Theor Appl Genet (2002) 104:436–450

Y. Matsuoka · S.E. Mitchell · S. Kresovich
M. Goodman · J. Doebley

Microsatellites in *Zea* – variability, patterns of mutations, and use for evolutionary studies

Received: 10 February 2001 / Accepted: 21 May 2001

Abstract To evaluate the performance of microsatellites or simple sequence repeats (SSRs) for evolutionary studies in *Zea*, 46 microsatellite loci originally derived from maize were applied to diverse arrays of populations that represent all the diploid species of *Zea* and 101 maize inbreds. Although null phenotypes and amplification of more than two alleles per plant were observed at modest rates, no practical obstacle was encountered for applying maize microsatellites to other *Zea* species. Sequencing of microsatellite alleles revealed complex patterns of mutation including frequent indels in the regions flanking microsatellite repeats. In one case, all variation at a microsatellite locus came from indels in the flanking region rather than in the repeat motif. Maize microsatellites show great variability within populations and provide a reliable means to measure intraspecific variation. Phylogeographic relationships of *Zea* populations were successfully reconstructed with good resolution using a genetic distance based on the infinite allele model, indicating that microsatellite loci are useful in evolutionary studies in *Zea*. Microsatellite loci show a principal division between tropical and temperate inbred lines, and group inbreds within these two broad germplasm groups in a manner that is largely consistent with their known pedigrees.

Communicated by A.L. Kahler

Y. Matsuoka · J. Doebley (✉)
Genetics Department, 445 Henry Mall, University of Wisconsin, Madison, WI 53706, USA
e-mail: jdoebley@facstaff.wisc.edu

S.E. Mitchell · S. Kresovich
Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853, USA

M. Goodman
Department of Crop Science, North Carolina State University, Raleigh, NC 27695, USA

Present address:
Y. Matsuoka, Fukui Prefectural University, Matsuoka-cho, Yoshida-gun, Fukui, 910–1195, Japan

Keywords Teosinte · SSR · Indel · Genetic diversity · Evolution

Introduction

Microsatellites, or simple sequence repeats (SSRs), consist of varying numbers of tandemly repeated units (1 to 6 base pairs each) and represent a class of repetitive DNA that is commonly found in eukaryotic genomes (Tautz and Renz 1984). They are characterized by great abundance (Condit and Hubbell 1991; Röder et al. 1995), high variability (Schug et al. 1998), and even distribution throughout the genomes in many species (Liu et al. 1996; Taramino and Tingey 1996; Röder et al. 1998). Microsatellites are typically multi-allelic loci, and loci with more than five alleles are commonly observed in plants (Innan et al. 1997; Senior et al. 1998) and animals (MacHugh et al. 1997). In addition, automated PCR-based techniques, which enable high-throughput data collection and good analytical resolution at a low cost, have been developed for microsatellites (Kresovich et al. 1995; Mitchell et al. 1997). Because of these qualities, microsatellites are now one of the preferred genetic markers in plants and animals.

Microsatellites have been exploited as tools to measure genetic distance and diversity in evolutionary studies (Bruford and Wayne 1993; Goldstein and Pollock 1997). Their power for these analyses comes from their characteristically high allelic diversity, which in turn is a product of their high rate of stepwise mutation due to replication slippage (Levinson and Gutman 1987). For this reason, several measures of genetic distance have been developed for microsatellites on the basis of the stepwise mutation model (SMM; Kimura and Crow 1964). The SMM assumes that alleles mutate back and forth by small numbers of repeats, and thus the same allelic states are created repeatedly over time. The SMM-based genetic distances for microsatellites have successfully been applied in evolutionary studies in animals (Goldstein et al. 1995, 1999). An alternative model is the infinite alleles

model (IAM; Ohta and Kimura 1973), which assumes that each mutation creates a new allele in the population.

Zea is the genus for maize (*Zea mays* ssp. *mays*) and its wild relative, teosinte. There are four species (*Zea diploperennis*, *Zea perennis*, *Zea luxurians*, and *Zea mays*) in this genus and they are all native to Mexico and Central America (Doebley and Iltis 1980; Iltis and Doebley 1980). Four subspecific taxa have been recognized in *Z. mays*, namely, *Z. mays* ssp. *huehuetenangensis*, *Z. mays* ssp. *mays*, *Z. mays* ssp. *mexicana*, and *Z. mays* ssp. *parviglumis*. The inter- and intra-specific relationships have been studied extensively by morphological and molecular systematic methods (Doebley 1990; Buckler and Holtsford 1996). Among the three wild subspecies of *Z. mays*, *Z. mays* ssp. *parviglumis* is thought to be the progenitor of cultivated maize (Doebley et al. 1984; Wang et al. 1999).

Previous studies on maize microsatellites have shown that they are abundant and highly variable (Senior and Heun 1993; Chin et al. 1996; Taramino and Tingey 1996), that they provide powerful tools for genome mapping, individual genotyping and germplasm evaluation (Smith et al. 1997; Senior et al. 1998), and that some microsatellites from maize are applicable to teosinte (Lübberstedt et al. 1998). To further evaluate the utility of microsatellites for evolutionary studies in *Zea*, we tested the performance of 59 microsatellites on a diverse array of populations and inbred lines of *Zea*. Our analyses show that microsatellites are powerful phylogenetic markers for both inter- and intra-specific studies in this genus. However, microsatellite variability in *Zea* does not fit the SMM because much of the observed variation appears to result from indels in regions flanking the repeat motif.

Material and methods

Plant materials

Sixteen open-pollinated populations of teosinte and maize landraces (six plants each) representing all the diploid species of *Zea* and 101 inbreds (one plant each) representing three major germplasm sources (Tropical inbreds, USA inbreds, and Canada/Europe inbreds) were analyzed (Table 1). DNA was extracted from individuals using a modified CTAB method (Saghai-Maroof et al. 1984; Doebley and Stec 1991).

Microsatellite markers and PCR amplification

The microsatellites and multiplex sets used in this study are listed in Table 2. Primer sequences, chromosomal location, and repeat motifs are available from the maize database (http://www.agron.missouri.edu/ssr.html). Thirteen microsatellites were tested but failed to consistently amplify and were dropped from our study: namely, phi011, phi014, phi022, phi024, phi041, phi069, phi070, phi078, phi100175, phi101049, phi108411, phi265454 and phi346482. Multiplex PCRs for each amplification set were performed in 20-µl volumes containing 25 ng of template DNA, 1–4 pmol of each forward and reverse primer, 1×PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% Triton X-100), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 0.5 U of *Taq* polymerase. Thermocycling consisted of initial denaturation of the template DNA at 95°C for 4 min followed by 25 cycles of 95°C for 1 min, 55°C for 2 min, and 72°C for 2 min. In the last cycle, extension time at 72°C was increased to 10 min.

Electrophoresis and detection

Samples containing 0.5 µl of PCR products, 0.1 µl of GeneScan 500XL size standard (PE Biosystems) and 1.0 µl of loading buffer (1 part de-ionized formamide: 4 parts 500 mg/ml blue dextran, 25 mM EDTA) were heated at 92°C for 5 min, then placed on ice. Denatured samples (0.6 µl) were immediately loaded on 5% denaturing (6 M urea) LongRanger (FMC) gels (36 cm well-to-read) in 1×TBE buffer (89 mM Tris, 89 mM borate, 2 mM EDTA pH 8.3) and electrophoresed at 3,000 V for 3 h on an automated DNA sequencer (PE Biosystems, model 377). Gels were run in either a 64-or 96-well format.

Fragment analysis

Fragment sizes were determined based on migration relative to the internal lane size standards using GeneScan v. 2.1 software (PE Biosystems) and the "local Southern" sizing algorithm. Data were scored and compiled using Genotyper v. 3.0 (PE Biosystems). The GeneScan and Genotyper software provide estimates of fragment size to two decimal places. The distribution of fragment sizes for each microsatellite locus was not continuous, but possessed "natural" discontinuities or breaks which we used to define sets of bins, i.e., alleles. For example, fragments of sizes 157.34 to 157.85 might be designated as allele "157" when relatively large breaks separated them and neighboring bins. We applied two criteria in

Table 1 Teosinte and maize landraces used in this study

Taxon	Source	Accession	Country	OTU
Z. diploperennis	H. Iltis	1,250	Mexico	DIP-LJ1
Z. luxurians	CIMMYT	9,478	Guatemala	LUX-IP1
Z. mays ssp. *huehuetenangensis*	CIMMYT	9,479	Guatemala	HUE-SA1
Z. mays ssp. *mexicana*	CIMMYT	8,771	Mexico	MEX-CZ1
	CIMMYT	11,400	Mexico	MEX-LR1
	CIMMYT	8,749	Mexico	MEX-MF1
	CIMMYT	11,387	Mexico	MEX-NO1
	CIMMYT	11,369	Mexico	MEX-UR1
Z. mays ssp. *parviglumis*	CIMMYT	8,756	Mexico	PAR-MZ1
	CIMMYT	8,755	Mexico	PAR-MZ2
	CIMMYT	8,781	Mexico	PAR-TL1
	CIMMYT	8,780	Mexico	PAR-TL2
Z. mays ssp. *mays*	CIMMYT	MEX 48	Mexico	MAY-CQ1
	INIA	GTO 73	Mexico	MAY-CN1
	ICA	VEN 604	Venezuela	MAY-CA1
	ICA	VEN 453	Venezuela	MAY-CO1

defining bins. First, the range of fragment sizes within a bin should not exceed 1.00 bp. Second, the gap between bins should be substantially greater than the gaps between fragment sizes within bins. With a few exceptions for which some judgement was needed, the fragment sizes sorted naturally into well-defined bins or alleles.

