ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 18, Issue, 04, pp.36770-36774, April, 2026
DOI: https://doi.org/10.24941/ijcr.50347.04.2026

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

MINING OF MICROSATELLITE MARKERS IN LEGUME CROPS

Jinal Naik and *Jitendra K Pal

Department of Biotechnology, Vanita Vishram Women's University, Surat

ARTICLE INFO

Article History:
Received 14th January, 2026
Received in revised form
24th February, 2026
Accepted 25th March, 2026
Published online 30th April, 2026

Keywords:
Comparative Genomics Microsatellite Markers Genetic Diversity, Breeding.

**Corresponding author: Jitendra K Pal*

ABSTRACT

Microsatellites or Simple Sequence Repeats (SSRs) are short tandem repeats widely distributed across plant genomes and are considered highly informative molecular markers due to their abundance, co-dominant inheritance, and high polymorphism. In legume crops, SSR markers are extensively utilized for genetic diversity studies, linkage mapping, and marker-assisted breeding. However, comprehensive genome-wide mining and comparative analysis of SSRs across major legume species remain limited. In the present study, genome-wide mining of SSR markers was performed using publicly available genomic sequences of major legume crops, including Glycine max, Cicer arietinum, Phaseolus vulgaris, and Vigna radiata. Bioinformatics tools were employed to identify SSR loci, classify repeat motifs (di-, tri-, tetra-, penta-, and hexa-nucleotides), and analyze their genomic distribution. Primer pairs were designed for selected SSR loci to evaluate their potential applicability in molecular studies. A large number of SSR loci were identified across all analyzed genomes, with di-nucleotide and tri-nucleotide repeats being the most predominant classes. Tri-nucleotide repeats were mainly enriched in coding regions, indicating their possible functional and evolutionary significance, while di-nucleotide repeats were more abundant in non-coding regions, suggesting higher variability. Comparative analysis revealed species-specific differences in SSR density and motif distribution. The designed SSR primers demonstrated high potential for cross-species transferability and polymorphism, highlighting their utility in genetic diversity analysis and breeding applications.

Citation: Jinal Naik and Jitendra K Pal. 2026. "Mining of microsatellite Markers in legume crops.". ***International Journal of Current Research, 18, (04), 36770-36774.***

INTRODUCTION

Molecular markers based on simple sequence repeats (SSRs) are extensively utilized in genetic and genomic studies owing to their co-dominant expression, widespread distribution across genomes, reproducibility, high polymorphism, and their potential for application across different species (Gaitán-Solís *et al*., 2002; Thiel *et al*., 2003; Saha *et al*., 2004). SSRs derived from coding or expressed regions of the genome tend to exhibit greater conservation, as these sequences are functionally important and subject to selective constraints. Consequently, markers developed from transcript sequences are more likely to be applicable across taxonomically related species than those originating from non-coding regions (Yu *et al*., 2004; Zhang *et al*., 2005; Parida *et al*., 2006). The evaluation of molecular markers developed in one species for their effectiveness in related taxa has become a routine strategy in comparative genomics and genetic mapping (Paterson *et al*., 2000). Numerous studies have reported successful cross-species amplification of SSR markers in plants, demonstrating their utility beyond the species of origin (Dayanandan *et al*., 1997; Brondani *et al*., 1998; Collevatti *et al*., 1999). For example, investigations involving wheat-derived EST-SSR markers across several members of the Poaceae family have revealed that transferability is closely associated with evolutionary relatedness among species (Gupta *et al*., 2003; Wang *et al*., 2005). In addition to enabling comparative genomic analyses, such marker systems also reduce the time and cost associated with marker development and genotyping, thereby facilitating large-scale genetic studies. Pigeonpea (*Cajanus cajan* (L.) Millsp.) is a key legume crop cultivated in tropical and subtropical agro-ecosystems. Although it was previously regarded as a neglected crop, recent advances have significantly improved the availability of genomic resources, including sequence data and molecular markers, along with the release of a draft genome assembly (Odeny *et al*., 2009; Varshney *et al*., 2010; Raju *et al*., 2010; Dutta *et al*., 2011; Bohra *et al*., 2011; Singh *et al*., 2011). Multiple lines of evidence, such as the existence of wild relatives, genetic diversity patterns, archaeological findings, and linguistic records, indicate that pigeonpea was domesticated in the Indian subcontinent (Fuller, 2006). Understanding the cross-transferability of genic SSR markers is particularly important for broadening the genetic base of cultivated pigeonpea through the incorporation of useful traits from related species. Moreover, such studies contribute to comparative genome research, as legumes are known to share a considerable degree of genomic conservation across both model and cultivated species (Weeden *et al*., 1992; Choi *et al*., 2004). The family Fabaceae represents one of the largest groups of angiosperms,

