Microsatellite markers in maize: challenges and guidelines for implementing multiplex SSR analyses

Alessandra Koltun[1,2], Priscila Alves da Silva[1,2], Isabela Yamauchi Torres[3], Fernando Bonifácio-Anacleto[4] and Juliana Erika de Carvalho Teixeira Yassitepe[1,2,5*]
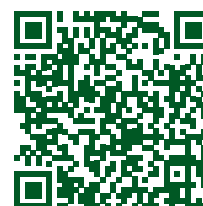
Abstract: *Microsatellites have been widely used to genotype individuals and to address a myriad of biological questions in many research fields for decades. However, when implementing a microsatellite marker analysis routine from scratch, various problems can arise throughout the process from DNA extraction to allele scoring, including inputting errors in the database, decreasing the reliability of results and having profound negative impacts on the derived decisions. Therefore, correctly assigning a genotype to a sample is crucial and dependent on acquiring knowledge of the technique steps, such as the chemistry of reactions, software, and data curation. This study tested two previously constructed simple sequence repeat (SSR) sets containing ten primer pairs each (ten-plex) in 1142 maize genotypes. Here, we describe the challenges faced when implementing this microsatellite-based genotyping protocol in our laboratory and possible ways to overcome them, hopefully aiding other novice research teams in this field.*

Keywords: *Multiplex PCR, manual allele scoring, pull-up, fluorophore interference, tropical maize*

*Corresponding author:
E-mail: juliana.yassitepe@embrapa.br
ORCID: 0000-0002-6862-3031

[1] Universidade Estadual de Campinas, Genomics for Climate Change Research Center, Cidade Universitária Zeferino Vaz, Barão Geraldo, 13083-970, Campinas, SP, Brazil
[2] Universidade Estadual de Campinas, Centro de Biologia Molecular e Engenharia Genética, Cidade Universitária Zeferino Vaz, Barão Geraldo, 13083-970, Campinas, SP, Brazil
[3] SEMPRE AgTech, Rodovia GO-164, km 25, 75920-000, Santa Helena de Goiás, GO, Brazil
[4] Universidade de São Paulo, Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, Av. Bandeirantes, 3900, Vila Monte Alegre, 14049-900, Ribeirão Preto, SP, Brazil
[5] Embrapa Agricultura Digital, Universidade Estadual de Campinas, Avenida Doutor André Tosello, 209, Cidade Universitária, 13083-886, Campinas, SP, Brazil

INTRODUCTION

Microsatellites or simple sequence repeats (SSRs) are tiny DNA sequences (1-6 bases long) that appear as tandem (side by side) repeats in genomes (Rafalski et al. 1996). These sequences are highly prone to mutations, resulting in a variable number of repetitions. Therefore, sequence length polymorphisms occur across individuals – typically between 5 and 40 repeats (Selkoe and Toonen 2006). Due to their high allelic diversity, genetically codominant nature, and advantages over other molecular markers (Selkoe and Toonen 2006, Varshney et al. 2008), microsatellite-based markers have been widely applied in genetic studies. For instance, SSRs may be used to investigate parentage, linkage, population genetics, genetic purity, fingerprinting, and evolutionary history, as well as to characterize inbred lines and varieties of cultivated plants and guide crossings and selection in breeding programs (Vigouroux et al. 2002 and references therein, Alzate-Marin et al. 2020, Kyi et al. 2022). Recently, with advances in sequencing technologies, other molecular markers, such as single-nucleotide polymorphisms (SNPs), have become more broadly used, mainly due to the reduction in the genotyping cost per sample. Nevertheless, using

microsatellite markers may still be the best genotyping alternative in some cases, such as characterizing germplasms in small-scale breeding programs.

The DNA surrounding a microsatellite locus – the flanking region – is highly conserved in a species and even across taxa (Glenn and Schable 2005). Therefore, microsatellite loci can be genotyped by polymerase chain reaction (PCR) amplification targeting their flanking regions. Fluorophore-labeled primers can be used, resulting in laser-induced fluorescent products. DNA fragments comigrating with an internal size standard are detected on a capillary electrophoresis system, converting these signals into digital data (Flores-Rentería and Krohn 2013). Fragments vary in size according to the number of microsatellite repeat units. Using different size ranges to distinguish fragments marked with the same fluorescent dye and different colors to discriminate fragments in near size ranges, it is possible to multiplex more than 20 markers in a single PCR and capillary injection (Guichoux et al. 2011, Flores-Rentería and Krohn 2013). In maize, the few published studies exploring multiplex PCR were restricted to 2-5 primers (Gethi et al. 2002, Wang et al. 2003, Tsonev et al. 2013), except for the method tested in this study with ten primer pairs designed by Wang et al. (2007).