Microsatellite sequencing

Microsatellite loci were amplified by PCR using PCR Supermix (BRL) and 10 pmol of each primer as described by Chin et al (1996). PCR products were purified with a column (Qiagen) and sequenced with the primers used for PCR. Sequencing was done for both strands with a BigDye terminator cycle sequencing kit (ABI) and an automated sequencer at the University of Wisconsin Biotechnology Center (Madison). All sequences have been submitted to Genbank (Accessions: AY033441–AY033471).

Gene diversity, genetic distances, and tree construction

Gene diversity was calculated at each locus according to $2n(1-\Sigma p^2_i)/(2n-1)$ for teosinte and maize landraces, and $n(1-\Sigma p^2_i)/(n-1)$ for inbreds, where n is the number of samples and p_i is the frequency of the i^{th} allele (Nei 1973). We used CLUSTAL W (Thompson et al. 1994) for sequence alignment, Microsat (http://lotka.stanford.edu/microsat/microsat.html) for Nei's standard genetic distance calculation (Nei 1972), Ldist (Kejun Liu, unpublished) for Rogers' distance (Rogers 1972), Cavalli-Sforza and Edwards's chord distance (Cavalli-Sforza and Edwards 1967), and Nei et al.'s chord distance (Nei et al. 1983) calculation, and PHYLIP (Felsenstein 1993) for tree construction using the Fitch-Margoliash method.

Results

Fidelity of microsatellite amplification

A total of 59 microsatellites were tested, of which 46 were found to be usable in terms of the low frequency of null phenotypes (<5%) and high reproducibility. Thirty eight microsatellites amplified well (<5% nulls) for all *Zea* samples tested, while four worked well for teosinte and maize landraces (open-pollinated lines) but not for the inbreds, and four others worked well for the inbreds but not the teosintes. These data are available at http://www.stat.ncsu.edu/~panzea. There were 528 alleles at 42 loci in the open-pollinated lines and 288 alleles at 42 loci in the inbreds. Null phenotypes were observed both in teosinte and maize landraces, and inbreds. Thirty two out of the 59 loci exhibited some null phenotypes with the frequency of null plants ranging from 1.0 to 62.5% for teosinte and maize landraces (average 6.6%) and from 1.0 to 49.5% for the inbreds (average 5.5%).

Null phenotypes pose a problem for data analysis. They may indicate either failed PCR reactions or mutations in the primer sites (i.e., null alleles). In the former case, designation of a failed reaction as a null allele is inappropriate. In the latter case, independent mutations that alter either primer site will be inappropriately scored as the same null allele. An additional problem is that plants appearing homozygous for a scorable allele may ac tually represent a heterozygote with a scorable and a null allele. These types of errors will be negligible when null phenotypes are infrequent. Although the criterion (<5% nulls) that we set for including a microsatellite locus in our study is arbitrary, the inclusion of loci with up to 5% nulls does not appear to bias the data significantly (see below).

Of the 42 useful microsatellite loci, 18 amplified more than two products in a total of 33 out of 96 plants of teosinte and maize landraces (1.3% of all data points). In the inbreds, one plant possessed three alleles for one microsatellite (0.02% of all data points). Because maize has a large number of duplicated loci (Helentjaris et al. 1988), the observation of three or more alleles may represent case s where more than one locus is being amplified in the PCR reaction. Alternatively, they may be the result of DNA template contamination and/or PCR artifacts. In either case, because the true alleles could not be distinguished from the aberrant ones, plant-microsatellite combinations with more than two products were treated as missing data.

Mutation patterns in *Zea* microsatellites

We examined the distributions of allele sizes (in base pairs) for the 46 microsatellite loci and compared those distributions to the expectation that allelic differences are the result of changes in the number of repeats at the locus. Contrary to expectations, most microsatellite loci do not fit the model that allelic differences would be in stepwise multiples of the number of base pairs in the repeat unit. Rather, of the 46 microsatellite loci, only two exhibited a purely stepwise allelic distribution, while four were nearly stepwise, 13 mixed (stepwise and continuous), eight nearly continuous and 19 continuous (Table 2). The representative distributions of each of the categories are shown in Fig. 1. The microsatellite loci of the last three categories (40 out of 46) failed to show patterns which fit the expected pattern for microsatellites, while those of the first two categories fit the expected pattern, either perfectly or almost perfectly. These results suggest that most of the microsatellite loci have not evolved in a stepwise fashion by changes in copy number of the repeat.

The distribution patterns of allele sizes also showed the presence of outliers, i.e., alleles that are substantially different in size from the majority of alleles at a particular microsatellite locus (Fig. 1). It was not clear whether the outliers were the alleles of the target locus, as they might have been derived from duplicate loci elsewhere in the genome.

To elucidate what underlies microsatellite variation in *Zea*, 31 alleles of six microsatellite loci from each distribution pattern were sequenced: phi059 (8 alleles), phi083 (5), phi033 (4), phi96100 (5), phi102228 (5), and phi121 (4) (Genbank accessions AY033441-AY033471). The DNA sequences revealed complex mutation patterns resulting from insertions/deletions (indels) in the regions flanking the repeat motifs. For example, in the case of phi059 (mixed), although one microsatellite repeat change was observed (CML254), the indels in the flanking regions were more frequent and

Table 2 Types of allele size distribution, multiplex sets, numbers of alleles, allelic range, and gene diversity for 46 microsatellite markers used for teosinte and maize landraces (TM) and inbreds (IN)

Marker	Multiplex	Plant type	Type of allele size distribution	Number of alleles (TM/IN)	Allelic range (TM/IN)	Gene diversity (TM)	Gene diversity (IN)
bngl653	2–3	TM/IN	Continuous	24/8	140–176/151–160	0.88	0.76
phi015	1–3	TM/IN	Continuous	21/11	76–113/83–104	0.84	0.70
phi029	2–2	TM/IN	Continuous	20/7	139–176/148–161	0.83	0.73
phi032	1–3	TM/IN	Continuous	13/5	232–246/233–242	0.76	0.73
phi033	1–1	TM/IN	Continuous	16/12	237–270/224–263	0.80	0.49
phi034	1–2	TM/IN	Nearly continuous	13/8	123–160/123–148	0.84	0.74
phi050	1–3	TM/IN	Mixed	7/4	77–87/81–87	0.77	0.49
phi051	1–1	TM/IN	Continuous	13/8	137–154/139–148	0.82	0.71
phi053	1–2	IN	Mixed	9	169–212	N/A[a]	0.74
phi056	1–3	TM/IN	Continuous	19/6	231–278/241–258	0.87	0.72
phi059	2–4	TM/IN	Mixed	11/7	117–161/117–161	0.64	0.59
phi062	1–1	TM/IN	Nearly stepwise	7/2	158–178/161–164	0.65	0.46
phi064	1–3	TM/IN	Continuous	20/14	75–121/75–110	0.92	0.89
phi072	1–2	TM/IN	Continuous	19/9	134–163/143–163	0.86	0.79
phi073	2–2	TM	Mixed	9	186–203	0.80	N/A
phi076	2–4	TM	Mixed	10	147–179	0.73	N/A
phi079	–	IN	Mixed	6	181–195	N/A	0.70
phi083	1–3	TM/IN	Nearly continuous	11/6	124–367/126–139	0.81	0.76
phi085	1–2	TM	Mixed	14	233–266	0.85	N/A
phi093	1–2	TM/IN	Continuous	19/12	272–296/284–294	0.91	0.85
phi115	1–1	TM/IN	Mixed	4/3	291–305/292–312	0.47	0.47
phi121	1–2	TM/IN	Stepwise	5/2	93–105/99–102	0.22	0.21
phi127	1–1	TM/IN	Nearly continuous	10/7	105–128/112–128	0.80	0.70
phi101249	2–1	TM	Nearly continuous	15	114–161	0.75	N/A
phi102228	2–3	TM/IN	Nearly stepwise	4/3	124–132/124–132	0.60	0.45
phi104127	2–3	TM/IN	Nearly continuous	11/6	132–170/132–165	0.73	0.66
phi109188	2–1	TM/IN	Continuous	17/10	148–180/148–171	0.87	0.60
phi109275	2–2	TM/IN	Continuous	15/7	121–149/122–139	0.88	0.81
phi109642	–	IN	Nearly stepwise	4	135–149	N/A	0.55
phi159819	3–5	TM/IN	Nearly continuous	9/6	119–139/123–139	0.78	0.75
phi213984	3–5	TM/IN	Mixed	9/3	287–320/287–304	0.25	0.29
phi233376	3–3	TM/IN	Continuous	15/10	140–177/137–159	0.81	0.84
phi251315	3–2	IN	Mixed	4	127–137	N/A	0.35
phi308090	3–4	TM/IN	Stepwise	9/5	190–226/211–226	0.73	0.59
phi330507	3–4	TM/IN	Mixed	5/5	133–161/128–145	0.41	0.55
phi333597	3–2	TM/IN	Mixed	5/4	207–227/213–226	0.50	0.65
phi335539	3–3	TM/IN	Nearly continuous	7/5	92–115/92–149	0.49	0.18
phi339017	3–5	TM/IN	Nearly continuous	10/5	138–159/148–157	0.43	0.47
phi389203	3–2	TM/IN	Nearly stepwise	6/5	301–313/301–314	0.72	0.47
phi402893	3–1	TM/IN	Continuous	23/10	205–243/209–240	0.87	0.66
phi427434	3–2	TM/IN	Nearly stepwise	10/6	124–147/124–143	0.76	0.64
phi427913	3–1	TM/IN	Continuous	9/9	117–135/117–207	0.81	0.54
phi448880	3–5	TM/IN	Mixed	7/4	174–191/174–188	0.62	0.48
phi453121	3–3	TM/IN	Continuous	19/10	209–233/209–227	0.91	0.77
phi96100	2–1	TM/IN	Continuous	18/11	219–301/235–300	0.83	0.84
phi96342	2–2	TM/IN	Continuous	20/10	223–256/233–250	0.85	0.78
Average				12.6/6.9		0.73	0.62