comprising more than 700 genera and approximately 20,000 species. From an agricultural perspective, legumes rank second only to cereals in global importance, contributing nearly 27% of total crop production (Graham & Vance, 2003). Based on phylogenetic relationships, this family is divided into three principal subfamilies: Mimosoideae, Caesalpinioideae, and Papilionoideae (Doyle *et al*., 2003). The Papilionoideae subfamily encompasses most economically significant legume crops and is further categorized into four major evolutionary lineages: Dalbergioid (e.g., groundnut), Genistoid (e.g., lupins), Hologalegina (e.g., chickpea, lentil, pea, and alfalfa), and Millettioid (e.g., soybean, pigeonpea, mungbean, and common bean). Major legume crops such as pigeonpea, cowpea, chickpea, lentil, common bean, and groundnut are integral to the livelihoods of farmers in developing regions. However, their productivity is frequently affected by environmental stresses and biological constraints. This highlights the need for efficient molecular tools to identify and utilize genes associated with stress tolerance. Previous research has demonstrated that SSR markers developed from model species like *Medicago truncatula* can be successfully applied to related genera such as *Pisum*, *Cicer*, and *Lens*, indicating substantial cross-species compatibility (Gutierrez *et al*., 2005). Nevertheless, maximizing the utility of genomic resources requires reciprocal exchange of information between model systems and crop species (Gepts *et al*., 2005). In view of these considerations, the present study was designed to evaluate the extent of cross-genera transferability of genic SSR markers within the Papilionoidea subfamily of Fabaceae, with the goal of facilitating their application across diverse *Cajanus* species and other agriculturally important legumes.

MATERIAL AND METHODS

Plant material: Major legume crops *Cajanus* species and other, including pea, chickpea, mungbean, cowpea, soybean, lentil, and groundnut.

Materials and Methods

Plant Material and DNA Extraction: Fresh or newly expanded leaves collected from 25 genotypes were used for genomic DNA isolation. For that, total genomic DNA was extracted using the standard protocol of cetyltrimethylammonium bromide (CTAB) described by Murray and Thompson (1980). The purity of the extracted DNA was evaluated by measuring absorbance at 260/280 nm using a spectrophotometer, while DNA concentration was estimated by agarose gel electrophoresis in comparison with DNA samples of known concentrations.

PCR Amplification: Polymerase chain reaction (PCR) amplifications were performed in a total reaction volume of 10 µL using a PTC-225 Gradient Thermal Cycler (MJ Research). Each reaction mixture contained 20 ng of genomic DNA, 1× PCR buffer (20 mM Tris–HCl and 50 mM KCl), 2.5 mM $MgCl_2$, 0.125 mM of each deoxynucleotide triphosphate (dNTP), 0.75 µM of each forward and reverse primer, and 0.5 U of Taq DNA polymerase. Amplification was carried out using a touchdown PCR protocol. The thermal cycling conditions included an initial denaturation step at 94°C for 3 minutes, followed by 5 cycles of denaturation at 94°C for 20 seconds, annealing starting at 60°C and decreasing by 1°C per cycle for 20 seconds, and extension at 72°C for 30 seconds. This was followed by 40 cycles consisting of denaturation at 94°C for 20 seconds, annealing at 56°C for 20 seconds, and extension at 72°C for 30 seconds. A final extension step was carried out at 72°C for 10 minutes.

Gel Electrophoresis and Fragment Analysis: PCR products were separated on 4% Metaphor agarose gels (Lonza, Rockland, ME, USA), depending on the expected fragment size. The gels were stained with ethidium bromide and visualized using a gel documentation system (Alpha Imager 2200, Alpha Innotech Corp., USA). Fragment sizes were estimated by comparison with a GeneRuler 100 bp DNA ladder (Hi Media)

RESULTS

Identification and Distribution of SSR Markers: A genome-wide study of microsatellite loci across selected legume species resulted in the identification of a substantial number of SSR markers, confirming their high abundance and genome-wide distribution. A total of ~1000 SSR loci were detected from genomic and transcriptomic datasets, indicating their suitability for marker development.