Allele sizing and scoring and subsequent data analyses are performed by external software. Some of the significant problems that lead to allele miscalling include stutters; false peaks generated by Taq polymerase slippages (Taberlet and Luikart 1999); null alleles; nonamplifying alleles (Dakin and Avise 2004, Pompanon et al. 2005); allelic dropout, caused by low DNA quantity and/or quality; and preferential amplification of one allele, usually the smaller one, over the other (Pompanon et al. 2005). Most of these errors are generally related to sample quality and PCR procedures. On the other hand, the pull-up effect occurs in capillary electrophoresis when one fluorophore signal intensifies the signal of another color peak, even background peaks, representing another critical source of allele miscalling (Wang et al. 2018). The automated process of allele scoring allows the analysis of a massive number of samples and markers (Flores-Rentería and Krohn 2013). However, it may not always be possible to automate this step, and manual allele scoring must be performed or at least used to double-check the results. In any case, several errors may occur, as already mentioned, and the sources must be identified and mitigated to genotype individuals accurately (Hoffman and Amos 2005). A reliable dataset depends on wisely choosing labels, performing PCR in good chemical conditions to avoid poor amplification, and cautiously scoring alleles.

SSR markers have significantly contributed to characterizing the genetic diversity of crops for different purposes. For instance, knowledge of the relationship among breeding materials in maize is crucial for developing lines and planning crosses in hybrid production, speeding up the entire process (Patto et al. 2004 and references therein). In this article, we implemented a multiplex SSR-based marker protocol for maize and reported problems that may arise during this procedure, along with some tips to overcome them. Hopefully, our guidelines will help other novice research teams willing to use this genetic approach to increase their data accuracy, strengthening their scientific conclusions in diverse projects.

MATERIAL AND METHODS

Aiming to establish a maize genotyping pipeline in our laboratory, in collaboration with the company, we identified 1142 genotypes, including several inbred lines and hybrids, with SSR-based molecular markers. To this end, we searched published literature for existing microsatellite markers and previously tested primers. Very robust screening has already been conducted in maize by Wang et al. (2007), who proposed a straightforward and low-cost protocol using two ten-plex SSR sets, in which a group of reactions is simultaneously undertaken in a single reaction tube, enabling and facilitating large-scale maize genotyping (Supplementary Table 1).

Maize seeds provided by Sempre AgTech (https://sempre.agr.br/win, Santa Helena de Goiás, GO, BR) were germinated in a greenhouse for 10 days. Leaf samples were collected, and DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega). Tailed primers with fluorescent labels (Table 1) were synthesized by the Exxtend company (Exxtend Soluções em Oligos, Brazil).

Amplification reactions were optimized with the Qiagen Multiplex Kit (Qiagen) as follows: Qiagen Multiplex PCR Master Mix 1x, 0.03 µM of each primer, 1X Q-solution, and 60 ng template DNA, in a final volume of 20 µL. The thermocycler was programmed as follows: one cycle at 94 °C for 15 min; 40 cycles of 94 °C for 30 s, 55 °C for 90 s, and 72 °C for 60 s; and a final extension cycle at 60 °C for 30 min.

After DNA amplification, samples were diluted ten times (2 µL of the PCRs and 18 µL water), and 0.5 µL of the diluted PCR product was mixed with 9.5 µL of formamide:size standard mix (20:1). The prepared samples were denatured at 95 °C for 2 min and then run on an ABI3730 Genetic Analyzer for automated capillary electrophoresis fluorescence detection. GeneScan 500 LIZ™ dye (Thermo Fisher Scientific) was used as the size standard, with LIZ as the fluorescent dye. Allele scoring was manually performed using Peak Scanner (Applied Biosystems) to generate the electropherograms.

Allele frequency was manually calculated using Excel tools. The genetic variation of each locus was measured by determining the total number and frequency (%) of alleles per locus and the polymorphic information content (PIC). PIC was calculated using the formula PIC = 1-$\sum(p_i)^2$, where p_i is the frequency of each allele within the locus (Senior et al. 1998).

The presence and frequency of null alleles were estimated using the Micro-Checker program (Van Oosterhout et al. 2004). The error rate was estimated by comparing the individual and multiplex results in 117 samples, evaluating nonamplified/null alleles, differences in fragment size, and miscalled alleles. To detect incorrect records of allele sizes in the 19 SSR loci, based on inconsistency with the size of tandem repeats in each locus, we used LOCUS software (https://en-lifesci.tau.ac.il/profile/kosman, Kosman and Jokela 2019) in a subset of 370 individuals, removing those with missing data for more than three loci. The structure analysis was performed with a presence and absence binary matrix. All graphs were generated using GraphPad Prism version 8.0.