[a] N/A: Not applicable

thus responsible for most of the allele-size variation (Fig. 2a). The outlier allele for phi059 from A272 (117 bp) showed strong homology to other alleles, but had a large deletion (39 bp). In the case of phi96100 (nearly continuous), a large deletion (81 bp), which deleted the repeat unit itself, generated the outlier from PAR-MZ1 (219 bp) (Fig. 2b). Another large deletion (56 bp) was found in the outlier from CML333 (235 bp). Additional indels in the flanking regions affected the allele-size variation (Fig. 2b). For phi083 (nearly continuous), the microsatellite repeat itself was invariant and all variation was due to indels in the flanking regions (Fig. 2c). A microsatellite-like stretch was found in the flanking region for this locus. In the case of phi033 (continuous), two different motifs, $(CTT)^n$ and $(AGGCAG)^n$, were found, both contributing to the allele-size variation . For phi121 (stepwise), imperfect repeats were observed in all alleles, although the allele sizes fit the model of changes in the number of repeat units. For phi102228 (Fig. 2d), (nearly stepwise), size variation of sequenced alleles was explained fully by the differences in the numbers of repeat units.

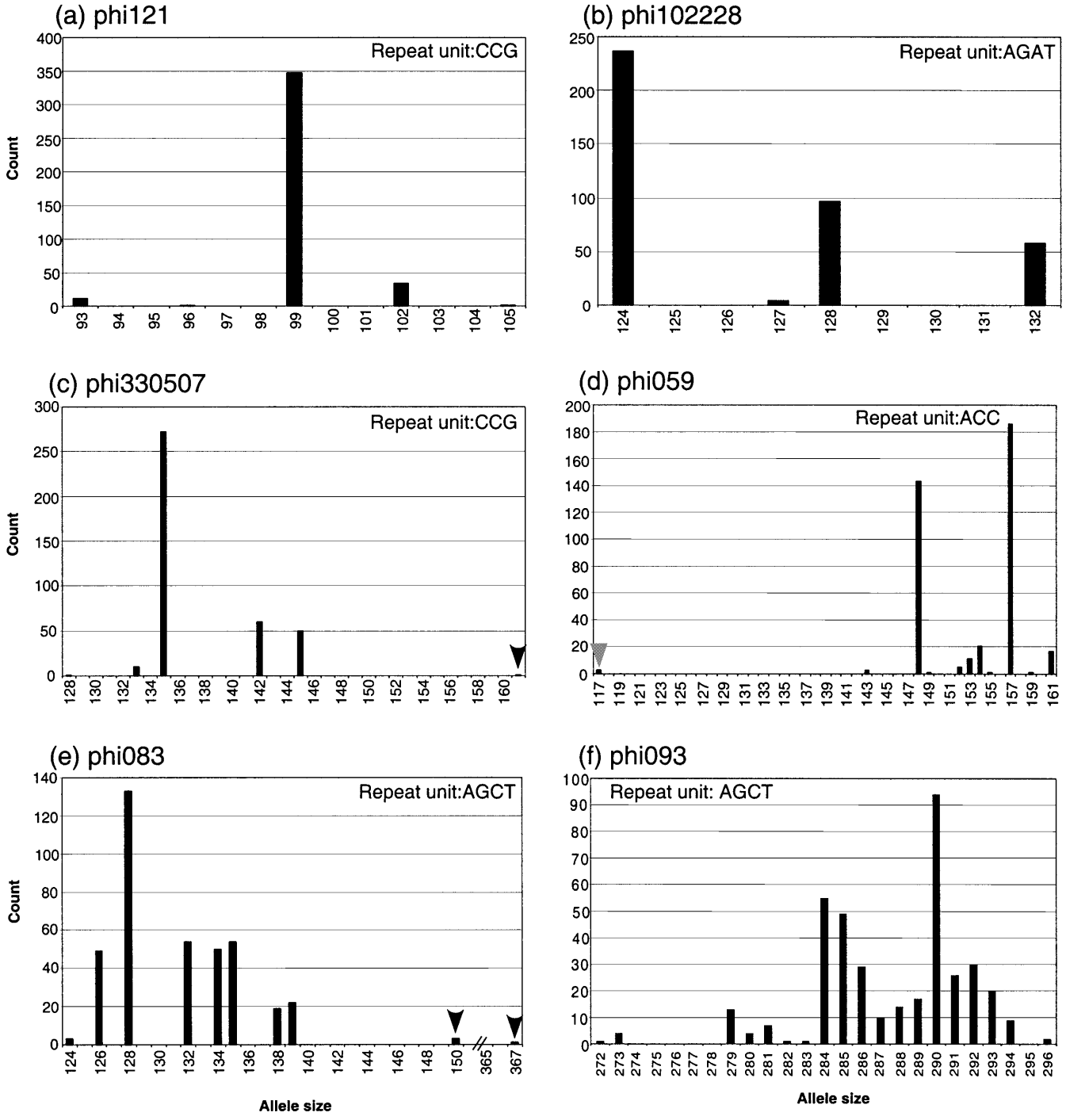


Fig. 1 Representatives of microsatellite allele-size distributions: stepwise (**a**), nearly stepwise (**b**), mixed (**c** and **d**), nearly continuous (**e**), and continuous (**f**). *Solid arrowheads* indicate outliers. Alleles including sequenced outliers are shown by *gray arrowheads*

The allele sequences indicate that the continuous nature of allele-size variation at most maize microsatellite loci is the result of indels in the regions flanking the repeats, and that outlier alleles result from large indels rather than amplification of duplicate loci. Indels in the flanking regions and imperfect repeats seem fairly frequent at maize microsatellite loci. In *Zea*, the mutational process at these loci is probably far more complex than expected from the simple model of changes in the number of repeat units.

Accuracy of allele-size determination and reproducibility

Microsatellite sequencing provided the opportunity to test the accuracy of allele-size determination by fragment analysis on the ABI377. In 20 out of 31 cases (64.5%), the allele sizes were in agreement between fragment analysis and sequencing. For the rest, the allele