Among the identified motifs, di-nucleotide repeats (42%) and tri-nucleotide repeats (35%) were the most prevalent, followed by tetra- (12%), penta- (6%), and hexa-nucleotide (5%) repeats (Table 1). Tri-nucleotide repeats were predominantly associated with coding regions, whereas di-nucleotide repeats were more abundant in non-coding regions. The predominance of tri-nucleotide repeats in coding sequences reflects selective pressure to maintain translational reading frames. The most frequent motifs identified were $(AG/CT)_n$ among di-nucleotides and $(AAG/CTT)_n$ among tri-nucleotides. Significant variation in SSR frequency and motif distribution was observed among species, suggesting genome-specific patterns of microsatellite evolution. The repeat lengths varied considerably, ranging from short motifs to highly expanded repeats, indicating their potential as highly polymorphic genetic markers.

Table 1. Distribution of SSR Motifs in Legume Crops

SSR Type	Number of SSRs	Percentage (%)
Di-nucleotide	426	42.6
Tri-nucleotide	347	34.7
Tetra-nucleotide	12.5	12.5
Penta-nucleotide	63	6.3
Hexa-nucleotide	54	5.4

The most frequent motifs identified were (AG/CT)n among di-nucleotides and (AAG/CTT)n among tri-nucleotides.

Marker analysis: A subset of SSR loci (n = 100 markers) was selected for primer design and experimental validation. The amplification success rate was high (>90%), indicating the robustness of primer design and conservation of flanking regions. The majority of markers produced clear, distinct, and reproducible banding patterns, confirming their reliability. Polymorphism analysis revealed substantial allelic diversity, with the number of alleles per locus ranging from 2 to 8 (mean = 4.5). The polymorphic information content (PIC) values

ranged from 0.25 to 0.85, with an average of 0.62, indicating moderate to high informativeness. Markers with PIC values greater than 0.60 were considered highly informative and suitable for genetic diversity and mapping studies.

Table 2. Polymorphism statistics of SSR markers

Parameter	Range	Mean
Alleles per locus	2–8	4.5
PIC	0.25–0.85	0.62
Expected heterozygosity (He)	0.30–0.88	0.65

Cross-Species Transferability of SSR Markers: The transferability of SSR markers was evaluated across multiple legume species to determine their cross-species applicability.
A high transferability rate (~90%) was observed within closely related species, whereas the rate declined to ~65–70% across genera, indicating a clear influence of phylogenetic distance. Despite this decline, a significant proportion of markers remained transferable across diverse taxa, highlighting the conserved nature of genic regions. These results demonstrate the utility of genic-SSR markers for comparative genomics and cross-species genetic studies.

Genetic Diversity and Population Structure: Genetic diversity analysis based on SSR markers revealed substantial variation among genotypes. Cluster analysis using the UPGMA method grouped genotypes into two major clusters, with further sub-clustering corresponding to species origin and genetic background. Analysis of molecular variance (AMOVA) revealed that 72% of genetic variation was attributed to within-population differences, while 28% was due to variation among populations, indicating high intra-population diversity.

Table 3. AMOVA results

Source of Variation	Variation (%)
Among populations	28
Within populations	72

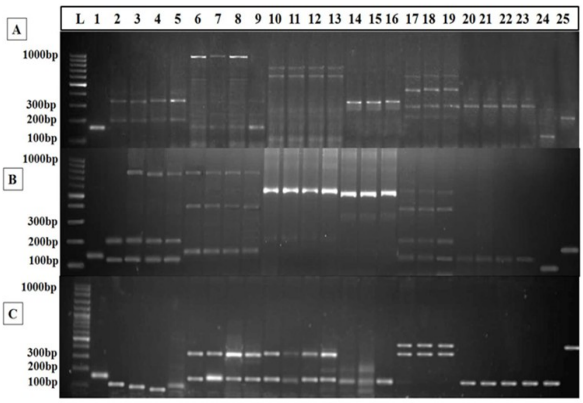

Fig. 1. Agarose gel electrophoretic profiles of PCR products from genic SSR markers: (A) ASSR19, (B) ASSR23, and (C) ASSR648. Lane 1: Pigeonpea (Asha); lanes 2–4: Field pea; lanes 6–9: Chickpea; lanes 14–16: Soybean; lanes 17–19: Cowpea; lanes 20–23: Mung bean; lanes 24–25: Groundnut. M = 100 bp DNA ladder

