

M. L. C. George · E. Regalado · W. Li · M. Cao ·  
M. Dahlan · M. Pabendon · M. L. Warburton ·  
X. Xianchun · D. Hoisington

## Molecular characterization of Asian maize inbred lines by multiple laboratories

Received: 20 October 2003 / Accepted: 3 February 2004 / Published online: 17 April 2004  
© Springer-Verlag 2004

**Abstract** This study focuses on the standardization of techniques across laboratories to enable multiple datasets to be compared and combined in order to obtain reliable and robust wide-scale patterns of diversity. A set of protocols using a core collection of simple sequence repeat (SSR) markers, reference lines and standard alleles, plus a common system of allele nomenclature, was adopted in the study of maize genetic diversity in a network of laboratories in Asia. Pair-wise allele comparisons of the reference lines, done to assess the general agreement between datasets from four laboratories, showed error rates (raw) ranging from 5.8% to 9.7%, which were reduced to less than 8% after adjustments of correctable errors, and further reduced to less than 6% after the exclusion of all markers with greater than 10% individual error rates. Overall, 45% of the total mismatches were due to frameshift errors, 39% to wrong allele size, 15% to failed amplification and 1% to “extra” alleles. Higher genetic

similarity values of the reference lines were achieved using fewer markers with data of higher quality rather than with more markers of questionable quality. Cluster analysis of the merged datasets showed the lines from southern China to be highly diverse, falling into six of the seven clusters observed and all well represented by tester lines. The lines from Indonesia fell into five of six groups, with two main groups represented by tester lines. The CIMMYT lines developed for the Asian region showed a relatively narrow genetic base, falling in two out of seven and in three out of six clusters in China and Indonesia, respectively. In contrast to the case in southern China where 95% of the lines clustered separately from the CIMMYT lines, lines in the Indonesian breeding program show a closer relationship with the CIMMYT lines, reflecting a long history of germplasm exchange.

Communicated by J.W. Snape

M. L. C. George (✉) · E. Regalado  
CIMMYT-Asian Maize Biotechnology Network,  
c/o IIRI, DAPO 7777, MetroManila, Philippines  
e-mail: m.george@cgiar.org  
Tel.: +63-2-5805600  
Fax: +63-2-5805699

W. Li · M. Cao  
Maize Research Institute,  
Sichuan Agricultural University,  
Xinkang Rd. 36, 625014 Ya’an, Sichuan, China

M. Dahlan · M. Pabendon  
Indonesia Cereal Research Institute,  
Agency for Agricultural Research and Development,  
Jl Ratulangi 274, 90514 Maros, South Sulawesi, Indonesia

X. Xianchun  
Institute of Crop Breeding and Cultivation,  
Chinese Academy of Agricultural Sciences,  
Zhongguancun South Street No.12, 100081 Beijing, China

M. L. Warburton · D. Hoisington  
CIMMYT (International Maize and Wheat Improvement Center),  
Apdo. Postal 6-641, 06600 Mexico D. F., Mexico

### Introduction

Information on the relationship between breeding materials is important for the efficient choice of parents in plant breeding programs. This is especially true in hybrid maize breeding, where the recognition and exploitation of heterotic patterns are vital for maximizing heterosis. Because crosses between genetically divergent lines generally produce better hybrids than crosses between closely related parents, success in hybrid breeding relies on clearly defined heterotic groups. When there are no established heterotic groups available in a large collection of germplasm, groups of genetically similar lines can first be identified based on DNA marker-based estimates of genetic distance. An optimum exploitation of heterosis can then be carried out by producing and evaluating diallel crosses among representative genotypes in each group (Melchinger 1999).

Molecular markers are able to characterize lines at the DNA level. With a sufficient number of markers, a clear structure of heterotic groups can be identified in a systematic way, and inbreds of unknown genetic origin can be assigned to established or new heterotic groups. Sev-

eral studies have successfully used RFLPs (restriction fragment length polymorphism) and SSRs (simple sequence repeats or microsatellites) to study maize genetic diversity and define heterotic groups in temperate maize germplasm (Messmer et al. 1992, 1993; Dubreuil et al. 1996; Lu and Bernardo 2001). However, tropical maize germplasm has not been as fully classified into heterotic groups. A study of tropical inbred lines and the populations from which these lines were developed revealed a large amount of diversity that made it difficult to find a clear-cut structure of the inbred lines (Warburton et al. 2002a). A similar situation was also observed in a regional diversity study of representative inbred lines from several Asian countries (George et al. 2004). However, in a small study involving tropical maize populations, Reif et al. (2003) did suggest some heterotic groupings as determined by SSR markers.

In Asia, the area planted in maize hybrids is expanding rapidly. In 1997, of the 19.6 million ha planted in tropical maize in seven Asian countries (China, India, Indonesia, Nepal, Philippines, Thailand, and Vietnam), 45% was planted with hybrids (Gerpacio 2001). Contributing to this advance of hybrids in the region is an international maize breeding system that includes the International Maize and Wheat Improvement Center (CIMMYT), national public breeding programs, and national and multinational private seed companies that engage in germplasm exchanges and collaborative varietal testing. As a focal point of maize germplasm improvement and exchange, CIMMYT has been active in producing improved materials and in promoting their dissemination. Since the 1980s, CIMMYT has been developing and releasing maize inbred lines (CIMMYT Maize Lines, or CMLs) that have played an important role in hybrid maize production in developing countries (CIMMYT 1998). The CMLs, derived from maize populations and pools that consisted of a wide range of tropical, subtropical, mid-altitude and highland materials mainly originating from Mexico, the Caribbean, Central and South America, India, Thailand, the Philippines and U.S. corn belt, have provided a very wide genetic and phenotypic base for breeders worldwide.

While pedigree information is useful for the assessment of the genetic relationships of breeding materials, this information is sometimes incomplete, unreliable, or unavailable. Furthermore, for maize breeding programs to be able to utilize new sources of germplasm effectively, it would be useful to have a knowledge of how the new inbred lines are related to the local lines for planning crosses for hybrid development. As a collaborative activity under the Asian Maize Biotechnology Network (<http://www.cimmyt.org/ambionet>), genetic diversity studies aimed at characterizing maize inbred lines have been conducted in several countries in Asia (Yuan et al. 2001; Pushpavalli et al. 2002; George et al. 2004). However, in the past, diversity studies involving the use of molecular markers that have been conducted in one laboratory have typically been self-contained experiments, and studies between different laboratories have rarely, if ever, been combined. This particular investi-

gation reported here focused on the wide-scale analysis of maize diversity to obtain a picture of diversity patterns within a country as well as to investigate the relationships between these lines and those in regional and international breeding programs. In this paper, we describe a methodology for the merging of molecular datasets from different laboratories as a coordinated network-wide activity, using the case of two countries (China and Indonesia) and CIMMYT as examples. An important aspect of this work is the standardization of the techniques to yield reproducible results across laboratories that would enable multiple datasets to be compared and combined into a common database.

The specific objectives of the investigation reported here were to (1) standardize protocols for maize fingerprinting for use in a network of laboratories; (2) combine datasets for wide-scale germplasm characterization; (3) analyze the genetic relationships between Asian national maize inbred line collections and CIMMYT lines developed for Asia.