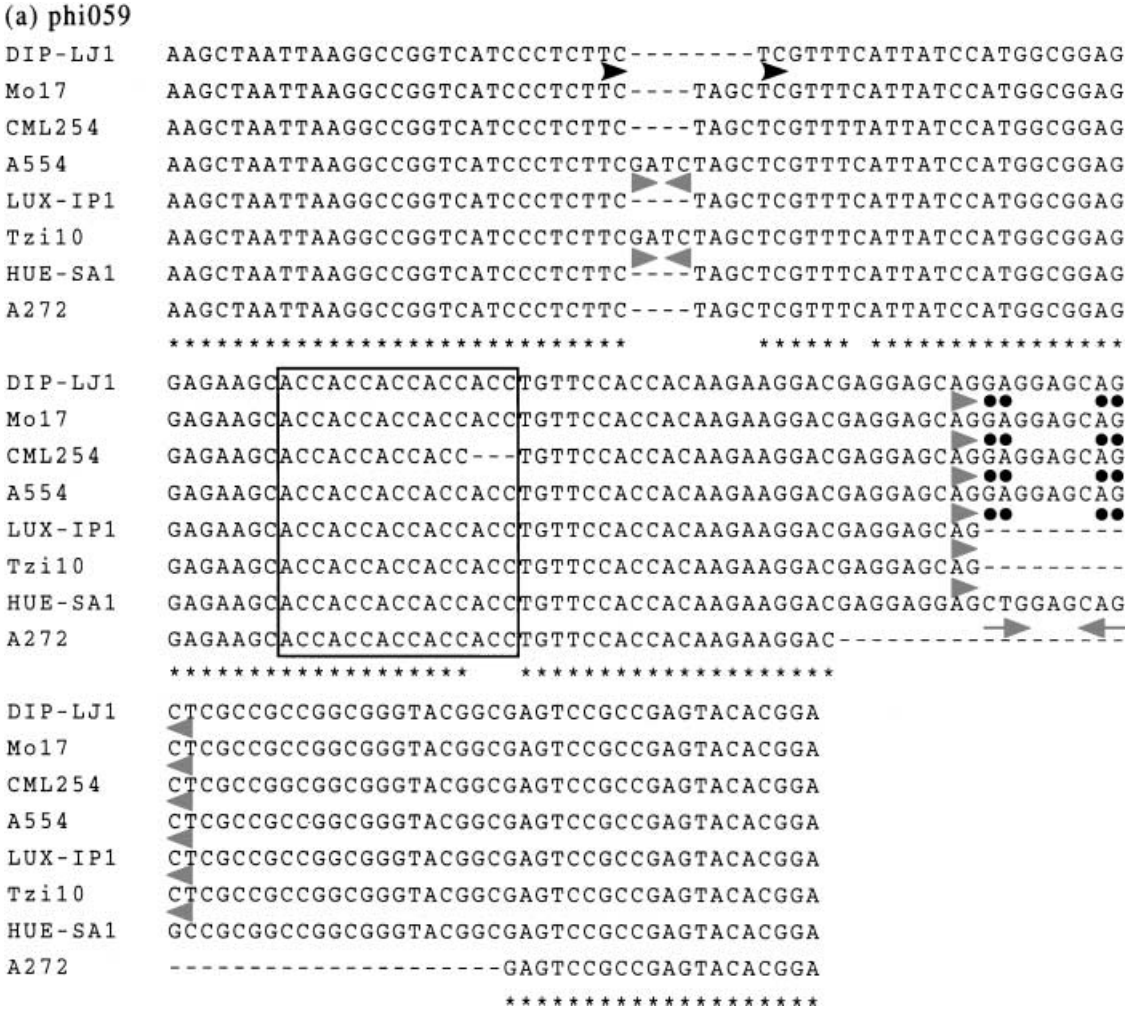


Fig. 2 Allele sequences of four microsatellites: (**a**) phi059, (**b**) phi96100, (**c**) phi083 and (**d**) phi102228. The stretches of microsatellite repeats are *boxed*. A microsatellite-like stretch found for phi083 is *boxed with a dashed line*. *Solid arrows, gray arrows* and *dots* indicate direct repeats, inverted repeats, and palindromes that are associated with indels in the regions flanking the microsatellite repeats. Note that some of the direct and inverted repeats are imperfect. *Asterisks* and *hyphens* show invariant sites and alignment gaps, respectively

sizes of fragment analysis were always larger than those of sequencing by 1 to 3 base pairs with two exceptions. These results suggest that allele-size determination of fragment analysis could be affected by the 3´ terminal extension activity of *Taq* polymerase, as *Taq* polymerase may produce one or a few nucleotide overhangs depending on the conditions (Holton and Graham 1991). Size differences larger than 1 bp were observed only in phi96100, suggesting that its primers might induce large overhangs. The allele size was smaller in fragment analysis than in sequencing for an unknown reason in two cases (phi059, A272 and phi083, F2834T).

Of the 59 microsatellite loci originally tested, 32 were used to test the reproducibility in genotyping. For each of these 32 loci, 4 to 32 plants (average 18.4) were re-analyzed and the results compared to the original or first-run data. This exercise demonstrated that the reproducibility of allele size estimates is significantly impaired when the allele sizes were over 320 bp. The percentage of perfect match or 1-bp difference between the original and re-run data was only 52% when the allele sizes over 320 bp were included, while it was 98.2% when the allele sizes were 320 bp or smaller. This may indicate that large DNA fragments (>320 bp) are not resolved well on gels with our conditions. For this reason, loci that had large allele sizes (>320 bp) were removed from our evolutionary genetic analyses (see below).

Reproducibility was also measured by comparing the original genotype for a plant-microsatellite combination to the genotype for that combination upon reanalysis. For the inbreds, allele size for 151 of 161 (94%) reruns were within 1 bp of the original scoring and five datapoints changed by 1 or 2 bp. These changes are likely all due to variance in the estimate of fragment sizes between runs. Two of 161 reruns were homozygous in one run but heterozygous in the other. This may be either PCR contamination, inconsistency in the amplification of heterozygotes (see below) or spillover between lanes. Three