DISCUSSION

Comparative genomics and marker-assisted approaches have significantly enhanced our understanding of genome organization, evolution, and diversity among plant species. In this context, comparative mapping has emerged as a powerful strategy for integrating genetic information across related taxa and for identifying conserved chromosomal segments (Nadeau and Sankoff, 1998; Paterson et al., 2000). The present study highlights the importance of genic simple sequence repeat (genic-SSR) markers as effective tools for cross-species analysis and comparative genomics in legumes. Genic-SSRs, being derived from expressed regions of the genome, are generally more conserved than genomic SSRs. This conservation enhances their potential for cross-species transferability and facilitates the development of orthologous markers. The findings of this study demonstrate that pigeonpea-derived genic-SSR markers exhibit substantial transferability across a wide range of related legume species and genera (Gutierrez, et al.,2005). This observation aligns with the fundamental concept that coding regions evolve more slowly than non-coding regions, thereby maintaining sequence homology across taxa. Recent advances in genomics further support the utility of SSR markers in cross-species applications. For instance, large-scale genome mining efforts have revealed millions of SSR loci across multiple legume genomes, emphasizing their abundance and widespread distribution. Similarly, next-generation sequencing (NGS)-based studies have facilitated the rapid identification of thousands of SSR markers, significantly expanding the available marker resources for legumes. These technological developments have strengthened the role of SSR markers in genetic mapping, diversity studies, and crop improvement programs (Singh et al.2011, Gupta, et al., 2003). The high transferability observed in the present study is consistent with recent reports in legume crops. A study on Vigna species demonstrated transferability rates exceeding 90% for SSR markers derived from whole-genome sequences, indicating strong conservation within the genus. In contrast, transferability across more distantly related genera tends to be lower, as reported in multi-species studies where cross-genus amplification ranged around 30%. These findings suggest that while genic-SSRs are highly transferable, their efficiency is influenced by phylogenetic distance—a trend also observed in the present investigation

The predominance of di-nucleotide and tri-nucleotide repeats identified in this study is in agreement with earlier and recent genomic analyses. For example, studies in lentil and other legumes have reported that di-nucleotide repeats are the most abundant, followed by tri-nucleotide repeats. Tri-nucleotide repeats are particularly enriched in coding regions due to their compatibility with the triplet codon structure of genes, which prevents frameshift mutations. This distribution pattern further supports the functional significance of genic-SSR markers and their evolutionary conservation. Polymorphism analysis in the present study revealed moderate to high levels of allelic variation and polymorphic information content (PIC), indicating the effectiveness of SSR markers in detecting genetic diversity. These findings are consistent with recent studies that have reported PIC values ranging from low to high across different legume species, confirming the reliability of SSR markers for diversity analysis and population structure studies. The ability of SSR markers to detect multi-allelic variation makes them particularly valuable for distinguishing closely related genotypes and for assessing genetic relationships (Paterson *et al*., 2000). The results of AMOVA in this study indicated that the majority of genetic variation resides within populations rather than among populations. This pattern has been widely reported in plant genetic studies and reflects the presence of significant intra-population diversity.

Such diversity may arise from factors such as gene flow, mutation, and recombination. The observed genetic structure is advantageous for breeding programs, as it provides a broad genetic base for selection and improvement. Cluster analysis and population structure analysis further revealed distinct grouping patterns among the studied genotypes. The formation of clusters corresponding to species or genetic backgrounds suggests that the SSR markers used in this study effectively capture genetic relationships. Similar observations have been reported in recent legume studies, where SSR-based clustering and Bayesian analysis successfully differentiated populations while also detecting admixture events. Another important observation in this study is the high quality and consistency of amplification profiles obtained using genic-SSR markers. The production of clear and reproducible bands, along with occasional larger-than-expected amplicon sizes, is a characteristic feature of EST-derived markers.