Table 1. Allele number, range, polymorphism information content (PIC), and the percentage of rare alleles found in this study for the SSR markers evaluated in a tropical maize population

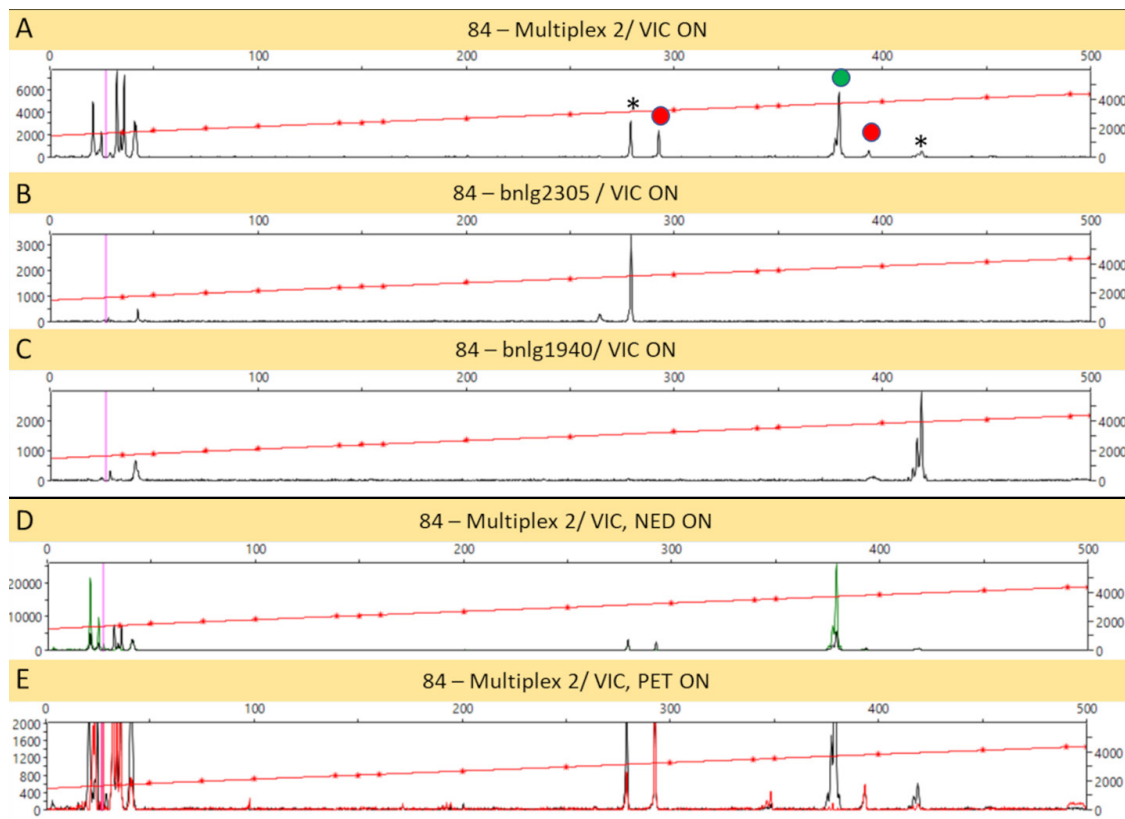| SSR marker	Allele number	Range	PIC	Rare alleles (%)
phi080	10	156-206	0.82	20
bnlg1754	19	201-237	0.81	42
bnlg1792	29	211-277	0.8	66
bnlg 161	30	145-211	0.6	60
umc1705	32	263-333	0.79	47
phi072	14	403-439	0.89	21
umc2163	39	123-217	0.94	36
phi125	43	219-333	0.94	47
phi065	6	397-427	0.62	50
phi116	23	139-187	0.7	52
umc2105	23	282-338	0.86	52
bnlg2331	35	345-439	0.8	53
bnlg1450	27	133-199	0.88	48
umc1741	37	277-373	0.88	46
bnlg1702	25	249-299	0.9	48
bnlg2291	23	365-413	0.9	35
umc2084	22	163-213	0.62	77
bnlg2305	23	249-301	0.88	57
bnlg1940	24	396-442	0.86	26

RESULTS AND DISCUSSION

Many steps exist between extracting DNA and entering a genotype into a database, and various errors can arise at each point. Genotyping errors can lead to a substantial number of incorrect genotypes in a large dataset (Hoffman and Amos 2005). Sources of error include poor amplification, misinterpreting stutter patterns or other artifact peaks for a real allele, contamination, mislabeling, or data entry errors, that is, technical and human causes (Bonin et al. 2004). In many cases, knowing the sources of error in the genotype data can allow correction for it, such as regenotyping individuals with no peaks to catch poorly amplified alleles and knowing how to discriminate true alleles from false peaks. For instance, amplifying each locus individually before initiating data collection for a multiplex reaction may guide identification of the abovementioned error sources and avoid the misidentification of alleles.