## Materials and methods

### Standard fingerprinting methods

The protocols for fingerprinting with SSR markers, including DNA extraction, amplification, electrophoresis, and gel visualization, that are in use in the AMBIONET Service Laboratory (ASL) in the Philippines, as described by George et al. (2004), were adopted as the standard for maize genetic diversity work in the network. These protocols were used in China (Maize Research Institute of the Sichuan Agricultural University, SAU) and Indonesia (Research Institute for Agricultural Biotechnology and Genetic Resources, RIABGR) as well as in other countries in the network. SSR primers were selected from a set of markers (Table 1), almost all which possessed a repeat unit greater than two nucleotides and included primers in use at the Applied Biotechnology Center (ABC) in CIMMYT (Warburton et al. 2002a). Amplification with all primers was done at a single annealing temperature (56°C for 1 min) using a PTC-100 Programmable Thermal Controller (MJ Research, Waltham, Mass.) or, occasionally in China, a HB/TS/53C thermocycler (Hybaid, UK). The primers used in the ASL and Indonesia were purchased from Research Genetics (Huntsville, Ala.), while those used in China were synthesized at SBS Genetech (Beijing, China). The protocols used at the ABC in CIMMYT were as described by Warburton et al. (2002a) and involved PCR analyses with a specific annealing temperature for each primer and an automatic DNA sequencer for allele detection.

Six genotypes (CML51, -292, -202, -206, -236, and -396) were included as samples in every run. The reference lines, which were obtained as seeds or as DNA from CIMMYT and provided as DNA to China and Indonesia, were used to check the agreement of results between laboratories.

### Allele nomenclature

A common system of allele nomenclature was based on band position relative to the  $\phi$ X174/*Hinf*I marker (Promega, Madison, Wis.) fragments. An allele name consisted of the name of the SSR locus followed by an alphanumeric designation that indicates the size range and relative size of the allele within that particular size range: a letter to refer to the size range (*d* for the 312- to 427-bp size range; *e* for 250–311 bp; *f* for 201–249 bp; *g* for 152–200 bp; *h* for 141–151 bp; *i* for 119–140 bp; *j* for 101–118 bp; *k* for 83–100 bp; *l* for 67–82 bp; and *m* for <67 bp), followed by a number to

**Table 1** SSR markers used in the study, their bin location, repeat type, number of alleles detected in the AMBIONET Service Laboratory (ASL) and size range as determined automatically with an ABI377 DNA Sequencer. The markers are marked “x” in the da-

taset where they were used; those that were merged with ASL are marked with “xx”, and those that were eventually merged with CIMMYT are marked with “xxx”

Locus	Bin no.	Number of alleles	Size range (bp)	China	Indonesia		ASL	CIMMYT	
					1	2		1	2
phi056	1.01	4	85–94 <sup>d</sup>	x <sup>e</sup>			x	x <sup>e</sup>	
phi109275	1.03	6 <sup>1</sup>	117–143 <sup>d</sup>	xx	xx		xx		x <sup>g</sup>
phi339017	1.03	5	148–163				xx		xx
umc1169 <sup>a</sup>	1.04	-	-	x					x
umc1124	1.05	3	74–82 <sup>d</sup>				x		
umc1122 <sup>a</sup>	1.06	-	-	x	x				x
phi002	1.08	3	159–165 <sup>d</sup>	x			x		x <sup>e</sup>
phi011	1.09	4	111–129 <sup>d</sup>	x <sup>e</sup>			x	x <sup>e</sup>	
phi308707 <sup>b</sup>	1.10	6	116–134	xx			xx		x
phi064	1.11	12 <sup>c</sup>	73–109			xxx	xxx	xxx	
phi227562	1.11	7	307–328	xxx	xxx		xxx	xxx	
phi96100 <sup>b</sup>	2.01	8	269–297	x	xxx		xxx	xxx	
umc1555 <sup>a</sup>	2.03	9	79–119	x		x			x
phi109642 <sup>b</sup>	2.03–2.04	3	136–144	xxx		x	xxx		xxx
phi083 <sup>b</sup>	2.04	5 <sup>c</sup>	125–137	x	xxx		xxx	xxx	
nc133 <sup>b</sup>	2.05	3	105–115	x			x	x	
phi127	2.08	10 <sup>c</sup>	96–140	xxx			xxx	xxx	
phi101049 <sup>b</sup>	2.10	11	226–274	xx	xx		xx		x
phi453121 <sup>a</sup>	3.00	-	-	x					x
phi104127 <sup>b</sup>	3.01	3	157–169				x		x
phi374118 <sup>b</sup>	3.02	6	217–238	x	xx		xx		x
phi029	3.04	7 <sup>c</sup>	148–162	xxx			xxx		xxx
phi053	3.05	8 <sup>c</sup>	167–195	xxx	xxx		xxx	xxx	
phi073	3.05	4	91–100 <sup>d</sup>	x <sup>e</sup>			x		x <sup>e</sup>
phi102228	3.06	3	123–131	xxx		xxx	xxx		xxx
umc1399 <sup>b</sup>	3.07	6 <sup>c</sup>	111–127	xx			xx	x	
phi046	3.08	2	62–66		xx		xx		
phi047	3.09	4	140–151 <sup>d</sup>				x		
umc1136 <sup>b</sup>	3.10	9 <sup>c</sup>	132–159	x		x	xx	xx	
phi072	4.00–4.01	6	143–167	xxx	xxx		xxx	xxx	
phi213984	4.01	2	287–305	xxx		xxx	xxx		xxx
phi079 <sup>b</sup>	4.05	6 <sup>c</sup>	180–200	xx	xx		xx		x
phi093 <sup>b</sup>	4.08	6	274–306	xxx		x	xxx	xxx	
umc1109	4.10	4	104–116	xxx			xxx		xxx
phi006 <sup>a</sup>	4.11	6	76–94					x	
phi076 <sup>b</sup>	4.11	5	161–179	x	xx		xx		x
nc130	5.00	4 <sup>c</sup>	140–148	xxx			xxx		xxx
phi024	5.01	5	168–186 <sup>d</sup>	xx		x	xx		
phi008 <sup>a</sup>	5.03	3	100–106						x
phi109188 <sup>b</sup>	5.03	8 <sup>c</sup>	148–174	x	xxx		xxx		xxx
phi113	5.03–5.04	6	120–284 <sup>d</sup>				x		
phi331888	5.04	4	127–136 <sup>d</sup>	xxx		xxx	xxx		xxx
phi085 <sup>a</sup>	5.06	-	-	x					
phi087	5.06	5	150–177	xxx	xxx		xxx		xxx
bnlg118	5.07–5.08	7	105–121				x		
umc1153	5.09	6 <sup>c</sup>	105–114	xxx	xxx		xxx		xxx
umc1143 <sup>b</sup>	6.00	8 <sup>c</sup>	77–87	x	xx		xx	x	
bnlg391	6.01	9	77–104 <sup>d</sup>	xxx		xxx	xxx		xxx
phi423796	6.01	6 <sup>c</sup>	121–141	xxx		xxx	xxx		xxx
phi031 <sup>a</sup>	6.04	-	-	x					x
phi452693 <sup>a</sup>	6.04	7	125–145	x	x				x
phi078	6.05	9	124–210 <sup>d</sup>	xx			xx		
phi070 <sup>a</sup>	6.07	-	-						x
phi123	6.07	4	146–154 <sup>d</sup>	xxx			xxx	xxx	
phi299852 <sup>b</sup>	6.07	10	111–147	x	xxx		xxx		xxx
phi089	6.08	2	87–95			xxx	xxx		xxx
umc1545 <sup>b</sup>	7.00	6 <sup>c</sup>	70–86	x	xxx		xxx		xxx
phi112 <sup>b</sup>	7.01	7	135–159 <sup>d</sup>	x			x		x <sup>f</sup>
phi034	7.02	8	113–146	xxx	xxx		xxx	xxx	
phi114	7.03	4	137–169	xxx		xxx	xxx		xxx
phi328175	7.04	5 <sup>c</sup>	100–130	xxx	xxx		xxx	xxx	
phi069 <sup>a</sup>	7.05	-	-	x					x
phi116 <sup>b</sup>	7.06	5	152–173 <sup>d</sup>	x			x		x <sup>f</sup>