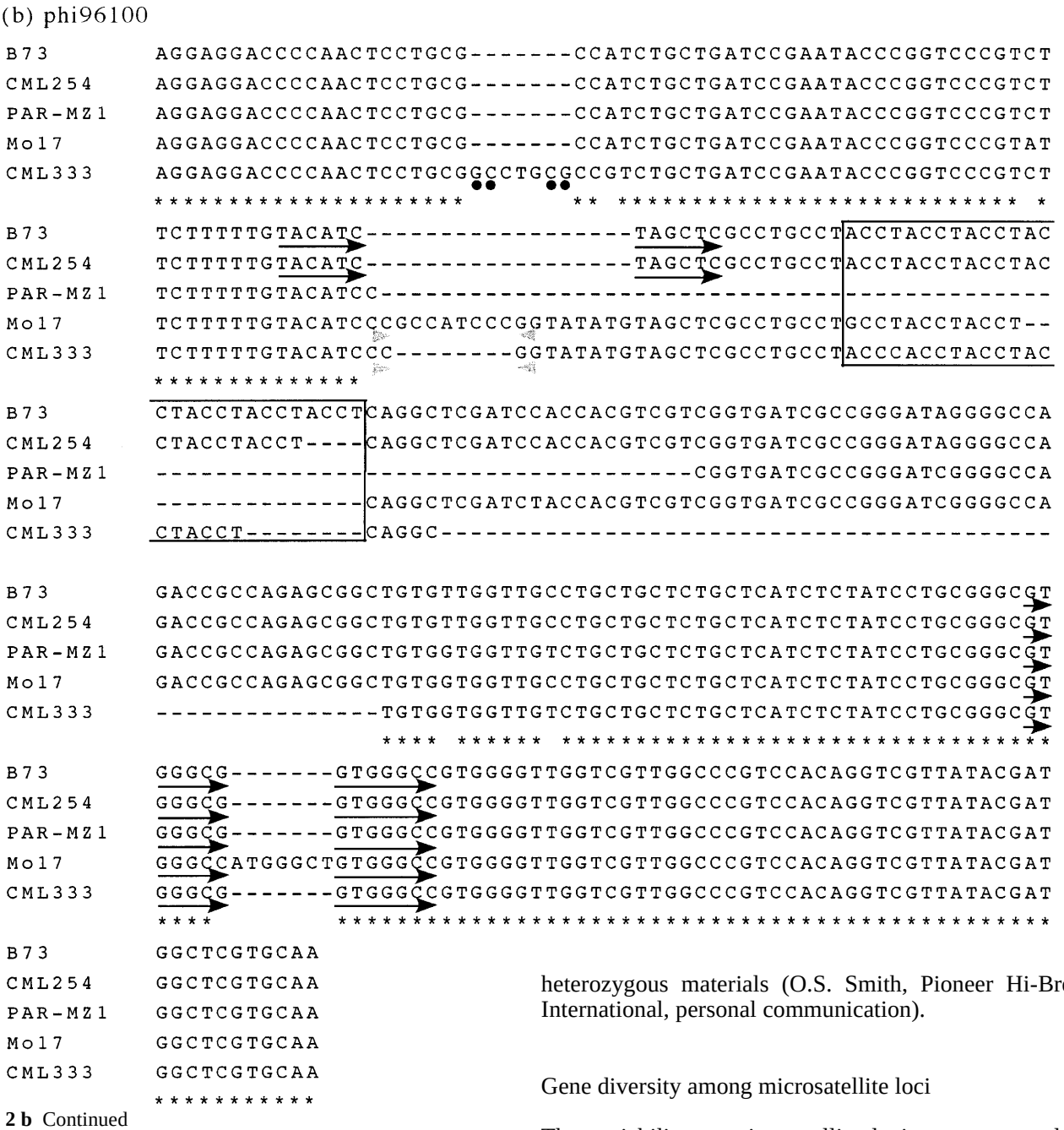


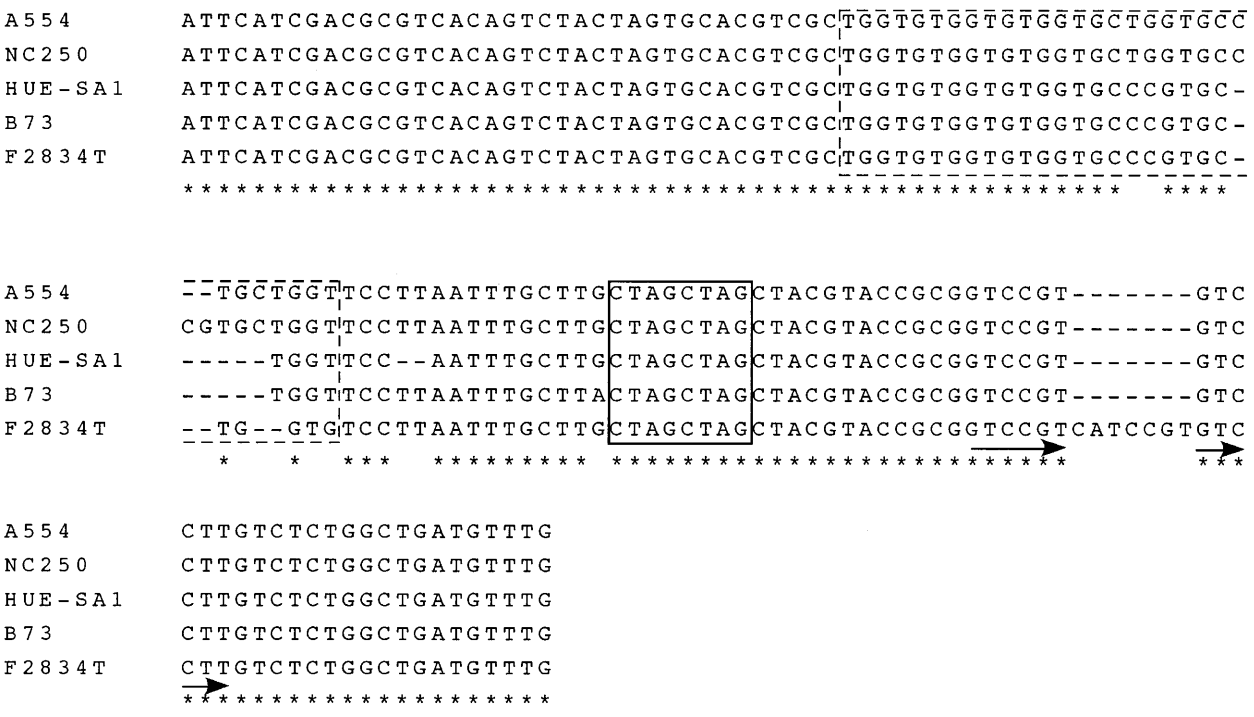
Fig. 2 b Continued

comparisons showed alleles differing by 9 bp or more between runs, which may be due to loading errors. For the open-pollinated materials, allele size for 343 of 350 (98%) reruns were within 1 bp in size of the original scoring and seven changed by 1 or 2 bp. There were an additional ten plants that were heterozygous in one run but homozygous in the other. We suspect that the amplification of both alleles in heterozygous plants is not faithful, especially considering that the primer for several microsatellite loci were multiplexed and the match between primer and template is likely not perfect with teosinte samples in many cases. Another laboratory has also observed inconsistent amplification of both alleles in heterozygous materials (O.S. Smith, Pioneer Hi-Bred, International, personal communication).

Gene diversity among microsatellite loci

The variability at microsatellite loci was measured in terms of the numbers of alleles and the gene diversity (Nei 1973) for teosinte and maize landraces and inbreds (Table 2) . In teosinte and maize landraces, the average number of alleles over 42 loci was 12.6 (range 4–24) and the average gene diversity was 0.73 (range 0.22–0.92). In inbreds, the average number of alleles over 42 loci was 6.9 (range 2–14), and the average of gene diversity was 0.62 (range 0.18–0.89). These values observed for inbreds were consistent with the results of Senior et al. (1998), where the average numbers of alleles was 5 (range 2–23) and the average gene diversity was 0.59 (range 0.17–0.92), and Smith et al. (1997), where the average gene diversity was 0.62. The average of gene diversity was also similar to that obtained by RFLP analy-

(c) phi083



(d) phi102228

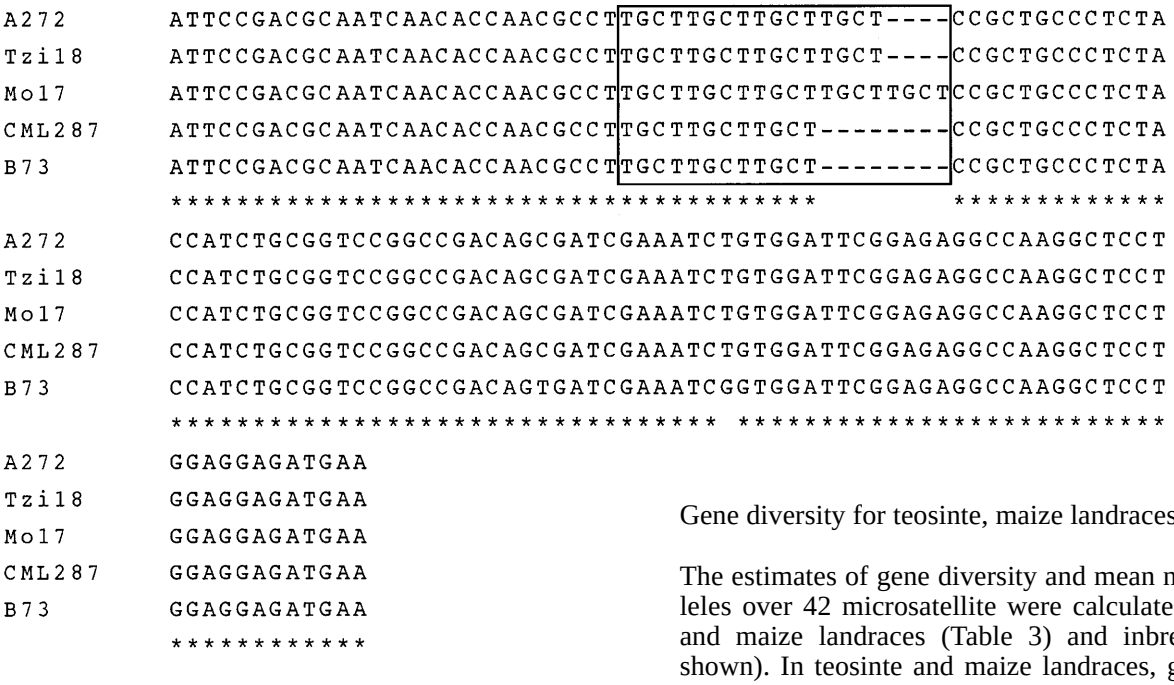


Fig. 2 c, d Continued

sis (Smith et al. 1997). An analysis of variance indicated that there was no association between gene diversity and chromosomal location of the microsatellites in teosinte and maize landraces (F=0.51; df=9, 31; P>0.80), and inbreds (F=0.24; df=9, 32; P>0.90).

Gene diversity for teosinte, maize landraces, and inbreds

The estimates of gene diversity and mean numbers of alleles over 42 microsatellite were calculated for teosinte and maize landraces (Table 3) and inbreds (data not shown). In teosinte and maize landraces, gene diversity varied among the 16 populations from 0.35 for *Z. luxurians* (LUX-IP1) to 0.68 for *Z. mays* ssp. *parviglumis* (PAR-MZ1). The populations of *Z. mays* and *Z. diploperennis* tended to show higher gene diversity (average 0.57) than those of *Z. luxurians* (0.35). This trend agrees with the results from isozyme studies (Doebley et al. 1984). Among the three wild subspecies of *Z. mays*, the populations *of Z. mays* ssp. *parviglumis* (0.65) and *Z. mays* ssp. *mexicana* (0.58) tended to show higher gene

Table 3 Gene diversity and mean number of alleles for teosinte and maize landraces

Taxon	OTU	No. of Plants	Gene diversity	Mean number of alleles
Z. *diploperennis*	DIP-LJ1	6	0.49	2.69
Z. *luxurians*	LUX-IP1	6	0.35	2.19
Z. mays ssp. *huehuetenangensis*	HUE-SA1	6	0.48	2.93
Z. mays ssp. *mexicana*	MEX-NO1	6	0.49	3.00
	MEX-CZ1	6	0.55	3.31
	MEX-UR1	6	0.65	4.21
	MEX-LR1	6	0.59	3.95
	MEX-MF1	6	0.60	3.76
mexicana, pooled		30	0.67	7.69
Z. mays ssp. *parviglumis*	PAR-MZ1	6	0.68	4.26
	PAR-MZ2	6	0.61	3.57
	PAR-TL1	6	0.63	4.12
	PAR-TL2	6	0.67	4.26
parviglumis, pooled		24	0.69	7.81
Z. mays ssp. *mays*	MAY-CN1	6	0.57	3.43
	MAY-CQ1	6	0.45	2.62
	MAY-CA1	6	0.51	2.95
	MAY-CO1	6	0.49	2.90
mays, pooled		24	0.61	5.60

Table 4 Numbers and percentages of specific alleles in *Zea*

Taxon or region	No. of populations or lines	No. of plants	Total No. of alleles	No. of specific alleles	Percentage of specific alleles
I. Teosinte and maize landraces					
Z. diploperennis	1	6	113	29	25.7
Z. luxurians	1	6	92	20	21.7
Z. mays					
ssp. *huehuetenangensis*	1	6	123	23	18.7
ssp. *mexicana*	5	30	323	68	21.1
ssp. *parviglumis*	4	24	328	78	23.8
ssp. *mays*	4	24	235	18	7.7
II. Inbred lines					
Tropic	41	41	231	34	14.7
USA	54	54	219	20	9.1
Canada/Europe	7	7	116	9	7.8

diversity than *Z. mays* ssp. *huehuetenangensis* (0.48). The pooled gene diversity of *Z. mays* ssp. *mays* (maize landraces) (0.61) was significantly lower than that of its close wild relatives, *Z. mays* ssp. *parviglumis* (0.69) (Wilcoxon' s signed rank test, $P<0.01$) and *Z. mays* ssp. *mexicana* (0.67) ($P<0.05$). The difference between *Z. mays* ssp. *parviglumis* and *Z. mays spp. mexicana* was not significant ($P>0.05$). These results indicated that the microsatellite loci were less variable in our maize sample than in the *mexicana* and *parviglumis* samples. A similar pattern of diversity between maize and teosinte is seen for the mean numbers of alleles.

