite markers, relying solely on the allelic differences derived from the changes in the number of repeat units may lead to an underestimation of the potential of a population as source of variability.

According to the review by Estoup et al. (2002), size homoplasy is usually higher among species than between populations of the same species, and rarer within populations. However, the average number of conformers *per* electromorph (C/E ratio) was higher than one in all the populations included in this study (Table 4), indicating that size homoplasy is very frequent within populations of maize landraces. The absence of significant differences among the C/E ratios of the populations being compared may be interpreted as a consequence of similar mutation rates operating on the generation of conformers. However, many conformers were shared among populations indicating they arose prior to population differentiation or that they were introduced by means of gene flow at similar rates.

The high levels of size homoplasy found within maize populations may be explained partially by the distinctive dynamics of maize cultivation and exchange in native communities. In maize research, the term *population* often refers to open-pollinated cultivars that are maintained by local farmers. Human activities are thus the primary means of dispersal, with seed exchange greatly increasing the levels of gene flow among landraces, and promoting the introduction of new alleles to broader geographic areas. Additionally, commercial inbreds are frequently used by farmers to improve ear quality in local germplasm, which also provides an alternative source of alleles contributing to the coexistence of different conformers under the same electromorph category.

Although it may be argued that the complex mutational processes described in this study represent an exception rather than the rule for maize microsatellite loci, it is worth mentioning that the majority of these loci are still commonly used in maize investigations and have been included recently in a series of genetic surveys aimed at detecting population structure and establishing phylogenetic relationships among maize landraces (e.g., Laborda et al. 2005; Pressoir and Berthaud 2004; Reif et al. 2004, 2005; Santacruz-Varela et al. 2004; Warburton et al. 2002). To what extent the findings presented here may alter the conclusions drawn from

microsatellite data in maize literature still remains to be determined. However, future studies will certainly benefit from preliminary surveys of size homoplasy, in order to avoid using non-homologous alleles when estimating population parameters in a phylogenetic framework, particularly when a small number of loci are being assessed.

Acknowledgments We thank the late Dr. C. Naranjo, Ing. J. Cámara Hernández and Ing. A.M. Miante Alzogaray for providing us with the material and for carrying out the taxonomic classification of the landraces included in this study. We are also in debt to the reviewers who have greatly improved this manuscript. Several grants from the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 02296), the University of Buenos Aires (EX–317) and the Agencia Nacional de Promoción Científica y Tecnológica (BID 1201 OC-AR PICT 04443) are gratefully acknowledged.

References

Adams RI, Brown KM, Hamilton MB (2004) The impact of microsatellite electromorph size homoplasy on multilocus population structure estimates in a tropical tree (*Corytophora alta*) and an anadromous fish (*Morone saxatilis*). Mol Ecol 13:2579–2588

Anderson TJC, Su X-Z, Roddam A, Day KP (2000) Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. Mol Ecol 9:1599–1608

Angers B, Bernatchez L (1997) Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Mol Biol Evol 14:230–238

Angers B, Estoup A, Jarne P (2000) Microsatellite size homoplasy, SSCP, and population structure: a case study. Mol Biol Evol 17:1926–1932

Beck NR, Double MC, Cockburn A (2003) Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees. Mol Biol Evol 20:54–61

Blankenship SM, May B, Hedgecock D (2002) Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. Mol Biol Evol 19:1943–1951

Brohede J, Primmer CR, Moller A, Ellegren H (2002) Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. Nucleic Acids Res 30:1997–2003

Clark RM, Tavare S, Doebley J (2005) Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. Mol Biol Evol 22:2304–2312

Colson I, Goldstein DB (1999) Evidence for complex mutations at microsatellite loci in *Drosophila*. Genetics 152:617–627

Culver M, Menotti-Raymond MA, O´Brien S (2001) Patterns of size homoplasy at 10 microsatellite loci in Pumas (*Puma concolor*). Mol Biol Evol 18:1151–1156

Dellaporta IK, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21

Di Rienzo A, Peterson AC, Garza JC et al (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Dieringer D, Schlotterer C (2003) Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. Genome Res 13:2242–2251

Doyle JJ, Morgante M, Tingey SV, Powell W (1998) Size homoplasy in chloroplast microsatellites of wild perennial relatives of Soybean (*Glycine* subgenus *Glycine*). Mol Biol Evol 15:215–218

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445

England PR, Usher AV, Whelan RJ, Ayre DJ (2002) Microsatellite diversity and genetic structure of fragmented populations of the rare, fire-dependent shrub *Grevillea macleayana*. Mol Ecol 11:967–977

Estoup A, Cornuet JM (1999) Microsatellite evolution: inferences from population data. In: Goldstein D, Schlotterer C (eds) Microsatellites: evolution and applications. Oxford University Press, New York, pp 49–65

Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604

Hale ML, Borland AM, Gustafsson MHG, Wolff K (2004) Causes of size homoplasy among chloroplast microsatellites in closely related *Clusia* species. J Mol Evol 58:182–190

Higgins DG, Thompson JD, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Kimura M, Ohta T (1978) Stepwise mutation model and distribution of allelic frequencies in finite populations. Proc Natl Acad Sci USA 75:2868–2872

Laborda PR, Oliveira KM, Garcia AAF, Paterniani ME, de Souza AP (2005) Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers? Theor Appl Genet 111:1288–1299

Le Clerc V, Bazante F, Baril C, Guiard J, Zhang D (2005) Assessing temporal changes in genetic diversity of maize varieties using microsatellite markers. Theor Appl Genet 110:294–302

Lia VV, Confalonieri VA, Ratto N, Cámara Hernández J, Miante-Alzogaray AM, Poggio L, Brown TA (2007) Microsatellite typing of ancient maize: insights into the history of agriculture in southern South America. Proc Biol Sci 274(1609):545–554

Makova K, Nekrutenko A, Baker RJ (2000) Evolution of microsatellite alleles in four species of mice (Genus *Apodemus*). J Mol Evol 51:166–172

Matsuoka Y, Mitchell SE, Kresovich S, Goodman MM, Doebley JF (2002a) Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450

Matsuoka Y, Vigouroux Y, Goodman MM et al (2002b) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084

Mogg R, Batley J, Hanley S, Edwards D, O'Sullivan H, Edwards KJ (2002) Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. Theor Appl Genet 105:532–543

Navascués M, Emerson BC (2005) Chloroplast microsatellites: measures of genetic diversity and the effect of homoplasy. Mol Ecol 14:1333–1341

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T (1989) Detection of polymorphisms of human DNA by gel electrophoresis as Single-Strand Conformation Polymorphisms. Proc Natl Acad Sci USA 86:2766–2770

Orti G, Pearse DE, Avise JC (1997) Phylogenetic assessment of length variation at a microsatellite locus. Proc Natl Acad Sci USA 94:10745–10749

Pressoir G, Berthaud J (2004) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94

Primmer CR, Saino N, Moller AP, Ellegren H (1998) Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. Mol Biol Evol 15:1047–1054

Reif JC, Xia XC, Melchinger AE et al (2004) Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. Crop Sci 44:326–334

Reif JC, Hamrit S, Heckenberger M et al (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyses of individuals and bulks. Theor Appl Genet 111:906–913

Saltonstall K (2003) Microsatellite variation within and among North American lineages of *Phragmites australis*. Mol Ecol 12:1689–1702

Santacruz-Varela A, Widrlechner MP, Ziegler KE et al (2004) Phylogenetic relationships among North American Popcorns and their evolutionary links to Mexican and South American Popcorns. Crop Sci 44:1456

Stumm G, Vedder H, Schlegel J (1997) A simple method for isolation of PCR fragments from silver stained polyacrylamide gels by fine needle scratching. Elsevier Trends Journals Technical Tips On Line. http://www.tto.biomednet.com/cgi-bin/tto/pr

Sunnucks P, Wilson A, Beheregaray LB et al (2000) SSCP is not so difficult: the application and utility of single-stranded conformation polymorphism in evolutionary biology and molecular ecology. Mol Ecol 9:1699–1701

Symonds VV, Lloyd AM (2003) An analysis of microsatellite loci in *Arabidopsis thaliana*: mutational dynamics and application. Genetics 165:1475–1488

Taylor JS, Sanny JSP, Breden F (1999) Microsatellite allele size homoplasy in the Guppy (*Poecilia reticulata*). J Mol Evol 48:245–247

van Oppen JH, Rico C, Turner GF, Hewitt GM (2000) Extensive homoplasy, nonstepwise mutations and ancestral polymorphisms at a complex microsatellite locus in lake Malawi Ciclids. Mol Biol Evol 17:489–498

Viard F, Franck P, Dubois M, Estoup A, Jarne P (1998) Variation at microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. J Mol Evol 47:42–51

Vigouroux Y, Jaqueth JS, Matsuoka Y et al (2002) Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol 19:1251–1260

Warburton ML, Xianchua X, Crossa J et al (2002) Genetic characterization of CIMMYT inbred lines and open pollinated populations using large scale fingerprinting methods. Crop Sci 42:1832–1840

Xie H, Sui Y, Chang F-Q, Xu Y, Ma R-C (2006) SSR allelic variation in almond (*Prunus dulcis* Mill.). Theor Appl Genet 112:366–372

Yokoyama J, Fukuda T, Yokoyama A, Nakajima M (2004) Extensive size homoplasy at a microsatellite locus in the Japanese bumblebee, *Bombus diversus*. Entomol Sci 7:189–197