**Table 1** (continued)

Locus	Bin no.	Number of alleles	Size range (bp)	China	Indonesia		ASL	CIMMYT	
					1	2		1	2
phi420701 <sup>a</sup>	8.00	4	291–300	x	x				x
umc1304	8.02	5	125–141	xxx		xxx	xxx		xxx
phi121	8.03	2	97–100	xxx		xxx	xxx		xxx
phi100175 <sup>a</sup>	8.03	8	117–141	x					x
phi014	8.04	5	150–173 <sup>d</sup>	xx			xx		
umc1161 <sup>b</sup>	8.06	8 <sup>c</sup>	134–158	x	x		xx		xx
phi015	8.08	7	83–107 <sup>d</sup>	xx			xx	x <sup>f</sup>	
phi080	8.08	7	143–173 <sup>d</sup>				x		
phi233376	8.09	8 <sup>c</sup>	139–163	xxx	xxx		xxx	xxx	
umc1279	9.00	4	92–101	xxx	xxx		xxx		xxx
phi033	9.01	6	245–263 <sup>d</sup>				x		
phi022	9.03	-	-						
phi065	9.03	4	131–151	xxx		xxx	xxx		xxx
phi032	9.04	5	233–249	xxx	xxx		xxx	xxx	
phi108411	9.05	5 <sup>c</sup>	113–129	xxx			xxx		xxx
phi448880	9.06–9.07	7	173–197	xxx		xxx	xxx		xxx
umc1277 <sup>a</sup>	9.07–9.08	3	133–138	x	x			x	
phi041 <sup>a</sup>	10.00	4	196–216	x	x				x
phi059 <sup>b</sup>	10.02	4	147–162	xx			xx		x
phi063	10.02	10 <sup>c</sup>	147–223	xxx		xxx	xxx		xxx
phi96342 <sup>b</sup>	10.02	5	234–250	x		x	xx	xx	
umc1152	10.02	7	155–219	xxx			xxx		xxx
phi050 <sup>b</sup>	10.03	3 <sup>c</sup>	80–86	xx	xx		xx		x
phi062 <sup>b</sup>	10.04	2	161–164	xx			xx		x
phi084	10.04	4	153–162	xxx			xxx		xxx
umc1061	10.06	5	89–110	xxx		xxx	xxx		xxx
umc1196 <sup>b</sup>	10.07	6	137–161	xx	xx		xx	x	

<sup>a</sup> Marker cannot be compared because it had a lot of missing data or did not amplify in ASL

<sup>b</sup> Marker had more than a 10% error rate

<sup>c</sup> Includes allele(s) that do not follow repeat type

<sup>d</sup> Approximate sizes were determined manually on denaturing PAGE

<sup>e</sup> Alleles cannot be compared due to differences in the size range of the amplified products

<sup>f</sup> Alleles cannot be standardized because of stutters or the marker being a di- or compound repeat

<sup>g</sup> Alleles cannot be compared because they do not have sequencer-derived molecular weight-based names

refer to the relative size (1 for the largest of the alleles; 2 for the second largest of the alleles; 3 for the third largest of the alleles and so on). For example, the allele *umc1399-il* refers to the largest allele found in the 119- to 140-bp size range produced with the SSR primer umc1399 (Fig. 1). These  $\phi$ X174/*Hinf*I names of the alleles were later replaced by names based on the molecular sizes of the alleles.

#### Conversion of $\phi$ X174/*Hinf*I-based allele names to molecular weight (MW)-based names

Alleles representing all of the different alleles produced from each SSR primer (unique alleles) were collected, then reamplified and detected using the ABI 377 DNA sequencer (Perkin Elmer, Boston, Mass.) as described by Warburton et al. (2002a) in order to determine their molecular sizes and enable datasets to be merged with the CIMMYT ABC where a fluorescence-based, automated allele-sizing technology is used. Briefly, DNA from the line that contained the allele of interest was amplified using primers labeled at the 5' end with 6-carboxyfluorescein (6-FAM), tetrachloro-6-carboxyfluorescein (TET) or hexachloro-6-carboxyfluorescein (HEX) (Gibco, Gaithersburg, Md.). The PCR reactions were performed using a Peltier Thermal Cycler (MJ Research) following the amplification conditions and annealing temperatures specific for each primer combination. Samples containing two to five PCR products were mixed with 0.3–0.4  $\mu$ l GeneScan-350 or -500 internal lane standard labeled with *N, N, N, N*-tetramethyl-6-carboxyrhodamine (TAMRA) and 30% formamide, denatured at 95°C for 5 min and electrophoresed on 4.5% denaturing acrylamide in 1 $\times$  TBE buffer

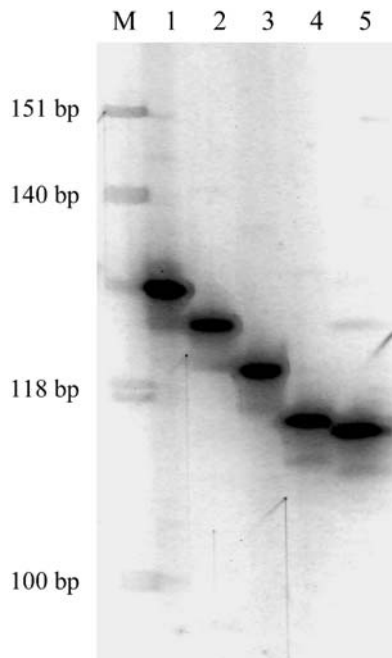
at constant voltage (3.00 kV) for 1.8–2.5 h. Fragment sizes were automatically calculated using GENESCAN 3.1 (Perkin Elmer/Applied Biosystems, Foster City, Calif.).

#### Standard alleles

To ensure that the alleles observed in the individual laboratories and the names given to them matched their designated  $\phi$ X174/*Hinf*I- or molecular size-based names, the ASL developed standard allele sets for each marker. The standard alleles consisted of a collection of the unique alleles detected at each SSR locus in the maize genetic diversity studies conducted in several countries in the network. The standard alleles were arranged as a “ladder” molecular weight marker and were packaged as a kit to the individual laboratories for use as reference markers (Fig. 1). The kits were provided as PCR products in microfuge tubes or as air-dried DNA samples in a 96-well microtiter plate for amplification with a specific SSR primer. The kits also contained a gel photograph of the  $\phi$ X174/*Hinf*I molecular weight marker and the alleles with their  $\phi$ X174/*Hinf*I -based names (and later with their molecular size-based names) and information on the PCR, electrophoresis and detection conditions.

#### Datasets used in this study

Each dataset contained the name of the lines fingerprinted, the name of the SSR markers used and the alleles detected as a 0/1



**Fig. 1** Standard alleles of SSR locus *umc1399* in 4.5% PAGE. *M* Molecular weight marker consists of fragments of  $\phi$ X174/*Hinf*I. Lanes (sequencer-derived molecular weight-based names are indicated in parenthesis): 1 *umc1399-i1* (*umc1399-127*), 2 *umc1399-i2* (*umc1399-123*), 3 *umc1399-i3* (*umc1399-119*), 4 *umc1399-j1* (*umc1399-115*), 5 *umc1399-j2* (*umc1399-113*). The allele *umc1399-j2* (*umc1399-113*) did not follow the tetramer repeat pattern and was not recorded (considered as missing data in the combined dataset) in the sequencer-based fingerprinting method of CIMMYT

matrix (0 for absence, 1 for presence). Missing data were indicated by a 9.