In inbreds, gene diversity was lower than 0.01 in most cases, indicative of their self-pollinating nature. Two inbreds (KUI21 and 38–11) showed rather high gene diversity (0.27 and 0.24, respectively), probably because of residual heterozygosity or contamination. Mean numbers of alleles per locus were smaller than 1.10 with four exceptions (KUI21, F2, CML91, 38–11). The pooled gene diversity (0.62) and mean number of alleles (6.88) over all inbreds were comparable to those of maize landraces (0.61 for the pooled gene diversity and 5.60 for the pooled mean number of alleles). Similar estimates of the pooled gene diversity were observed among the three germplasm sources for inbreds (0.59 for Tropical, 0.60 for USA, and 0.58 for Canada/Europe), while the estimated mean number of alleles was relatively small in the Canada/Europe (2.76) inbreds as compared to in the Tropical (5.55) and the USA (5.21) lines, probably because of a smaller sample size.

Taxon-specific alleles

Numbers and percentages of taxon-specific alleles are shown for teosinte and maize landraces (Table 4-I) and inbreds (Table 4-II). In teosinte and maize landraces, the percentage of specific alleles in maize landraces (7%)

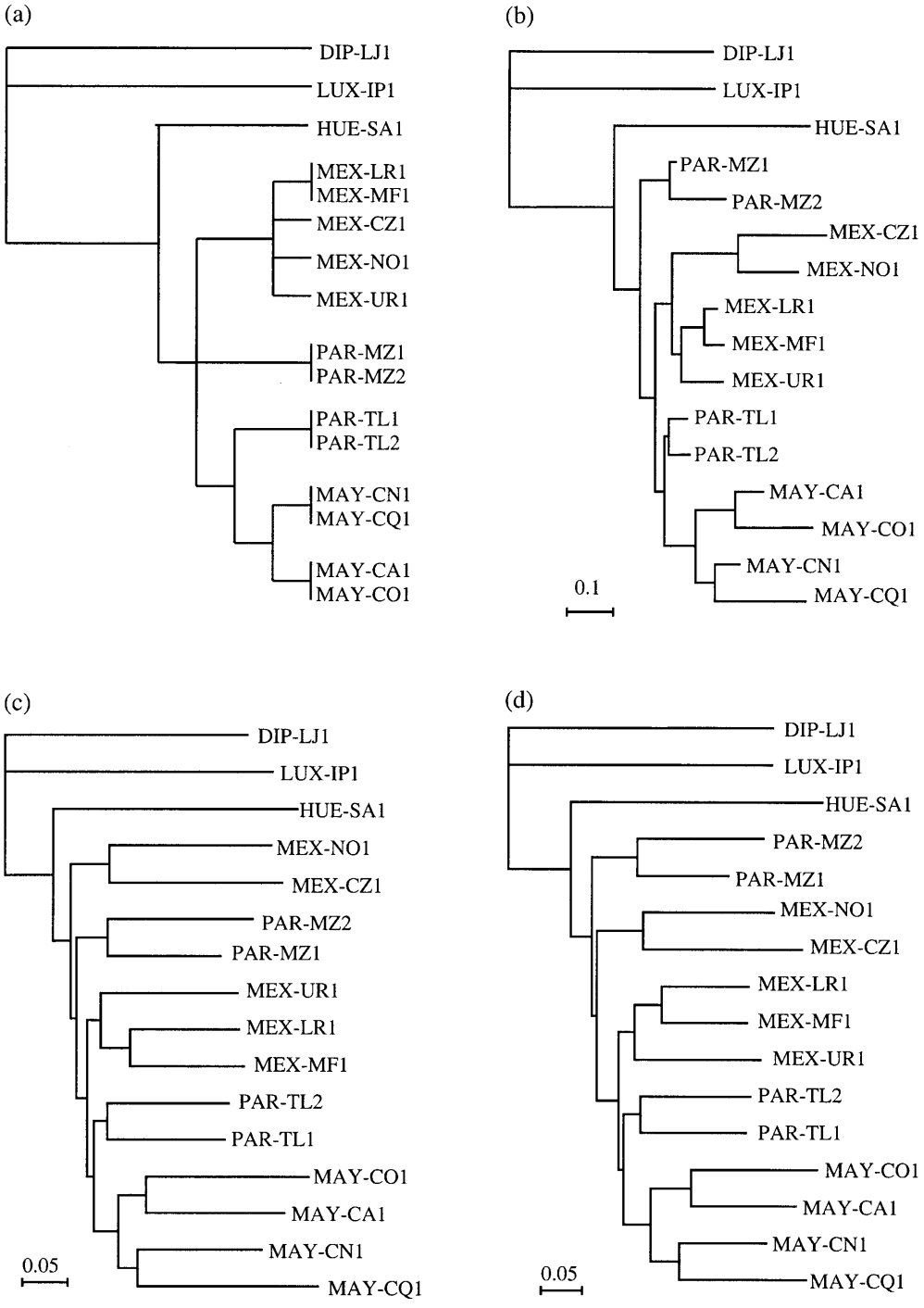


Fig. 3 Expected (**a**) and Fitch-Margoliash (**b, c**, and **d**) trees for 16 populations of teosinte and maize. The expected tree is based on the taxonomy of *Zea* (Doebley and Iltis 1980; Iltis and Doebley 1980), the geographic relationships of the populations (Table 1), and molecular evidence (Doebley 1990; Buckler and Holtsford 1996). The separation of the Mazatlan populations of *Z. mays* ssp. *parviglumis* from the Teloloapan populations is suggested by unpublished data (Ed Buckler, personal communication). The trees were constructed with Nei's standard distance (**b**), Rogers' distance (**c**), and Nei et al.'s chord distance (**d**). Trees with Cavalli-Sforza and Edwards's chord distances have the same topology as the Rogers' distance tree (data not shown)

was greatly reduced compared to that of wild taxa (25.7% for *Z. diploperennis*, 21.7% for *Z. luxurians*, 21.1% for *Z. mays* ssp. *mexicana*, and 23.8% for *Z. mays* ssp. *parviglumis*), indicating that most of the alleles in the maize landraces (>92%) are common to its wild relatives. This may reflect that maize was recently derived from teosinte (approximately 7,500 years). Tropical inbreds showed a higher percentage of specific alleles (14.7%) than USA inbreds (9.1%) and Canada/Europe inbreds (7.8%). The presence of many specific alleles in tropical inbreds was in agreement with the results of Senior et al. (1998). These specific alleles can be useful in maize breeding programs because they may be diagnostic for specific lineages or inbreds.

Performance of microsatellites for species and population divergence

To evaluate the performance of maize microsatellites for analyses of interspecific, subspecific and populational relationships, phylogenetic trees were constructed for te-