Here, we present practical information on applying previously developed primers for SSR markers in maize, specifically, two sets of multiplex PCR combinations based on a five-color fluorescence capillary detection system that was reported by Wang et al. (2007) (Supplementary Table 1 and Supplementary Figure 1). Multiplex PCR combined with automated fluorescence detection may significantly increase SSR analysis throughput and reduce costs associated with large-scale SSR genotyping. Using these 2 ten-plex sets, we reduced the cost for PCR amplification and fragment analysis by almost 10x compared with single marker analysis. However, even though it is a great option to decrease costs, allele scoring can be tricky.

A high-quality genetic dataset starts with good DNA extraction, substantially reducing subsequent technical difficulties with amplification. Then, after optimizing PCR amplification and running the samples through a fluorescence detector, several programs are available to perform fragment analysis of microsatellite electropherograms. Unfortunately, most are not open source and require the purchase of expensive licenses for unrestricted use. Applied Biosystems provides a

Figure 1. Electropherograms from simple sequence repeat (SSR) analysis visualized using Peak Scanner showing the results from sample 84 in the multiplex reaction and the individual primers for each marker. (A) Sample 84 in multiplex 2. Asterisks indicate true alleles, while red and green circles indicate pull-up effects from PET and NED, respectively. (B) Sample 84 was individually amplified with bnlg2305 and (C) bnlg1940, confirming the true alleles. (D) Sample 84 in multiplex 2 with VIC and NED dyes and (E) VIC and PET turned on, showing the pull-ups from NED and PET, respectively.

simple electropherogram viewer (Peak Scanner) for examining individual samples; however, analyses must be performed manually, as in this study.

The genetic material used in a given study should guide the researcher's expectations for the result; lines should exhibit homozygosity in the majority of loci, whereas in hybrids, heterozygote deficits should be a red flag (see Selkoe and Toonen 2006 for practical guidance). Here, in our first analysis, we found many alleles for most loci and individuals, which was unexpected since most genotypes are maize lines. Then, we identified that bleed-through (pull-up) signals were occurring (Supplementary Figure 2).

As pull-ups were frequent among dyes, especially VIC (green) over NED (black) and FAM (blue) over PET (red), we performed single reactions with each primer in a subset (117) of samples to identify some of the true alleles. Then, we carefully turned on multiple dyes simultaneously to identify the pull-ups and correctly assign the genotypes (Figure 1). The highest peaks were not necessarily the true alleles. The pull-up effect occurs when a color peak reaches intensity saturation and intensifies another color peak's signal, even background signals (Flores-Rentería et al. 2013). Previous studies have focused on pull-up problems (Flores-Rentería et al. 2013, Wang et al. 2018). For instance, the HEX fluorophore introduced extra signals/peaks in the ROX fluorophore (standard) channel, hampering size calibration and leading to allele miscalling (Wang et al. 2018). Possible solutions were provided, such as avoiding overloading the PCR products in capillary electrophoresis.