#### Country datasets

The country datasets from China and Indonesia were generated using the standard protocols adopted by the network. The dataset from China (130 lines  $\times$  75 SSRs, Table 2) had  $\phi$ X174/*Hinf*I-based allele names and consisted of tropical inbred lines developed at the Sichuan Agricultural University Maize Research Institute as well as local lines (pedigree information unavailable) collected from several provinces in China that make up the tropical/subtropical corn belt of the country. The inbred lines from Indonesia (Table 3) had molecular size-based names and consisted of tropical lines developed at the Indonesian Cereal Research Institute in Maros, South Sulawesi and the Research Institute for Agricultural Biotechnology and Genetic Resources in Bogor, West Java as well as lines obtained from CIMMYT. The data were partly generated at the ASL in the Philippines (dataset 1 with 79 lines  $\times$  30 SSRs, Table 2) and at the RIABGR in Indonesia (dataset 2 with 41 of the 79 lines  $\times$  20 SSRs, Table 2). These two datasets, each using a different set of markers, were merged into a composite dataset prior to being merged with others.

#### Regional dataset

A regional dataset from the ASL in the Philippines was made up of 102 inbred lines analyzed with 74 SSRs (Table 2). This dataset was generated using representative lines from the national programs in China, India, Indonesia, Philippines, Thailand and Vietnam, from the CIMMYT Asian Regional Maize Program (ARMP) in Thailand

and the Maize Program in Mexico as well as temperate lines from the USA and Germany (George et al. 2004). These data, having both manually derived  $\phi$ X174/*Hinf*I-based allele names as well as automatically derived molecular size-based names, served as the reference dataset to which the country and CIMMYT data were compared.

#### CIMMYT dataset

SSS marker data from the CIMMYT ABC were generated using an automatic DNA sequencer and had molecular size-based names (Warburton et al. 2002a). The CIMMYT dataset analyzed in this study is a composite derived from two separate datasets. CIMMYT dataset 1 (44 lines  $\times$  24 markers, Table 2) included nine downy mildew-resistant CMLs (CML425–433; Table 3) developed for the Asian tropical environments as well as representative subtropical, tropical mid-altitude and highland CMLs (Warburton et al. 2002b). CIMMYT dataset 2 (103 lines  $\times$  54 markers; Table 2) consisted of subtropical, tropical mid-altitude and highland CMLs developed by the CIMMYT Maize Program for these environments.

#### Comparison of datasets

The general agreement between datasets was assessed by calculating the error rates between the datasets in pair-wise comparisons. The calculation of error rates between ASL and the two countries was based on six genotypes used as reference lines (CML51, CML292, CML202, CML206, CML236, and CML396). Error rates were calculated separately in the comparison of ASL with the two datasets from Indonesia as well as with the two datasets from CIMMYT. There were three reference lines used for calculating the error rates between CIMMYT dataset 1 and ASL (CML51, CML292 and CML281) and six reference lines for ASL and CIMMYT dataset 2 (CML281, CML289, CML452, CML453, CML385 and CML387).

For calculating the error rates, mismatches were first counted by aligning the reference lines of the two datasets being compared using the SSR markers and allele names as the basis for alignment, comparing each pair of alleles and then counting each case where there was no 1/1 or 0/0 match. The alleles used in the alignment consisted of all expected alleles based on the repeat type within the size range observed in the datasets being compared, including alleles not yet in the standard allele collection. All alleles of the SSR locus not reported in the two datasets being compared were considered as matches (0/0 match). When an allele was reported as a 9 in one dataset and a 0 in another, a comparison was not considered possible since a match or a mismatch could not be declared—thus, the allele was subtracted from the total number of pair-wise comparisons. The overall (raw) error rate of the datasets was calculated by taking the number of pair-wise comparisons that were mismatched and dividing it by the total number of pair-wise comparisons between the two datasets. The individual error rates per marker were also calculated.

Allele mismatches were examined to determine if the error was correctable. For example, if all observed alleles of a locus were off by one repeat unit, the lines were assumed to have been scored this way and were corrected for a simple frameshift (Fig. 2). As well, some mismatches were corrected if the gel could be reviewed and rescored. An adjusted error rate was calculated, based on the remaining error (uncorrectable) after realignment of the frameshifts or rescoring of the data (Table 2). Errors that were considered to be uncorrectable included alleles with wrong designations, a failed amplification in one of the laboratories or a frameshift that was compounded by other errors (Fig. 2). All individual markers with more than 10% uncorrectable errors were excluded in the datasets prior to merging.

Also prior to the merging of the datasets, all lines with more than 20% heterozygosity and more than 15% missing data were excluded in all datasets. All markers with more than 15% missing data were excluded as well. Two or three datasets at a time were

**Table 2** Summary of the comparison of datasets with the AMBIONET Service Lab (ASL)

Comparison detail	Dataset compared with ASL [102 (lines) × 74 (markers)]				
	China	Indonesia		CIMMYT	
		Dataset 1	Dataset 2	Dataset 1	Dataset 2
Dataset (lines × markers)	130×75	79×30	41×20	44×24	103×54
Number of reference lines compared	6	6	6	3	6
Number of SSR markers in common	60	25	18	19	37
Total number of pair-wise comparisons	2,489	1,054	681	565	1,926
Number of pairs of alleles <sup>a</sup>	2,586	1,098	708	633	2,070
Amount of missing data	97	44	27	68	144
Total number of incorrect pair-wise comparisons	242	72	57	33	149
Number of differences due to frame-shift	173	34	34	0	6
Number of differences due to wrong size	45	20	12	18	121
Number of differences due to failed amplification	24	18	11	10	19
Number of differences due to extra allele	0	0	0	5	3
ERROR RATE BETWEEN DATASETS (RAW)	9.7%	6.8%	8.4%	5.8%	7.7%
Total number of correctable pair-wise comparisons	67	36	12	0	0
Simple frameshift	67	14	12	0	0
Rescored gel	0	22	0	0	0
Remaining number of incorrect pair-wise comparisons	175	36	45	33	149
Adjusted error rate <sup>b</sup>	7.0%	3.4%	6.6%	-	-
Number of SSR markers excluded due to a more than 10% error	14	1	4	4	9
Final error rate <sup>c</sup>	2.1%	3.1%	0.6%	4.2%	5.6%
Number of markers that were merged with ASL	45	24	14	15	28
Merged dataset	64×45	49×24	41×14	44×15	103×28

<sup>a</sup> Based on the total number of alleles from two sets of reference lines being compared, including all predicted alleles that follow the stepwise pattern of the SSR within the size range observed

<sup>b</sup> After the adjustment of frameshift errors and/or rescoring of bands

<sup>c</sup> After removing markers with more than a 10% error rate

merged to form a matrix of binary data, with the columns equal to genotypes (inbred lines) and the rows equal to the alleles of each primer.

#### Diversity analysis

Combined datasets were analyzed using NTSYS-PC ver. 2.02 (Rohlf 1999). Matrices of genetic similarities were constructed from the binary data using Jaccard's coefficient. Cluster analysis was carried out with the unweighted pair group method using arithmetic averages (UPGMA), and the relationships between inbred lines were visualized in a dendrogram. To assess the relative stability of the clusters, we ran bootstrap analyses with the WINBOOT program (Yap and Nelson 1994) using 400 repeated samplings with replacement.

## Results

### Correspondence between datasets

The overall error rates between datasets based on raw data ranged from 5.8% to 9.7% (average = 7.7%, summarized in Table 2). The mismatches between the ASL dataset and the five datasets were due to frameshifts (45% of total mismatches), wrong size (39% of total mismatches), failed amplification (15% of total mismatches), and “extra” alleles (1% of total mismatches).