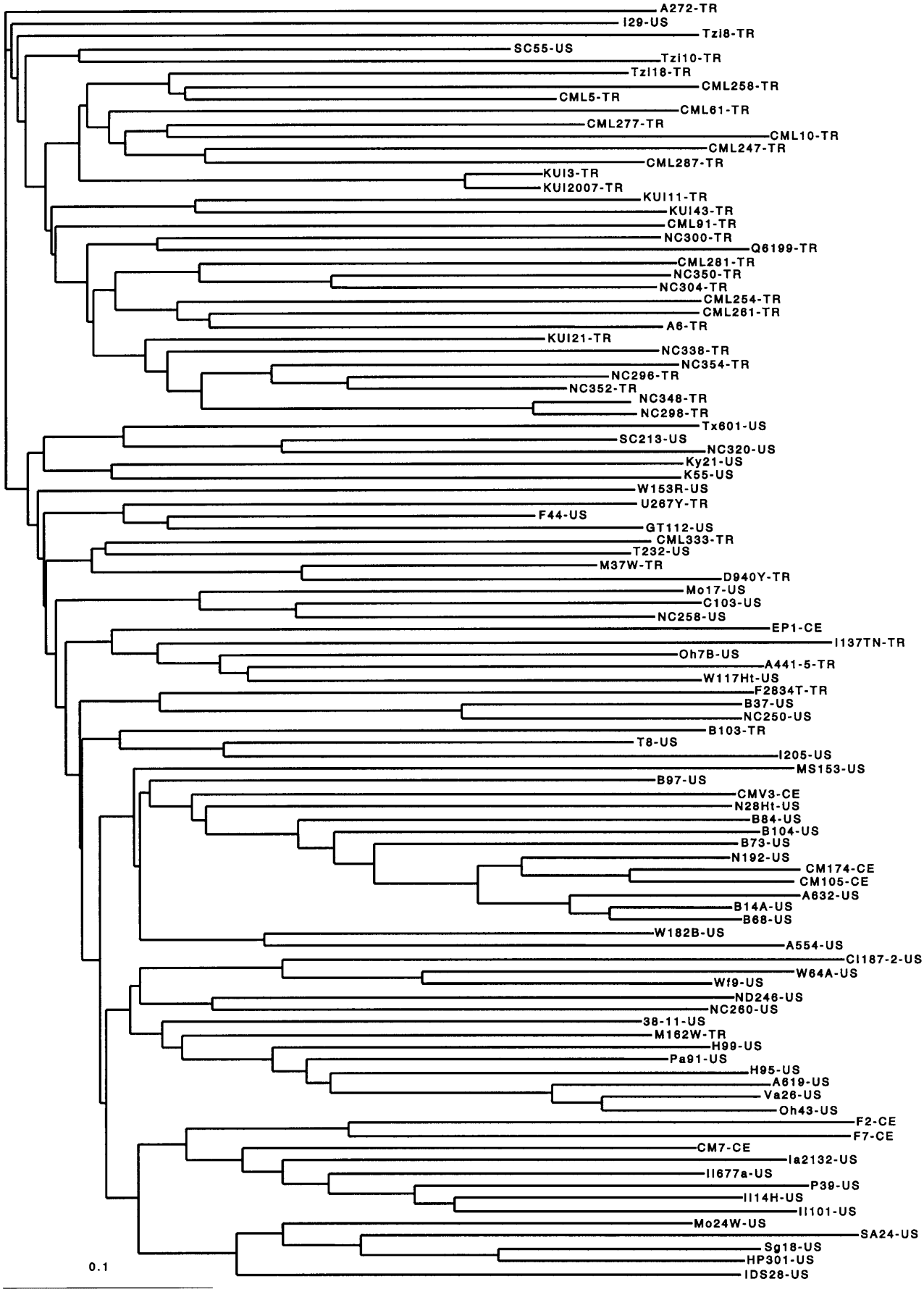


Fig. 4 Phylogenetic tree for 101 maize inbred lines based on Rogers' distance and the Fitch-Margoliash method of tree construction

osinte and maize landraces and inbreds. Because inter- and intra-specific relationships have already been proposed in *Zea* (Doebley 1990; Buckler and Holtsford 1996), we have a null hypothesis (Fig. 3a) against which the microsatellite-based tree can be compared. The 16 populations of teosinte and maize landraces were selected to represent different levels of divergence from separate species to populations of the same species separated by just a few kilometers. The utility of microsatellites for predicting very recently diverged (within 100 years) lines can be tested using the inbreds by comparing the microsatellite-based trees with known pedigrees (Smith et al. 1997).

Choosing an appropriate genetic-distance measure is critical in the analyses of inter- and intra-specific divergence. Because most of our maize microsatellite loci showed non-stepwise distribution patterns (see above), we employed four genetic distances that are free of the SMM assumptions: Nei's standard distance (Ds) (Nei 1972), Rogers' distance (Dr) (Rogers 1972), Cavalli-Sforza and Edwards's chord distance (Dc) (Cavalli-Sforza and Edwards 1967), and Nei et al.'s chord distance (Dn) (Nei et al. 1983). Ds is based on the IAM, whereas the latter three are Euclidean (or metric) distances that are based on multidimensional geometric models. These four distances were calculated respectively for teosinte and maize landraces, and inbreds, and trees were constructed using the Fitch-Margoliash algorithm (Fitch and Margoliash 1967).

The tree based on Ds (Fig. 3b) showed perfect agreement with the expected tree (Fig. 3a). The two distinct species, *Z. luxurians* and *Z. diploperennis*, are basal, ssp. *huehuetenangensis* is basal to other *Z. mays* subspecies, and the populations of *Z. mays* ssp. *mexicana* are monophyletic. The trees based on the Euclidean distances (Dr, Dc and Dn) are fairly similar, but they all failed to show a monophyletic cluster for *Z. mays* ssp. *mexicana* (Fig. 3c and d). At the population level, the trees based on all four distances performed well in that geographically close populations of the same subspecies clustered together: PAR-TL1 and PAR-TL2 from near the town of Teloloapan, PAR-MZ1 and PAR-MZ2 from near the town of Mazatlan, MEX-LR1 and MEX-MF1 from the Valley of Mexico, MAY-CO1 and MAY-CA1 from South America, and MAY-CN1 and MAY-CQ1 from central Mexico. Finally, all trees showed the close relationship between the Teloloapan populations of *Z. mays* ssp. *parviglumis* and maize landraces, supporting the observation that *Z. mays* ssp. *parviglumis* is the progenitor of maize (Doebley et al. 1984; Wang et al. 1999).

Phylogenetic analysis of microsatellite data for inbreds revealed many groupings that are consistent with the known pedigrees of these lines (Fig. 4). There is a general separation of the tropical and temperate lines. Within the temperate lines, several anticipated subgroups appear. A619, Oh43 and Va26, from the Lancaster heterotic group, represent one of them. These three lines were commercially important in the 1970 s; A619 and Va26 (and H95) were largely derived from Oh43. W64A was derived from Wf9 and CI187–2. C103, Mo17 and NC258 represent a Lancaster cluster. All three lines have been important commercially. B37 and NC250 are closely related; NC250 was derived from a backcross to B37. The remaining 25% of NC250 is unrelated tropical germplasm. A group of lines derived from the Iowa State Stiff Stalk Synthetic encompasses A632, B14A, B68, CM105, CM174, N192 (all derived in large part from B14), B73, B104, and B84. These form the largest and most-widely used group of lines. All of the sweetcorns (Il101, Il14H, P39, Il677a and Ia2132) grouped together, as did all of the popcorns (HP301, Sg18, SA24 and IDS28). In addition, the pair of widely used French lines F2 and F7, derived from the same source population, group together.

Within the tropical lines, NC298, NC348, NC296, NC352, NC354 and NC336 represent a temperate-adapted, all-tropical breeding group, the first four of which represent two sets of near-sister lines [(NC298, NC348); (NC296), (NC352)]. NC304 and NC350 represent two near-sister lines of all-tropical derivation. KUI3 and KUI2007 represent two Suwan-1 lines from Thailand. The closest grouping which appears to be of unrelated materials is that of SC213 (our source of this line may not be representative) and NC320. SC213 is most closely related by pedigree to GT112, while NC320 is closely related to SC76 (not included here), which was derived from a cross of two open-pollinated varieties (Hastings Prolific and Yellow Tuxpan).

Discussion

Application of maize microsatellites to other *Zea* species

Ascertainment bias is a concern when applying microsatellites developed for one species (the focal species) to other closely related (non-focal) species (Hutter et al. 1998). Previous studies have shown that microsatellites tend to be more variable in the focal species than in non-focal species to which they are applied (Ellegren et al. 1995). This is due to an ascertainment bias that is caused by non-random cloning and selection for variable microsatellites in the original species (Ellegren et al. 1995; Forbes et al. 1995; Hutter et al. 1998). Because our microsatellites are all derived from maize and selected non-randomly from genomic libraries (Taramino and Tingey 1996) and databases (Chin et al. 1996), their variability could have been affected by ascertainment bias when they were applied to teosinte. While we can not conclude that ascertainment bias has not affected our estimates of gene diversity, the relative levels of genetic diversity within and among species observed with microsatellites and previous isozymes analyses are consistent. For example, while there is lower gene diversity and fewer alleles in *Z. luxurians* relative to maize (Table 3), this difference parallels the reduced variability of isozymes in *Z. luxurians* (Doebley et al. 1984), and supports the idea that *Z. luxurians* is less variable than other species in *Zea*. Furthermore, *Z. diploperennis* showed gene diversi-

ty and a mean number of alleles that are comparable to those of the populations of *Z. mays* (Table 3), which also agrees with the results from isozyme studies (Doebley et al. 1984).

Zea is well-known to have a polyploid ancestry (Helentjaris et al. 1988). Polyploidy gives rise to duplicated loci that may, or may not, amplify depending on how well primer binding sites are conserved. This problem would make the data interpretation complex. In this study, we observed amplification of more than two products in a plant at relatively low frequencies (1.8% for teosinte and maize landraces, and 0.02% for inbreds). These errors may represent the amplification of duplicate loci, contamination of the PCR, or some of error. The low frequency of this problem indicates that maize microsatellite primers are specific enough to routinely amplify only a single locus despite the duplicate nature of many maize loci. The presence of such extra "alleles" for a single microsatellite has also been observed in a fully diploid organism (Goldstein and Clark 1995). By treating plant-microsatellite combinations with more than two peaks as missing data, data analyses and results should not be biased.

Complex and frequent mutations in *Zea* microsatellites

It has been shown that mutation patterns at microsatellites are often complicated and thus use of the SMM may not be justified. For example, the distribution of mutations in the microsatellite repeats was suggested to be asymmetric: some tend to gain repeats (Weber and Wong 1993; Eichler et al. 1994), while others tend to lose repeats (Zhang et al 1994). Other observations suggest that some form of allele size-constraint prevents microsatellites from increasing to a large number of repeats (Zhivotovsky et al. 1997), and that interruptions of stretches of microsatellite repeats tend to stabilize microsatellite loci (Ashley and Warren 1995; Goldstein and Clark 1995; Kruglyak et al. 1998).

In addition to these complexities, mutations in the region flanking microsatellite repeats can make the evolution of microsatellite loci even more complex. Several authors reported interspecific and intraspecific size variation at microsatellite loci that was caused by indels in the flanking regions (Angers and Bernatchez 1997; Grimaldi and Crouas-Roy 1997; Buteler et al. 1999; Colson and Goldstein 1999). In *Drosophila*, about 60% of microsatellite loci examined showed divergence in the lengths of flanking regions between species (Hutter et al. 1998; Colson and Goldstein 1999).