Identifying genotyping errors and estimating error rates can be performed via several approaches (Hess et al. 2012). One involves regenotyping a subset of randomly selected individuals, beginning with DNA extraction, to compare the
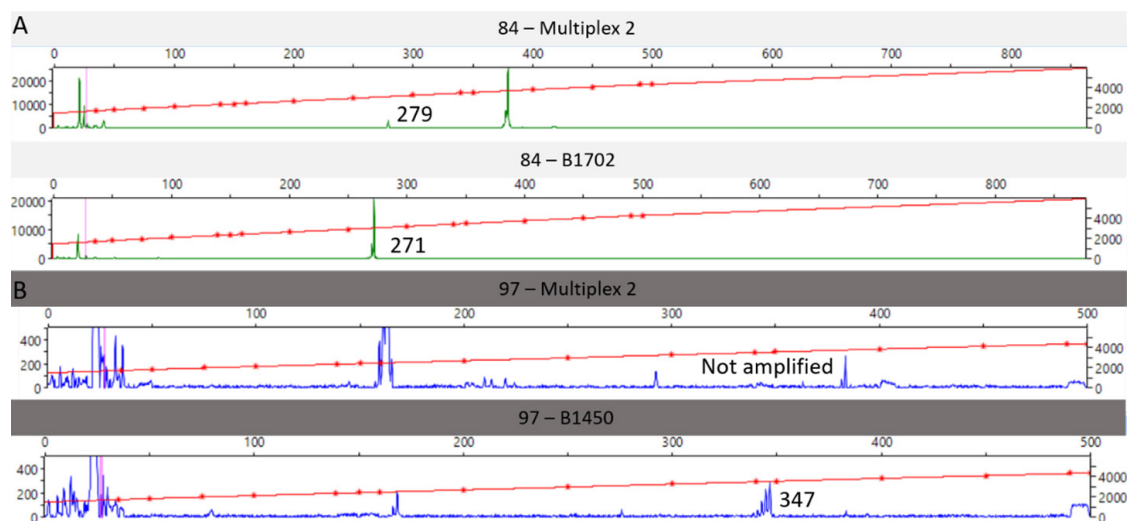
Figure 2. Electropherograms from SSR analysis visualized using Peak Scanner showing the results from (A) sample 84 in the multiplex reaction and individual primers for marker bnlg1702, exemplifying different allele sizes amplified by each reaction. (B) Results from sample 97 in multiplex 2 and individually amplified with bnlg1450 represent a case of lack of amplification of the true allele.

two datasets, although this method is laborious (Ewen et al. 2000). Bonin et al. (2004) recommended that at least 5–10% of samples be replicated. Here, the error rate was estimated by comparing the individual (in approximately 10% of the population) and multiplex results. Three analyses were performed, called A1 to A3. A1 included all primers and the initial evaluation of data that generated multiple peaks/alleles for each locus, A2 concerns the values after pull-up curation, in which errors from dye interference were removed, and A3 only considered the best primers, that is, discarding markers that resulted in many errors in the subsample analyzed. For instance, primers umc2163 and phi125 from multiplex 1 were removed in A3 due to the many miscalled alleles. In addition, bnlg1702 from multiplex 2 was discarded due to different sizes between multiplex and individual analyses, showing high interreference using the multiplex approach (wrong allele calls) and no amplification in many cases (Figure 2). This discrepancy in allele size or lack of alleles may be due to interference in fragment amplification in the multiplex PCR.

The comparison between the results of multiplex versus individual amplification of 117 samples in the first (A1), second (after pull-up curation; A2), and third analyses (considering only the best primers totaling 99 samples; A3) showed that after pull-up curation, miscalled alleles dropped 6.8% (from 17.9% in A1 to 11.1% in A2) (Supplementary Figure 5). In addition, removing the loci that generated a high number of errors increased true alleles by 9.9% (from 70.9% in A1 and A2 to 80.8% in A3) in the subsample analyzed. Finally, even after allele correction, we found three alleles in some genotypes (rare cases), which may indicate duplicated regions in the genome. After pinpointing the true alleles, the ranges for each locus were redefined, and the entire dataset was reevaluated. This approach was crucial to pinpoint the errors and improve genotyping quality.

As previously stated, some primers did not work well due to a high percentage of null alleles (bnlg439 resulted in nearly no allele amplification) or error rates (size difference and miscalled alleles) and were removed at some point in the analyses. Nonamplification may occur as a result of mutations in the primer region (Paetkau and Strobeck 1995). In other cases, a high number of nonamplified alleles was observed, such as 62.3% for bnlg161, 41% for bnlg2331, 52.4% for phi116, and 57.5% for umc2084 (Figure 5). However, these samples resulted in normal amplification of other loci (suggesting that the problem was not due to poor-quality DNA). Moreover, we assessed the presence of null alleles by comparing observed and expected heterozygosity at each locus using four different methods that assume that heterozygote deficiency is caused by null alleles and not by other genotyping errors or deviations from panmixia (Supplementary Table S1). Null alleles were confirmed to be present in all SSR markers of the two multiplex sets, exhibiting an incidence rate of 18.12% (0.591 ± 0.107). All the information gathered may indicate the best SSR loci for further analyses.

In general, the 19 SSR markers used in this study were highly polymorphic. The total number of alleles per locus ranged from 6 to 43 (Supplementary Figure 3), amplifying 482 alleles in the maize population evaluated and exhibiting high polymorphic information content (0.82 on average, ranging from 0.60 to 0.94). PIC values estimate the discriminatory power of each locus since they take into account not only the number of alleles but also their frequency. Our results were similar to those obtained by Belicuas et al. (2009), who analyzed 21 SSR markers in maize and found eight to 36 alleles per locus and an average PIC of 0.85 (varying from 0.65 to 0.96). Our values were superior to those detected in a study carried out by Smith et al. (1997), who found an average PIC of 0.62 for 131 SSR markers in 58 maize lines and four hybrids. Here, we used markers already selected by Wang et al. (2007) for being highly informative.