The frameshift errors, which ranged from 47% to 72%, were observed mostly between ASL and the manually generated datasets (China and Indonesia). Up to half of these errors were correctable by the adjustment of simple

frameshifts or rescoring of the bands in cases where it was possible to review the gels. On the other hand, wrong-size errors occurred more frequently (55–81%) between ASL and the automatically generated datasets (CIMMYT), and none of these was correctable. A error range of 10% to 30% resulted from the “failed” amplification of an allele due to variation in experimental conditions or band interpretation in the different laboratories. For example, a “failed” amplification error may result when a faint band is considered to be an allele and scored as 1 in one laboratory but is ignored in another laboratory and consequently not scored as absent (0) or missing data (9). A small percentage of error (2–15%) was due to the “extra” alleles observed between the ASL and CIMMYT datasets. These “extra” alleles, whose polymorphisms do not follow stepwise mutation patterns (indels), were scored manually in ASL but could not be scored automatically and were thus ignored by CIMMYT.

The error rates were reduced to less than 8% (average = 5.6%) following adjustments of the correctable errors of the China and Indonesia datasets. There was a further reduction in the error rates of all datasets—to less than 6% (average = 3.3%)—following the exclusion of markers with more than a 10% individual error rate (Table 2). A total of 25 markers (Table 1), ranging in number from one (Indonesia dataset 1) to 14 (China dataset), were excluded due to high error rates. Based on the final error rates, the ASL and the China and Indonesia datasets showed a better correspondence (average error = 1.9%)

**Table 3** Genetic similarity (GS) of reference lines from datasets merged with the ASL dataset

Error rate	Genetic similarity						Average
	CML51	CML292	CML202	CML206	CML236	CML396	
China							
Raw = 9.7%	0.47	0.47	0.49	0.45	0.40	0.46	0.45
Adjusted = 7.0%	0.63	0.60	0.66	0.65	0.57	0.63	0.62
Final = 2.1%	0.88	0.92	0.88	0.88	0.80	0.86	0.87
Indonesia dataset 1							
Raw = 6.8%	0.58	0.64	0.66	0.70	0.64	0.59	0.64
Adjusted = 3.4%	0.73	0.82	0.85	0.89	0.73	0.82	0.81
Final = 3.1%	0.77	0.85	0.84	0.88	0.75	0.82	0.82
Indonesia dataset 2							
Raw = 8.4%	0.56	0.68	0.54	0.59	0.58	0.59	0.59
Adjusted = 6.6%	0.64	0.78	0.60	0.67	0.64	0.67	0.66
Final = 0.6%	1.00	0.93	1.00	1.00	0.88	1.00	0.97
							Average
ABC dataset 1							
Raw = 5.8%	CML51	CML292	CML281				0.56
Adjusted = no corrections made	0.52	0.52	0.62				
Final = 4.2%	0.60	0.56	0.78				0.64
							Average
ABC dataset 2							
Raw = 7.7%	CML452	CML453	CML385	CML387	CML 281	CML289	0.51
Adjusted = no corrections made	0.26	0.22	0.67	0.70	0.54	0.60	
Final = 5.6%	0.33	0.30	0.66	0.92	0.71	0.75	0.61

than those between the ASL and CIMMYT (average error = 4.9%).

#### Merging of datasets

The size of the dataset that could be merged was reduced to a large extent when the minimum standards of error rates (<10%), missing data (<15%) and heterozygosity (<20%) were applied. Approximately one-half of the lines from China and Indonesia were excluded from the datasets due to a high percentage of heterozygosity. After the removal of these lines and one marker with more than 15% missing data (*phi002*), the China dataset (64 lines) was merged with the ASL dataset (102 lines), making a combined China-ASL dataset of 166 lines × 45 markers (Table 2). For Indonesia, a composite dataset consisting of 37 lines × 38 markers was first prepared by combining the marker data of the lines fingerprinted in both datasets. Similarly, a composite dataset consisting of 17 lines × 43 markers was prepared for CIMMYT. Subsequently, each of the composite datasets from Indonesia and CIMMYT was merged with the ASL dataset.

The multiple combination of datasets from different laboratories resulted in larger datasets having more lines but fewer markers, as seen in the China-ASL-CIMMYT (183 lines × 32 markers) and Indonesia-ASL-CIMMYT (156 lines × 29 markers) datasets (Table 1). Although the markers used by the different laboratories were based on a common core set, the number of markers that could be combined was limited due to the different markers chosen

from the core set by each laboratory as well as the quality of the resulting data. Nevertheless, all chromosomes were represented in the remaining markers, with the coverage ranging from one to five markers per chromosome in the China-ASL-CIMMYT dataset and from two to four markers per chromosome in the Indonesia-ASL-CIMMYT dataset (Table 1).

#### Genetic similarity of reference lines

Similarity matrices were constructed from the binary data of the reference lines from the combined datasets in order to compare the correspondence of the lines at different error levels. There were substantial increases in the genetic similarity (GS) values of the reference lines when the errors were reduced. As the quality of the data improved, the GS values of the reference lines from the China-ASL dataset increased from an average of 0.45 when the error was 9.7% (raw, based on 60 markers) to 0.87 when the error was 2.1% (final, based on 46 markers). The same held true in the Indonesia-ASL dataset—an average GS of 0.64 when the error was 6.8% (raw, based on 25 markers) increased to 0.82 when the final error was 3.1% (based on 24 markers) in dataset 1, and from 0.59 (8.4% raw error, 18 markers) to 0.97 (0.6% final error, 14 markers) in dataset 2.

However, consistent with the only slightly reduced error rates in the ASL-CIMMYT datasets, there was also only a slight improvement in the GS values of the reference lines (average GS = 0.56 at a 5.8% error rate based

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		ASL lines						CIMMYT lines						# of mismatches
2		1	2	3	4	5	6	1	2	3	4	5	6	
3	<i>umc1399-127</i>	0	0	0	1	0	0	0	0	0	1	0	0	0
4	<i>umc1399-123</i>	0	0	0	0	0	0	1	1	1	0	1	0	0 (after adjustment)
5	<i>umc1399-119</i>	1	1	1	0	1	0	0	0	0	0	0	0	0 (after adjustment)
6	<i>umc1399-115</i>	0	0	0	0	0	0	0	0	1	0	0	1	2
7	<i>umc1399-113</i>	0	0	0	0	0	1	9	9	9	9	9	9	-
8	<i>phi072-163</i>	0	0	0	1	0	0	0	0	0	1	0	0	0
9	<i>phi072-155</i>	0	0	0	0	0	1	1	1	1	0	1	0	5
10	<i>phi072-151</i>	1	1	1	0	1	0	0	0	0	0	0	1	5
11	<i>phi072-147</i>	0	0	0	0	0	0	0	0	0	0	0	0	0
12														
13														

**Fig. 2** Examples of errors observed between datasets. An example of a correctable error is the simple frameshift (*B3:F5* and *H3:L5*), which can be adjusted by shifting the alleles of ASL lines 1–5 by one repeat unit (from *umc1399-119* to *umc1399-123*). Examples of uncorrectable errors are a mismatch due to wrong size (*G9:G10* and *M9:M10*), a mismatch due to failed amplification in one laboratory (*D6* and *J6*), and a mismatch due to a frameshift that is com-

pounded by another error (*B8:G10* and *H8:M10*). The allele *umc1399-113* (*B7:M7*), an “extra allele” observed in ASL but not in CIMMYT, was not included in the total number of alleles compared because a neither match nor a mismatch can be declared. All alleles of the SSR loci in the AMBIONET database but not reported in both comparisons (*B11:M11*) were considered as matches (0/0 match)

on 19 markers and average GS=0.64 at a 4.2% error rate based on 15 markers for dataset 1; for dataset 2, average GS=0.51 at a 7.7% error rate based on 37 markers and average GS=0.61 at a 5.6% error rate based on 28 markers). A better correspondence between the reference lines, as evidenced by higher GS values, was consistently achieved with fewer markers with higher quality of data rather than with more markers of questionable data.