This phenomenon may be attributed to the presence of introns or insertions in genomic DNA that are absent in the corresponding expressed sequences. Such features enhance the informativeness of these markers and contribute to their utility in genetic analysis. Recent developments in genomic technologies have further expanded the scope of SSR marker applications. The integration of whole-genome sequencing, bioinformatics tools, and SSR databases has enabled the identification and validation of large numbers of markers in a cost-effective and time-efficient manner. For example, comprehensive databases containing millions of SSR loci across multiple legume species have been developed to facilitate marker discovery and selection. These resources provide valuable platforms for comparative genomics and molecular breeding (Brondani *et al*., 1998). Despite the emergence of single nucleotide polymorphism (SNP) markers and high-throughput genotyping technologies, SSR markers continue to hold significant relevance, particularly in laboratories with limited resources (Gupta, *et al*., 2003). Their ease of use, cost-effectiveness, and high reproducibility make them accessible tools for a wide range of applications, including germplasm characterization, genetic diversity analysis, and marker-assisted selection. Furthermore, SSR markers are highly informative for studying genetic relationships and evolutionary patterns among closely related species. The findings of this study have important implications for legume improvement programs. The identification of cross-transferable SSR markers provides a valuable resource for genetic mapping and marker-assisted breeding. These markers can be used to identify genomic regions associated with important agronomic traits, such as yield, disease resistance, and stress tolerance. Moreover, the ability to transfer markers across species reduces the need for de novo marker development, thereby saving time and resources (Doyle and Luckow 2003). The Future research should focus on integrating SSR markers with high-throughput genomic technologies to further enhance their utility and to accelerate the development of improved legume varieties.

REFERENCES

Brondani, C., Rangel, P.H.N., Brondani, R.P.V., & Ferreira, M.E. (1998). Highly informative microsatellite markers for rice. *Theoretical and Applied Genetics*, 97, 488–495.

Choi, H.K., Mun, J.H., Kim, D.J., Zhu, H., Baek, J.M., Mudge, J., *et al*. (2004). Estimating genome conservation between crop and model legume species. *PNAS*, 101, 15289–15294.

Dayanandan, S., Bawa, K.S., &Kesseli, R. (1997). Conservation of microsatellites among tropical trees. *Proceedings of the National Academy of Sciences*, 94, 3124–3129.

Doyle, J.J., Doyle, J.L., Ballenger, J.A., Palmer, J.D. (1996). The distribution and phylogenetic significance of SSRs in plants. *Molecular Phylogenetics and Evolution*, 5, 249–259.

Doyle, J.J., Luckow, M.A. (2003). The rest of the iceberg: legume diversity. *Plant Physiology*, 131, 900–910.

Dutta, S., Kumawat, G., Singh, B.P., Gupta, D.K., Singh, S., Dogra, V., *et al*. (2011). Development of genic-SSR markers in pigeonpea. *BMC Plant Biology*, 11, 17.

Ellis, J.R., & Burke, J.M. (2007). EST-SSRs as molecular markers. *Molecular Ecology Resources*, 7, 379–390.

Fuller, D.Q. (2006). Agricultural origins and domestication in South Asia. *Journal of World Prehistory*, 20, 1–86.

Gaitán-Solís, E., Duque, M.C., Edwards, K.J., & Tohme, J. (2002). Development of SSR markers in common bean. *Genome*, 45, 43–50.

Gepts, P., Beavis, W.D., Brummer, E.C., Shoemaker, R.C., Stalker, H.T., Weeden, N.F., & Young, N.D. (2005). Legumes as model plant families. *Plant Physiology*, 137, 1228–1235.

Graham, P.H., & Vance, C.P. (2003). Legumes: importance and constraints. *Plant Physiology*, 131, 872–877.

Gupta, P.K., Varshney, R.K., Sharma, P.C., & Ramesh, B. (1999). Microsatellites in plants. *Current Science*, 77, 45–54.

Gupta, P.K., Rustgi, S., Sharma, S., Singh, R., Kumar, N., & Balyan, H.S. (2003). Transferability of SSR markers in cereals. *Theoretical and Applied Genetics*, 106, 912–922.

Gutierrez, M.V., Vaz Patto, M.C., Huguet, T., Cubero, J.I., & Moreno, M.T. (2005). Cross-species amplification of SSR markers. *Theoretical and Applied Genetics*, 110, 276–283.

Hamblin, M.T., & Di Rienzo, A. (2000). Detection of selective sweeps using microsatellites. *Genetics*, 155, 1855–1864.

Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., & Dhawan, A.K. (2011). Microsatellite markers: an overview. *Plant Cell Reports*, 30, 163–176.

Murray, M.G., & Thompson, W.F. (1980). Rapid isolation of plant DNA. *Nucleic Acids Research*, 8, 4321–4325.

Nadeau JH, Sankoff D (1998) Counting on comparative maps. *Trends in Genetics*, 14, 495-501

Odeny, D.A., Jayashree, B., Ferguson, M., Hoisington, D., Crouch, J., & Gebhardt, C. (2009). Development of SSR markers in pigeonpea. *Molecular Breeding*, 24, 399–411.

Parida, S.K., Kumar, K.A., Dalal, V., Singh, N.K., & Mohapatra, T. (2006). Unigene-derived SSR markers. *BMC Genomics*, 7, 177.

Paterson, A.H., Bowers, J.E., & Chapman, B.A. (2004). Ancient polyploidization in plants. *Nature*, 422, 433–438.

Peakall, R., & Smouse, P.E. (2012). GenAlEx 6.5 genetic analysis. *Bioinformatics*, 28, 2537–2539.

Powell, W., Machray, G.C., & Provan, J. (1996). Polymorphism in SSR markers. *Trends in Plant Science*, 1, 215–222.

Raju, N.L., Gnanesh, B.N., Lekha, P., Jayashree, B., Pande, S., Hiremath, P.J., *et al*. (2010). The first set of EST resource in pigeonpea. *Plant Biotechnology Journal*, 8, 436–445.

Rohlf, F.J. (2000). NTSYS-pc numerical taxonomy software. *Applied Biostatistics Inc.*, NY.

Saha, M.C., Mian, M.A.R., Eujayl, I., Zwonitzer, J.C., Wang, L., & May, G.D. (2004). SSR markers in forage crops. *Theoretical and Applied Genetics*, 109, 577–585.

Singh, N.K., Gupta, D.K., Jayaswal, P.K., Mahato, A.K., Dutta, S., Singh, S., *et al*. (2011). Pigeonpea genome analysis. *Nature Biotechnology*, 30, 83–89.

Singh NK, Gupta DK, Jayaswal P K, Mahato AK, Dutta S, Singh S, Bhutani S *et al*. (2011) The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol*, DOI 10.1007/s13562-011-0088-8.

Slatkin, M. (1995). Genetic differentiation and gene flow. *Genetics*, 139, 457–462.

Smouse, P.E., Peakall, R., & Gonzales, E. (2008). GenAlEx software. *Molecular Ecology Notes*, 8, 567–569.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6 molecular analysis. *Molecular Biology and Evolution*, 30, 2725–2729.

Thiel, T., Michalek, W., Varshney, R.K., & Graner, A. (2003). EST-derived SSR markers. *Theoretical and Applied Genetics*, 106, 411–422.

Varshney, R.K., Graner, A., & Sorrells, M.E. (2005). SSR marker development trends. *Trends in Biotechnology*, 23, 48–55.

Varshney, R.K., Hoisington, D.A., & Tyagi, A.K. (2006). Advances in cereal genomics. *Plant Physiology*, 141, 20–29.

Varshney, R.K., Song, C., Saxena, R.K., Azam, S., Yu, S., Sharpe, A.G., *et al*. (2013). Draft genome sequence of pigeonpea. *Nature Biotechnology*, 30, 83–89.

Weeden, N.F., Muehlbauer, F.J., &Ladizinsky, G. (1992). Comparative genetics of legumes. *Annual Review of Genetics*, 26, 105–131.

Wang, M.L., Barkley, N.A., & Yu, J.K. (2009). SSR markers in crop improvement. *Plant Breeding Reviews*, 31, 1–37.

Yu, J.K., La Rota, M., Kantety, R.V., & Sorrells, M.E. (2004). EST-derived SSR markers. *Genome*, 47, 805–818.

Zhang, L.Y., Bernard, M., Leroy, P., Feuillet, C., &Sourdille, P. (2005). High transferability of SSR markers. *Theoretical and Applied Genetics*, 111, 677–687.

Zane, L., Bargelloni, L., &Patarnello, T. (2002). Microsatellite markers: review. *Molecular Ecology*, 11, 1–16.

****