Interestingly, half of the alleles found in the germplasm had a <1% frequency. It is difficult to affirm whether they are errors due to the multiplex approach, although care was taken to minimize them to the lowest levels possible, or they truly represent the richness of the tropical population with more than a thousand genotypes evaluated in this study. Tropical and subtropical lines are more diverse and contain more rare alleles than temperate lines (Yan et al. 2009). Wang et al. (2023) found that more than half (1.4 M, 52.8%) of the variants were rare (minor allelic frequency <5%) in a study genotyping maize germplasm with 228 diverse accessions.

Additionally, Hamblin et al. (2007) evaluated the genetic diversity of sweet corn inbred lines and found an SSR dataset dominated by rare alleles. Finally, Laosatit et al. (2022) assessed the genetic diversity of 268 sweet maize lines using 20 SSR markers and found up to 30% rare alleles in subgroups of this population. Indeed, maize germplasm exhibits remarkable diversity and harbors distinctive and uncommon alleles, which can play a significant role in their preservation as a genetic resource. Favorable alleles that are absent in elite germplasm are invaluable sources for breeding programs, as they provide novel combinations for agronomic traits of interest.

Another probable cause for null alleles is genome structural variation. Maize genome complexity has been exploited in pangenomic analyses (Jin et al. 2016, Hufford et al. 2021). The pangenome of a species comprises core sequences (present in all individuals in the species) and dispensable sequences (present in only a subset of individuals within the species). In maize, it is estimated that there are at least 1.5x more genes in the pangenome than in an individual's genome (~20,000 more genes), with even more variation outside the gene regions (Brohammer et al. 2018). Growing evidence associates structural variation with phenotypic diversity and is hypothesized to contribute to the high levels of heterosis often observed in maize hybrids (Brohammer et al. 2018).

Additionally, the extreme genetic diversity of maize is well known and justifies its remarkable adaptation in tropical and temperate regions (Joets et al. 2018). This diversity includes small-scale variation, such as SNPs and insertions/deletions (indels), and copy number and presence/absence variations (Swanson-Wagner et al. 2010). For instance, Hufford et al. (2021) visualized a high structural diversity of 26 maize inbred genomes in both genic and repetitive regions. Therefore, the absence of a given SSR marker may be due to genetic diversity among the more than 1000 lines and/or structural rearrangements.

We used LOCUS software to identify erroneous records of allele sizes within 19 SSR loci. This identification was based on discrepancies between the size of tandem repeats in each locus within a subset of 370 individuals. Since an SSR mutation can be considered an error, researchers can run their database through this program to employ the infinite alleles model (IAM) and the stepwise mutation model (SMM). These models are employed to elucidate the evolution of new microsatellite alleles, with IAM assuming a random process where all possible alleles are equally likely to result from a mutation at a given microsatellite locus and SMM describing microsatellite evolution as a stepwise change involving the addition or subtraction of single or a few repeat units (Kosman et al. 2019). In contrast to SMM, IAM does not consider a specific mechanism of change in allele size. When the SMM approach is adopted, it is implied that their genetic distance influences the relative similarity in allele sizes of microsatellites between two individuals. In our data subset, we observed two incorrect allele sizes at four SSR loci: phi072, phi116, umc2105, and umc2084. For the marker phi072, alleles 425 and 411 were incorrectly detected in 15.7 and 10.5% of individuals, respectively. For phi116, alleles 165 (8.9%) and 175 (4.9%) were incorrect; for umc2105, alleles 290 (25.1%) and 324 (2.4%) were incorrect; and for umc2084, alleles 189 (28.4%) and 199 (5.9%) were incorrect. We also observed a high index of missing data for two of these markers, phi116 (17%) and umc2084 (24%), in addition to alleles exceeding the amplification range (Supplementary Table 2). Thus, when assessing the similarity or dissimilarity between the SSR genotypes of individuals, consideration should be given to allele sizes (SMM), which can provide a more robust analytical tool than solely comparing the number

of loci where individuals have different alleles (IAM) (Kosman and Jokela 2019).