#### Genetic relationships of inbred lines

A dendrogram showing the relationship between 58 inbred lines from six provinces in southern China and selected lines, including four temperate lines from northern China, nine downy mildew-resistant CMLs developed by the Maize Program of CIMMYT for Asia and two U.S. corn belt lines, is presented in Fig. 3. Seven major clusters could be discerned from the dendrogram in addition to one line from Guizhou province (SJP31) and two lines from CIMMYT (CML433 and CML431) that did not cluster with any group. Discernable subclusters were also found in the major clusters, including four subclusters in Clusters 1 and 4, three in Clusters 2 and 3 and two in Clusters 6 and 7. The U.S. corn belt lines—B73 of the BSSS heterotic group and Mo17 of the Lancaster heterotic group—were located in separate subclusters in Cluster 2 (GS = 0.24). Bootstrap values of the upper level branches of the dendrogram were high, indicating that these groupings are robust (Fig. 3).

The lines from southern China were a highly diverse group, being found in all of the seven major groups. However, the CIMMYT lines developed for the Asian

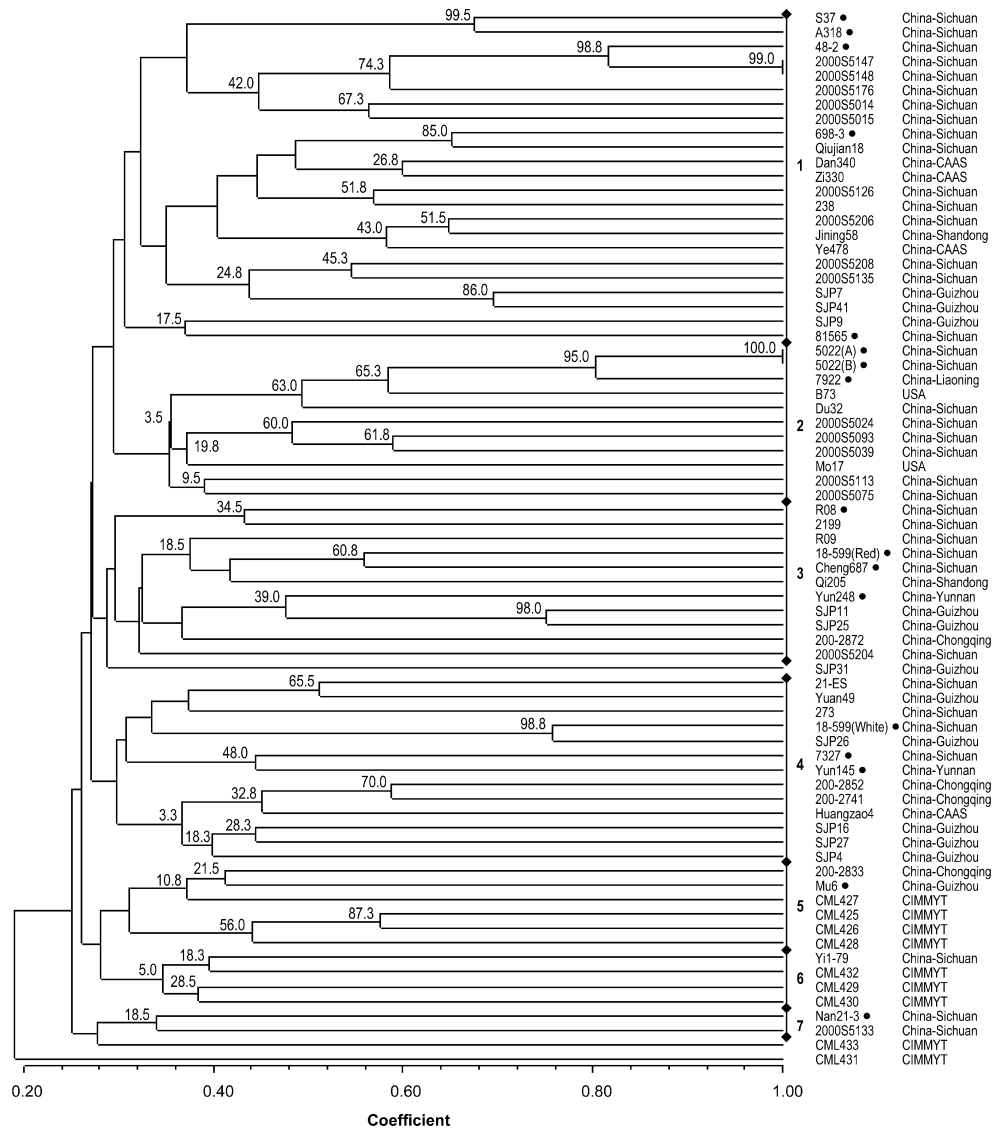
region showed a relatively narrow genetic base and were found in only two clusters. The CIMMYT lines formed a discrete germplasm group with only three lines from China, one each from the provinces of Chongqing, Guizhou and Sichuan, clustering with them. Two CIMMYT lines, CML431 and CML433, were each distinct and did not cluster with any group (Fig. 3).

Lines collected from Sichuan province (37 lines) were found in six major groups, those from Guizhou (12 lines) in four groups and those from Chongqing (4 lines), Shandong (2 lines), and Yunnan (2 lines) in two groups. One line from Liaoning province clustered with the lines from Sichuan and the USA. Some China lines closely related by pedigree, such as the sister lines 5022A and 5022B, clustered closely together as expected, but others like 18-599(R) and 18-599(W) did not. The dendrogram also shows the relationships of elite lines that are used as testers in China. Six of the seven major groups are well represented by the tester lines 48-2, S37, A318, 18-599, R08, 7327, 7922, 5022, 698-3, Cheng687, Nan21-3, Mu6, Yun145, Yun248 and 81565.

The temperate lines from northern China were found in three separate subclusters. The lines Dan340 and Zi330 of the Lu heterotic group and Ye478 of the PN heterotic group were in separate subclusters of Cluster 1 together with lines from the Sichuan and Guizhou provinces of southwest China and the more northern province of Shandong. The line Huangzao4 of the more distantly related Tangsipingtong heterotic group was found in Cluster 4 with lines from Sichuan, Guizhou and the more southern province of Chongqing. The placement of the temperate lines from northern China was generally con-



**Fig. 3** UPGMA dendrogram of maize inbred lines from China and downy mildew-resistant lines developed for Asia and CIMMYT (Warburton et al. 2002b) based on 32 SSR markers. Two U.S. corn belt lines (Mo17 and B37) were included for comparison. Inbred lines used as testers in China are indicated by a *solid circle*



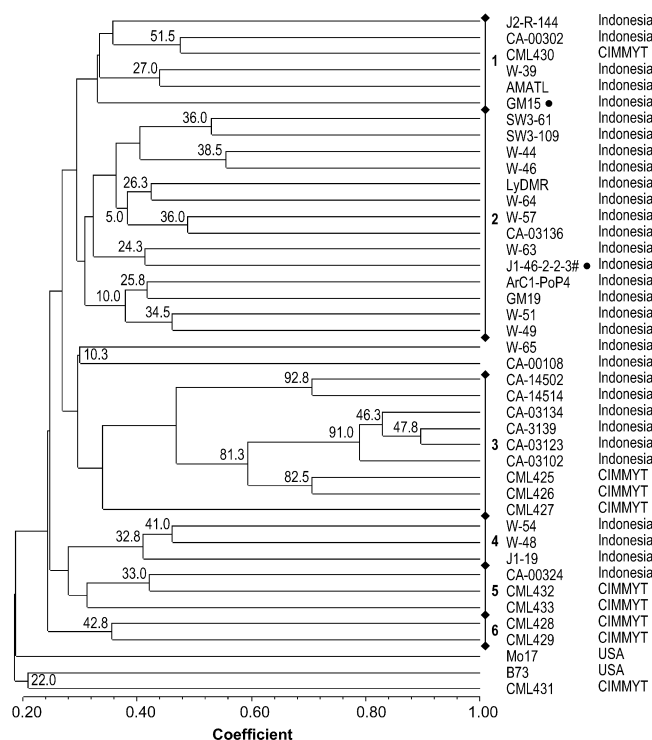
sistent with their pedigree and grouping as described in the study of Yuan et al. (2001).