In this study, we have shown that 40 out of 46 maize microsatellite loci have allele size distributions that do not fit the simple model of allelic variation based solely on changes in the number of repeat units. Sequencing of 32 alleles from six microsatellite loci revealed that indels are frequent in the region flanking the microsatellite repeats (Fig. 2). If this phenomenon is common to the 40 non-stepwise microsatellite loci, 87% of our loci are affected by indels with respect to their allele-size variation. Maize microsatellites were selected to show variability in size on agarose gels (Chin et al. 1996; Taramino and Tingey 1996). This may have biased the selection process to favor microsatellite loci with large differences between the sizes of alleles as caused by indels in the flanking regions. It may be more appropriate to refer to these markers as "IRRs" (Indel-Rich Regions) instead of microsatellites or SSRs since our data indicate that most of the variation is due to indels in the regions flanking the repeat motifs.

Among the 31 sequenced alleles, there are at least 38 indels in the flanking region, and 23 of them (60%) were associated with some types of characteristic sequences such as direct repeats, inverted repeats, or palindromes (Fig. 2). These characteristic sequences might have increased the likelihood of indel mutations through replication error.

Performance of maize microsatellites for evolutionary studies in *Zea*

In spite of the complex mutation patterns, maize microsatellite loci are useful for the analysis of inter- and intra-specific relationships in *Zea*. A great deal of variation was observed among and within populations in terms of numbers of alleles and gene diversity (Table 3), indicating that microsatellites provide an excellent source of intraspecific variation in this genus. The pattern of variation observed among the 16 populations of *Zea* agrees reasonably well with the results from isozyme studies (Doebley et al. 1984). Gene diversity and the mean number of alleles for inbreds closely parallels those values from previous studies (Smith et al. 1997; Senior et al. 1998). These facts indicate that maize microsatellite loci provide reliable estimates on intraspecific DNA variation. Furthermore, our data show that genetic variation is reduced in maize as compared to that of its close wild relatives, *Z. mays* ssp. *mexicana and Z. mays* ssp. *parviglumis* (Table 3). This probably reflects the "domestication bottle neck" through which maize once passed; thus, maize microsatellite loci are evolving slowly enough so that the evidence for this historical event has not been erased (Eyre-Walker et al. 1998).

Maize microsatellite loci can also be useful to clarify the relationships between inbreds, populations, and species in *Zea*. Because complex mutations in the flanking regions affect the size variation of maize microsatellites, the SMM is probably not appropriate for our data. In teosinte and maize landraces, the Fitch-Margoliash tree with Nei's standard distance (based on the IAM) reconstructed correctly both the taxonomic relationships of the species and the geographical relationships of the populations (Fig. 3). The microsatellite-based trees indicate that the Teloloapan populations of *Z. mays* ssp. *parviglumis* (PAR-TL1 and PAR-TL2) are placed basal to *Z. mays* ssp. *mays* in all Fitch-Margoliash trees, a result that is consistent with the view that maize originated from *Z.*

mays ssp. *parviglumis* (Doebley et al. 1984; Wang et al. 1999). Also the inbred tree is largely consistent with the known pedigrees of these lines. These results indicate that, despite their complex mutation patterns, microsatellite loci may have advantages in phylogeographical studies in *Zea*, especially considering that the number of available microsatellites for maize is near 2,000 (http://www.agron.missouri.edu/ssr.html).

In summary, this study establishes an automated format of multiplexed maize microsatellites as powerful tools for evolutionary studies in *Zea*. Maize microsatellites were applicable to other *Zea* species with no practical difficulties. The fact that up to ten microsatellite loci were assayed per lane on an ABI377 demonstrates the feasibility of assaying a large number of microsatellite loci on a comprehensive sample of *Zea* germplasm. This should allow sufficient data that one could generate phylogenetic and diversity data for individual chromosomes or chromosomal segments, and thus address questions about whether different regions of the genome have experienced different histories. In spite of their complex mutation patterns involving indels in the regions flanking the microsatellite repeats, the information obtained was useful to elucidate the intraspecific variation and to clarify the interspecific relationships. Understanding the mutational process at microsatellite loci will allow the development of appropriate statistical models that will improve their usefulness in evolutionary studies in *Zea* as well as in other organisms. The development of genetic-distance measures that combine features of the SMM and IAM may be desirable for maize microsatellite loci.

Acknowledgments This research was supported by NSF grant DBI-0096033. We thank Kejun Liu for making his computer software package Ldist available to us.

References

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

Ashley Jr CT, Warren ST (1995) Trinucleotide repeat expansion and human disease. Annu Rev Genet 29:703–728

Buckler IV ES, Holtsford TP (1996) *Zea* systematics: ribosomal ITS evidence. Mol Biol Evol 13:612–622

Bruford MW, Wayne RK (1993) Microsatellites and their application to population genetic studies. Curr Opin Genet Dev 3:939–943

Buteler MI, Jarret RL, LaBonte DR (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. Theor Appl Genet 99:123–132

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet 19:233–257

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

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

Condit R, Hubbell SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66–71

Doebley JF (1990) Molecular systematics of *Zea* (Gramineae). Maydica 35:143–150

Doebley JF, Iltis HH (1980) Taxonomy of *Zea* (Gramineae). I. A subgeneric classification with a key to taxa. Am J Bot 67:982–993

Doebley JF, Stec A (1991) Genetic analysis of the morphological differences between maize and teosinte. Genetics 129:285–295

Doebley JF, Goodman MM, Stuber CW (1984) Isoenzymatic variation in *Zea* (Gramineae). Systematic Bot 9:203–218

Eichler EE, Holden JJ, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994) Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. Nature Genet 8:88–94

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? Nat Genet 11:360–362

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA 95:4441–4446

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. Mol Biol Evol 12:1106–1113

Goldstein DB, Clark AG (1995) Microsatellite variation in north american populations of *Drosophila melanogaster*. Nucleic Acids Res 23:3882–3886

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J Hered 88:335–342

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Goldstein DB, Roemer GW, Smith DA, Reich DE, Bergman A, Wayne RK (1999) The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics 151:797–801

Grimaldi MC, Crouau-Roy B (1997) Microsatellite allelic homoplasy due to variable flanking sequences. J Mol Evol 44:336–340

Helentjaris T, Weber D, Wright S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragment length polymorphisms. Genetics 118:353–363

Holton TA, Graham MW (1991) A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. Nucleic Acids Res 19:1156

Hutter CM, Schug MD, Aquadro CF (1998) Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Mol Biol Evol 15:1620–1636

Iltis HH, Doebley JF (1980) Taxonomy of *Zea* (Gramineae). II. Subspecific categories in the *Zea mays* complex and a generic synopsis. Am J Bot 67:994–1004

Innan H, Terauchi R, Miyashita NT (1997) Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 1997:1441–1452

Kimura M, Crow JF (1964) The numbers of alleles that can be maintained in a finite population. Genetics 49:725–738

Kresovich S, Szewc-McFadden AK, Bliek SM, McFerson JR (1995) Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor Appl Genet 91:206–211

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Liu ZW, Biyashev RM, Saghai-Maroof MA (1996) Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor Appl Genet 93:869–876

Lübberstedt T, Dussle C, Melchinger AE (1998) Application of microsatellites from maize to teosinte and other relatives of maize. Plant Breed 117:447–450

MacHugh DE, Shriver MD, Loftus RT, Cunningham P, Bradley DG (1997) Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146:1071–1086

Mitchell SE, Kresovich S, Jester CA, Javier Hernandez C, Szewc-McFadden AK (1997) Application of multiplex PCR and fluorescence-based, semi-automated allele sizing technology for genotyping plant genetics resources. Crop Sci 37:617–624

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153–170

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet Res 22:201–204

Rogers JS (1972) Measures of genetic similarity and genetic distance. Studies in Genetics VII. University of Texas Publication 7213, Austin, Texas, pp 145–153

Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganal MW (1995) Abundance, variability and chromosomal location of microsatellites in wheat. Mol Gen Genet 246:327–333

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P, Ganal MW (1998) A microsatellite map of wheat. Genetics 149:2007–2023

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TFC, Aquadro CF (1998) The mutation rates of di-, tri- and tetra-nucleotide repeats in *Drosophila melanogaster*. Mol Biol Evol 15:1751–1760

Senior ML, Heun M (1993) Mapping maize microsatellites and polymerase chain reaction confirmation of the targeted repeats using a CT primer. Genome 36:884–889

Senior ML, Murphy JP, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci 38:1088–1098

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. Theor Appl Genet 95:163–173

Taramino G, Tingey S (1996) Simple sequence repeats for germplasm analysis and mapping in maize. Genome 39:277–287

Tautz D, Renz M (1984) Simple sequences are ubiquitous components of eukaryotic genomes. Nucleic Acids Res 12:4127–4138

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

Wang RL, Stec A, Hey J, Lukens L, Doebley J (1999) The limits of selection during maize domestication. Nature 398:236–239

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Zhang L, Leeflang EP, Yu J, Arnheim N (1994) Studying human mutations by sperm typing: instability of CAG trinucleotide repeats in the human androgen receptor gene. Nature Genet 7:531–535

Zhivotovsky LA, Feldman MW, Grishechkin SA (1997) Biased mutations and microsatellite variation. Mol Biol Evol 14:926–933