Structural variation analysis revealed 5 (bnlg2291) to 63% (bnlg161) absence of allele amplification (Supplementary Figure 7). The four markers that presented higher absence rates (bnlg161, phi116, bnlg2331, and umc2084) are located in chromosome arms. Grzybowski et al. (2023) discovered more than 46 million high-confidence sequence variants in 1,515 maize individuals. The group found that segregating SNPs were more frequent around pericentromeric regions, while segregating indels were more common on chromosome arms. Moreover, linkage disequilibrium was typically elevated in pericentromeric regions, likely reflecting lower recombination rates in these regions. In this study, the SSR markers with greater absence (bnlg161, phi116, bnlg2331, and umc2084) were located in chromosome arms, whereas markers with low absence (bnlg1792k8 and bnlg2291) were closer to the chromosome centromeres. The pattern of linkage disequilibrium within presence-absence variation dramatically differs from that of flanking regions and corroborates the intuition that these sequences may recombine less than other genomic regions. The presence and absence of variation exist in specific materials, indicating that structural features may be at the origin of adaptive traits involved in the success of a given material (Joets et al. 2018).

CONCLUSION

Microsatellite markers have been successfully used for decades in many research areas. Although this technique is highly established, implementing an approach from scratch is not trivial. Even when loci are carefully screened and selected, unexpected errors may arise when applying them to different populations, as seen here. Therefore, increased reporting on the problems and solutions stumbled upon during the process is extremely helpful and should be encouraged. This synthesis of practicalities is one more humble step toward building a manual of accessible information for newcomers willing to utilize SSR markers to address their biological questions based on a more reliable dataset.

ACKNOWLEDGEMENTS

REFERENCES

Alzate-Marin AL, Costa-Silva C, Rivas PMS, Bonifacio-Anacleto F, Santos LG, Moraes Filho RMD and Martinez CA (2020) Diagnostic fingerprints ISSR/SSR for tropical leguminous species *Stylosanthes capitata* and *Stylosanthes macrocephala*. **Scientia Agricola 77**: e20180252.

Belicuas SNJ, Guimarães CT and Magalhães JV (2009) Caracterização molecular de milho e sorgo para aplicação nos programas de melhoramento da Embrapa. Embrapa Milho e Sorgo, Sete Lagoas, 7p. (**Comunicado Técnico, 177**).

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C and Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. **Molecular Ecology 13**: 3261-3273.

Brohammer AB, Kono TJY and Hirsch CN (2018) The maize pan-genome. In Bennetzen J, Flint-Garcia S, Hirsch C and Tuberosa R (eds) **The maize genome. Compendium of plant genomes.** Springer, Cham, p. 13-29.

Dakin EE and Avise JC (2004) Microsatellite null alleles in parentage analysis. **Heredity 93**: 504-509.

Ewen KR, Bahlo M, Treloar SA, Levinson DF, Mowry B, Barlow JW and Foote SJ (2000) Identification and analysis of error types in high-throughput genotyping. **The American Journal of Human Genetics 67**: 727-736.

Flores-Rentería L and Krohn A (2013) Scoring microsatellite loci. In Kantartzi S (ed) **Microsatellites: Methods and protocols.** Human Press, Totowa, p. 319-336.

Gethi JG, Labate JA, Lamkey KR, Smith ME and Kresovich S (2002) SSR variation in important US maize inbred lines. **Crop Science 42**: 951-957.

Glenn TC and Schable NA (2005) Isolating microsatellite DNA loci. **Methods in Enzymology 395**: 202-222.

Grzybowski MW, Mural RV, Xu G, Turkus J, Yang J and Schnable JC (2023) A common resequencing-based genetic marker data set for global maize diversity. **The Plant Journal 113**: 1109-1121.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F and Petit RJ (2011) Current trends in microsatellite genotyping. **Molecular Ecology Resources 11**: 591-611.

Hamblin MT, Warburton ML and Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. **PloS ONE 2**: e1367.

Hess MA, Rhydderch JG, LeClair LL, Buckley RM, Kawase M and Hauser L (2012) Estimation of genotyping error rate from repeat genotyping, unintentional recaptures and known parent–offspring comparisons in 16 microsatellite loci for brown rockfish (*Sebastes auriculatus*). **Molecular Ecology Resources 12**: 1114-1123.

Hoffman JI and Amos W (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. **Molecular Ecology 14**: 599-612.

Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, Coletta RD, Tittes S, Hudson AI, Marand AP, Wei S, Lu Z, Wang B, Tello-Ruiz MK, Piri RD, Wang N, Kim DW, Zeng Y, O'Connor CH, Gilbert AM, Baggs E, Krasileva SV, Portwood II JL, Cannon EKS, Andorf CM, Manchanda N, Snodgrass SJ, Hufnagel DE, Jiang Q, Pedersen S, Syring ML, Kudrna DA, Llaca V, Fengler K, Schmitz RJ, Ross-Ibarra J, Yu J, Gent JI, Hirsch CN, Ware D and Dawe RK (2021) De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. **Science 373**: 655-662.