The relationship of 31 inbred lines from Indonesia, nine downy mildew-resistant CMLs developed for Asia and two U.S. corn belt lines is shown in Fig. 4. Lines coming from the same background tended to cluster together (Table 4). Six clusters could be discerned in the dendrogram, with lines from the Indonesia dataset located in five of the six clusters. Clusters 1, 2 and 4 consisted of lines developed in Indonesia, and Cluster 3 consisted of Indonesian lines originally obtained from CIMMYT (CA series). Cluster 1, which included some Indonesian lines with an AMATL (Asian mildew acid-tolerant late-maturity) background, contained the downy mildew-resistant CML430 that also has AMATL in its background. Cluster 2, the largest cluster, was made up of 14 Indonesian lines whose background include CIMMYT populations 28 and 31, AMATL and the popular Suwan series developed in Thailand for resistance to downy mildew. Clusters 1 and 2 were represented by two of the

tester lines in use in Indonesia, GM15 and J1-46, respectively. Lines W-65 and CA-0010, CML431 and the U.S. temperate lines Mo17 and B73 (GS = 0.19) were each distinct and did not group with any other lines.

The downy mildew-resistant CMLs were located in four clusters, three of which were shared with Indonesian lines (Clusters 1, 3 and 5). In Cluster 3, the downy mildew-resistant CML425 and -426 (from population 31) and CML427 (from population 145) grouped together with six Indonesian lines (originally from CIMMYT) having similar backgrounds. One cluster of two lines (CML428 and CML429) were distinct from the Indonesian lines. Overall, the bootstrap values of the branches in the dendrogram are not high.

In contrast to the case in southern China where 95% of the lines clustered separately from the CIMMYT lines, lines in the Indonesian breeding program showed a closer relationship with the CIMMYT lines. This was expected, as Indonesia has a long history of germplasm exchange



**Fig. 4** UPGMA dendrogram of maize inbred lines from Indonesia and downy mildew-resistant lines developed for Asia in CIMMYT (Warburton et al. 2002b) based on 29 SSR markers. Two U.S. corn belt lines (Mo17 and B37) were included for comparison. Inbred lines used as testers in Indonesia are indicated by a *solid circle*

with CIMMYT, and many inbred lines in Indonesia have CIMMYT materials in their background.

## Discussion

In the study reported here we assessed the feasibility of an SSR marker method that was standardized for the analysis of maize genetic diversity in a network experiment involving four laboratories. We also compared and combined manually generated country and regional data with automatically generated data from an international center (CIMMYT) for a wide-scale analysis of genetic diversity.

Of the PCR-based molecular markers, amplified fragment length polymorphisms (AFLPs) and SSRs are considered to be more robust (Powell et al. 1996; Pejic et al. 1998) than random amplified polymorphic DNAs (RAPDs) for which reproducibility is an issue even under the most controlled conditions. SSR markers are particularly useful because they are highly polymorphic, are usually mapped and are easy to use. More importantly, as SSR scores can be described using allele sizes rather than bands, genotype information can be exchanged between different laboratories as numbers rather than images.

However, there are important reproducibility problems, especially in experiments involving several laboratories. In network studies involving several European laboratories, scoring differences of a few basepairs among

SSR fragments have been reported (Jones et al. 1997; Bredemeijer et al. 2002; Röder et al. 2002). In the present study, several types of errors occurred between datasets. Overall, frameshift errors were the most frequent, making up almost one-half of the observed mismatches. However, they were also frequently correctable. Errors in allele sizing, which also occurred at a high frequency, are easier to prevent than to correct. Most of the cases of wrong-size errors in the manually generated datasets occurred when the standard alleles of some primers were not used or when the standard allele nomenclature was not followed. With respect to automatically generated data, sizing is more precise, although some differences in allele sizes may occur as a result of differences in methodology. There was also a loss of information since fragment size differences that are not a multiple of the repeat unit was recorded as missing data. On the other hand, the use of standard alleles during manual sizing of the alleles provides a reasonable degree of accuracy that can reduce the occurrence of misidentified alleles and make it easier to confirm the identity of fragments across different studies.

This study shows that by using a standardized methodology and appropriate controls, it is feasible to compare and combine SSR fingerprint information for certain SSR markers from different laboratories. For diversity studies conducted in different laboratories to be compared, common reference lines and SSR markers as well as the use of the more repeatable markers are essential. Comparison is not possible when there are no reference lines on which to base the correspondence of the data being compared. Furthermore, the quality of the diversity analysis depends on having a sufficient number of markers, thus the choice of markers in the individual studies is critical. Although individual datasets have adequate markers for analyses, there can be a sharp decrease in the number of markers when two or three datasets with only a few common markers are combined. A closer coordination of activities prior to laboratory analyses would increase the pool of common markers.

The results of the wide-scale study of maize genetic diversity reported here could assist breeders to more efficiently choose genetically diverse parents for breeding programs and to systematically introgress traits such as downy mildew resistance from new germplasm. Furthermore, molecular marker information can help to monitor the level of genetic diversity in breeding materials as well as the purity of inbred lines.

This protocol and general approach provides a system for diversity studies which would work well in a network mode of collaboration and would enable many breeding programs to combine their data with others to obtain value-added information that could enhance breeding efficiency. Since the standard alleles, which have a dual nomenclature, enable manually sized datasets to be compared with sequencer-sized datasets, this methodology could be incorporated into national programs for routine use, thereby enabling those with limited facilities to make use of information from a global database. This is of particular value in crops like maize where substantial

**Table 4** Maize inbred lines from Indonesia, ASL and CIMMYT Applied Biotechnology Center

Maize line	Origin/pedigree <sup>a</sup>	Type <sup>b</sup>	Data source
AMATL	AMATLCO	Tropical Y F	Indonesia
ArC1-PoP4	Suwan 2	Tropical Y F	Indonesia
CA-00108	CIMMYT	Tropical Y F	Indonesia
CA-00302	AMATLCO	Tropical Y F	Indonesia
CA-00324	AMATLCO	Tropical Y F	Indonesia
CA-03102	Pop 31 (CIMMYT)	Tropical Y F	Indonesia
CA-03123	Pop 31 (CIMMYT)	Tropical Y F	Indonesia
CA-03134	Pop 31 (CIMMYT)	Tropical Y F	Indonesia
CA-03136	Pop 31 (CIMMYT)	Tropical Y F	Indonesia
CA-14502	Pop 145 (CIMMYT)	Tropical Y F	Indonesia
CA-14514	Pop 145 (CIMMYT)	Tropical Y F	Indonesia
CA-3139	Pop 31 (CIMMYT)	Tropical Y F	Indonesia
GM15	Malang Composite 9	Tropical Y F	Indonesia
GM19	Pop 28 (CIMMYT)	Tropical Y F	Indonesia
J1-19	Malang Synthetic J1, Harapan, Kalingga, Wiyasa, Malang Composite 9, 11, Muneng Synthetic, CIMMYT Pop 27, -28	Tropical Y F	Indonesia
J1-46-2-2-3#	Malang Syntetic J1	Tropical Y F	Indonesia
J2-R-144	Malang Syntetic J1	Tropical Y SD	Indonesia
LyDMR	Suwan 2	Tropical Y F	Indonesia
SW3-61	Suwan 3	Tropical Y F	Indonesia
W-39	AMATLCO	Tropical Y F	Indonesia
W-51	Pop 28 (CIMMYT)	Tropical Y F	Indonesia
W-44	Pop 28 (CIMMYT)	Tropical Y F	Indonesia
W-54	AMATLCO	Tropical Y F	Indonesia
W-64	KSX3602 (KU, CIMMYT)	Tropical Y F	Indonesia
W-46	AMATLCO	Tropical Y F	Indonesia
W-49	Pop 28 (CIMMYT)	Tropical Y F	Indonesia
W-65	Pop 345 (CIMMYT)	Tropical Y F	Indonesia
W-63	KSX3602 (KU, CIMMYT)	Tropical Y F	Indonesia
W-48	Population 345 (CIMMYT)	Tropical Y F	Indonesia
W-57	PIO3011 (CIMMYT, Pioneer)+C16	Tropical Y F	Indonesia
SW3-109	Suwan-3/SW3-109-3-2-2-2-##	Tropical Y F	Indonesia
CML425	P31C4S5B-6-##-BBBB	Tropical Y F	CIMMYT
CML426	P31C4S5B-38-##-2-BBB	Tropical Y F	CIMMYT
CML427	SW92145-2P9S2-##-4-BBBB	Tropical Y F	CIMMYT
CML428	SW91145-2P3S2-##-3-BBBB	Tropical Y F	CIMMYT
CML429	EY-DMR-G-C5-S2-B-B-3-1-BBBB	Tropical Y F	CIMMYT
CML430	AMATLCOHS169-1-1-1-1-2-2-1-B*3	Tropical Y F	CIMMYT
CML431	AMATLCOHS170-2-3-2-1-1-1-1-B*3	Tropical Y F	CIMMYT
CML432	KTX3752F2-7-1-1-1-B-B-B	Tropical Y F	CIMMYT
CML433	KTX3753F2-5-1-1-2-BBB	Tropical Y F	CIMMYT
B73	US: Lancaster Sure Crop (LSC)	Temperate Y D	ASL
Mo17	US: Iowa Stiff Stalk Synthetic C5 (BSSS)	Temperate Y D	ASL

<sup>a</sup> Pedigree: Pop or P, Population; C, cycle; HC, full sibs; B, selfed and bulked; -1,-2,-3, ear to row; #, sibling; AMATL, Asia mildew acid-tolerant late-maturity

<sup>b</sup> Type: Y, Yellow grain; D, Dent; SD, Semindent; F, Flint

resources are devoted to the extensive field-testing of crosses to identify lines with superior combining ability. Armed with a knowledge of how local lines relate to others, breeders would be able to make more informed choices of crosses, permitting a more efficient use of genetic resources in breeding programs. In addition, this methodology can be applied to a wide-scale study of diversity of many crops at various locations around the world.

**Acknowledgements** This work was supported by the Asian Development Bank, Sichuan Agricultural University, the Indonesia Agency for Agricultural Research and Development, and CIMMYT. We thank M. Khairallah (formerly in CIMMYT) for her contributions in the initial part of this study, and Sutrisno (RIABGR) for the use of the institute's facilities for the molecular work in Indonesia.

## References

- Bredemeijer GMM, Cooke RJ, Ganai MW, Peeters R, Isaac P, Noordijk Y, Rendell S, Jackson J, Röder MS, Wendehake K, Dijcks M, Amelaine M, Wickaert V, Bertrand L, Vosman B (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. *Theor Appl Genet* 105:1019–1026
- CIMMYT (1998) A complete listing of improved maize germplasm from CIMMYT. CIMMYT, Mexico, D.F.
- Dubreuil P, Dufour P, Drejci E, Causse M, de Vienne D, Gallais A, Charcosset A (1996) Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. *Crop Sci* 36:790–799
- George MLC, Regalado E, Warburton M, Vasal S, Hoisington D (2004) Genetic diversity of maize inbred lines in relation to downy mildew. *Euphytica* 135:145–155
- Gerpacio R (ed) (2001) Impacts of international maize breeding research in Asia 1966–1998. CIMMYT, Mexico, D.F.

- Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, Bredemeijer G, Vosman B, Matthes M, Daly A, Brettschneider R, Bettini P, Buiatti M, Maestri E, Malcevski A, Marmiroli N, Aert R, Volckaert G, Rueda J, Linacero R, Vazquez A, Karp A (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by network of European laboratories. *Mol Breed* 3:381–390
- Lu H, Bernardo R (2001) Molecular marker diversity among current and historical maize inbreds. *Theor Appl Genet* 103:613–617
- Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) *The genetics and exploitation of heterosis in crops*. ASA, CSSA, and SSSA, Madison, Wis., pp 99–118
- Messmer MM, Melchinger AE, Boppenmaier J, Brunklaus-Jung E, Herrmann RG (1992) Relationships among early European maize inbreds: I. Genetic diversity among flint and dent lines revealed by RFLPs. *Crop Sci* 32:1301–1309
- Messmer MM, Melchinger AE, Herrmann RG, Boppenmaier J (1993) Relationships among early European maize inbreds: II. Comparison of pedigree and RFLP data. *Crop Sci* 33:944–950
- Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. *Theor Appl Genet* 97:1248–1255
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* 2:225–238
- Pushpavalli SNCVL, Sudan C, Mohammadi SA, Nair SK, Prasanna BM, Gadag RN, Singh NN (2002) Analysis of simple sequence repeat (SSR) polymorphism in Indian maize inbred lines. *J Genet Breed* 56:229–236
- Reif JC, Melchinger AE, Xia XC, Warburton ML, Hoisington DA, Vasal SK, Srinivasan G, Bohn M, Frisch M (2003) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Sci* 43:1275–1282
- Röder M, Wendehake K, Korzun V, Bredemeijer G, Laborie D, Bertrand L, Isaac P, Rendell S, Jackson J, Cooke R, Vosman B, Ganai M (2002) Construction and analysis of a microsatellite-based database of European wheat varieties. *Theor Appl Genet* 106:67–73
- Rohlf FJ (1999) *NTSYS-PC: numerical taxonomy system*, ver. 2.1. Exeter, Setauket, N.Y.
- Warburton ML, Xianchun X, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002a) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci* 42:1832–1840
- Warburton ML, Xianchun X, Ambriz S, Diaz L, Vasal S, Gonzalez F (2002b) Genetic relationships among CIMMYT maize lines developed for Asia and from other CIMMYT breeding programs. In: *Proc 8th Asian Regional Maize Workshop*. Bangkok, Thailand
- Yap I, Nelson RJ (1996) WINBOOT: a program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendrograms. *IRRI Discussion Paper Series No. 14*. International Rice Research Institute, Manila, Philippines
- Yuan LX, Fu JH, Zhang SH, Liu X, Peng Z, Li XH, Warburton ML, Khairallah M (2001) Heterotic grouping of maize inbred lines using RFLP and SSR markers. *Acta Agron Sin* 27:149–156