Jin M, Liu H, He C, Fu J, Xiao Y, Wang Y, Xie W, Wang G and Yan J (2016) Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. **Scientific Reports 6**: 1-12.

Joets J, Vitte C and Charcosset A (2018) Draft assembly of the F2 European maize genome sequence and its comparison to the B73 genome sequence: a characterization of genotype-specific regions. In Bennetzen J, Flint-Garcia S, Hirsch C and Tuberosa R (eds) **The maize genome. Compendium of plant genomes.** Springer, Cham, p. 3-12

Kosman E and Jokela J (2019) Dissimilarity of individual microsatellite profiles under different mutation models: Empirical approach. **Ecology and Evolution 9**: 4038-4054.

Kyi S, Win KK, Than H, Win S, Htwe N and Hlaing A (2022) DNA Fingerprinting of selected maize (*Zea mays* L.) genotypes using SSR markers. **Environmental and Rural Development 13**: 158-163.

Laosatit K, Amkul K, Somta P, Tanadul O, Kerdsri C, Mongkol W, Jitlaka C, Suriharn K and Jompuk C (2022) Genetic diversity of sweet corn inbred lines of public sectors in Thailand revealed by SSR markers. **Crop Breeding and Applied Biotechnology 22**: e431322410

Paetkau D and Strobeck C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. **Molecular Ecology 4**: 519-520.

Patto MV, Satovic Z, Pêgo S and Fevereiro P (2004) Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. **Euphytica 137**: 63-72.

Pompanon F, Bonin A, Bellemain E and Taberlet P (2005) Genotyping errors: causes, consequences and solutions. **Nature Reviews Genetics 6**: 847-859.

Rafalski DJA, Vogel JM, Morgante M, Powell W, Andre C and Tingey SV (1996) Generating and using DNA markers in plants. In Birren B and

Lai E (eds) **Non mammalian genomic analysis: a practical guide**. Academic Press, p. 75-134.

Selkoe KA and Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. **Ecology Letters 9**: 615-629.

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

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S and 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. **Theoretical and Applied Genetics 95**: 163-173.

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D and Springer NM (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. **Genome Research 20**: 1689-1699.

Taberlet P and Luikart G (1999) Non-invasive genetic sampling and individual identification. **Biological Journal of the Linnean Society 68**: 41-55.

Tsonev S, Velichkova M, Todorovska E, Avramova V and Christov NK (2013) Development of multiplex primer sets for cost efficient SSR genotyping of maize (*Zea mays*) mapping populations on a capillary sequencer. **Bulgarian Journal of Agricultural Science 19**: 5-9.

Van Oosterhout C, Hutchinson WF, Wills DP and Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. **Molecular Ecology Notes 4**: 535-538.

Varshney RK, Thiel T, Sretenovic-Rajicic T, Baum M, Valkoun J, Guo P, Grando S, Ceccarelli S and Graner A (2008) Identification and validation of a core set of informative genic SSR and SNP markers for assaying functional diversity in barley. **Molecular Breeding 22**: 1-13.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC and Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. **Molecular Biology and Evolution 19**: 1251-1260.

Wang F, Zhao J, Dai J, Yi H, Kuang M, Sun Y, Yu X, Guo J and Wang L (2007) Selection and development of representative simple sequence repeat primers and multiplex SSR sets for high throughput automated genotyping in maize. **Chinese Science Bulletin 52**: 215-223.

Wang FG, Zhao JR Guo JL, Chen G, Liao Q, Sun SX and Chen RM (2003) Serial study on the establishment of DNA fingerprint of new maize varieties in China III: The use of multiplex PCR technique in maize SSR primer amplification. **Maize Science (in Chinese) 11**: 3-6.

Wang W, Guo W, Le L, Yu J, Wu Y, Li D, Wang Y, Wang H, Lu X, Qiao H, Gu X, Tian J, Zhang C and Pu L (2023) Integration of high-throughput phenotyping, GWAS, and predictive models reveals the genetic architecture of plant height in maize. **Molecular Plant 16**: 354-373.

Wang ZF, Dai SP, Lian JY, Chen HF, Ye WH and Cao HL (2018) Allele size miscalling due to the pull-up effect influencing size standard

calibration in capillary electrophoresis: A case study using HEX fluorescent dye in microsatellites. In Abdurakhmonov IY (ed) **Genotyping**. IntechOpen, London, p. 31-46.

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD and Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. **PloS ONE 4**: e8451.