Original Research Article

Genetic Diversity of Fifty Coffee Genotypes using Simple Sequence Repeat (SSR) Markers.

ABSTRACT

Coffee plays an important role in the global economy, human health and society. It is one of the major cash crops constituting the backbone of the Nigerian economy before the emergence and re-emergence of petroleum oil.Molecular markers are reliable and valuable tool for analyzing polymorphism, screening and selection of desirable genotypes in breeding programs. The study aimed at determining the genetic diversity among coffee genotypes.Fifty genotypes of coffee (*coffeaspp*)collected from CRIN germplasm were assessed using 12 simple sequence repeat (SSR) markers. ~~The~~ Findings revealed thatpolymorphic information content (PIC) ranged between 0.311and 0.882 with an average of 0.545.~~However, the~~ The gene diversity (0.602) was high andthe unweighted pair group method with arithmeticUPGMA dendrogram grouped the genotypes into four ~~major~~ mainclusters with subclusters.The results revealed high levels of polymorphism indicating genetic diversity and confirm uniqueness among the coffee genotypes.

Keywords: Germplasm; Genotypes; Polymorphic; Dendrogram; Clusters

Comment [p1]: The abstract should be revised to capture the following:
Introduction (aim of the study)
Methodology
Key findings
Conclusion

Comment [p2]: The species studied should be stated

Comment [p3]: This should be written in full and the acronym written in bracket.

Comment [p4]: This should be written in full and the acronym written in bracket.

1. INTRODUCTION

Coffee is the world's second most popular beverage after water, with global commerce exceeding US$10 billion [1, 2].In actuality, Coffee is a very healthystimulant that should be consumed daily in amounts of at least 3 to 4 cups. It has healthy nutrients and antioxidants that help us feel better overall. Its primary effects include stimulating the circulatory, respiratory, and central neurological systems, acting as a diuretic and delaying weariness [3~~, 4, 5, 6,~~ 7~~,~~]. Regular coffee use has been linked to a lower risk of mortality and chronic diseases like cancer [3, 8]. Due to the presence of a variety of biological substances including caffeine, diterpenes, caffeic acid, polyphenols as well as volatile aroma and heterocyclic substances,the beneficial effects of coffee are supported by a number of conceivable processes [Ref??]. Recent research indicates that coffee consumption is unrelated to prostate, pancreatic, and ovary cancers but is

linked to a lower risk of liver, kidney, premenopausal breast and colorectal cancers [Ref??]. Drinking coffee might still lower the death rate from liver cancer [3, 9]. ~~There~~ Even though~~aren't~~ coffee lacksmany nutrients~~ in coffee~~, ~~but~~ however, the crop possesses~~it does include~~ thousands of naturally occurring substances, including phenolic compounds, minerals, lipids, and potentially healthy amounts of carbohydrates and fats [3]. Chorogenic acids, feruloyl quinic acids, various iridoid glycosides, isomers of monoester, diester caffeoylquinic acids, and dioxoanthracene [10, 11] are among the several key substances found in coffee that have therapeutic use. It is essential to the neurological, dermatological, nervous, and gastrointestinal systems [12, 13]. Additionally, it lessens memory loss, type-2 diabetes and protect from Alzheimer's and Parkinson's disease [10, 14].

Despite its importance, Nigeria has had a long-term fall in coffee production which is blamed on a variety of issues, one of which being the ~~absence~~ lackof ~~better~~ improvedplanting materials and more advanced technologies. The goal of the coffee breeding programme is to ~~create~~ developcultivars with ~~the~~ desiredagronomic and ~~technological~~ qualitycharacteristics that growers and other end-users in the coffee product value chain ~~want~~need, together with great potential for productivity, adaptability to various producing regions, and superior cup quality [15]. In breeding operations, evaluating genetic variance is crucial, utilization requires a thorough awareness of the variety of accessible germplasm, the formulation of suitable breeding techniques for the creation of enhanced planting materials can be guided by accurate knowledge of the diversity.

Molecular markers are extremely helpfulfor genetic improvement, help to monitor the level of genetic diversity and thereby enhance selection of breeding materials.~~A study using SSR marker was conducted by~~ Ahmed et al.[16]~~to study~~ studiedthe molecular characterization and validation of eighty sixgenotypes of sunflower (*Helianthus annuus* L.) hybrids and reported the applicability of simple sequence repeats (SSR)~~SSR~~ markers in sunflower for hybrid authentication. Khatab et al.[17] reported genetic diversity of the twenty four genotypes of soybean using ~~simple sequence repeats (~~SSR~~)~~ marker and revealed that SSR markers are efficient for measuring genetic diversity and relatedness as well as identifying varieties of soybeans.Considering the nutritional, medicinal, economic and scientific importance of *Coffea*, assessment of the extent of genetic variation in *Coffea species* is essential and highly significant for its improvement.

Comments

The introduction needs to be improved. It should be re-organized such that each paragraph should cnstitute a sentence topic followed by sentences of similar thoughts.

Formatted: Font: (Default) Arial, 12 pt, Portuguese (Brazil)

Formatted: Font: (Default) Arial

The literature on the key words in the title such as genetic diversity is not well articulated. Past efforts on morphological and molecular marker studies on the crop should be included.

Moreover, the introduction lacks objective(s).

Formatted: Font: (Default) Arial, 12 pt

2.MATERIALS AND METHODS

2.1. Sample collection

Fifty genotypes of *Coffeaspp* in a germplasm collection at Cocoa Research Institute of Nigeria, Ibadan, Nigeria, were used in this study. Fresh young leaf samples of each of the selected fifty genotypes of *coffea* materials were harvested into well labelled and tightly covered sample bags. The samples were placed on ice pack and immediately conveyed to the biotechnology laboratory, Where? for DNA extraction and genetic profiling using Simple Sequence Repeat Marker (SSR) procedure.

Comment [p5]: How many species studied? How many genotypes per species studied? Any GIS info on collection sites? No information on germplasm status utilized in your study.
Were the collected germplasm established at Cocoa Research Institute and leaf samples collected from them or leaf samples were collected on coffee farmers' fields?
If the genotypes were established at the experimental site, which experimental design was utilized?
What were the environmental (climatic) conditions that prevail in the experimental area.
When were the genotypes established in the field?

Table 1: List of the genotypes used in the study

S/N	Genotypes	Collection site
1	A81	CRIN Ibadan
2	C36	CRIN Ibadan
3	C96	CRIN Ibadan
4	C105	CRIN Ibadan
5	C107	CRIN Ibadan
6	C108	CRIN Ibadan
7	C111	CRIN Ibadan
8	D57	CRIN Ibadan
9	E1	CRIN Ibadan
10	E106	CRIN Ibadan
11	M10	CRIN Ibadan
12	M53	CRIN Ibadan
13	H139	CRIN Ibadan
14	T24	CRIN Ibadan
15	T204	CRIN Ibadan
16	T921	CRIN Ibadan
17	T1049	CRIN Ibadan
18	W109	CRIN Ibadan
19	A118	CRIN Ibadan
20	TG181	CRIN Ibeku
21	TG405	CRIN Ibeku
22	TG107	CRIN Ibeku

23	TG149	CRIN Ibeku
24	TG468	CRIN Ibeku
25	TG375	CRIN Ibeku
26	TG211	CRIN Ibeku
27	TG216	CRIN Ibeku
28	TG202	CRIN Ibeku
29	TG126	CRIN Ibeku
30	CPR	CRIN Mambilla
31	TH-F1-12-2	CRIN Mambilla
32	T992	CRIN Mambilla
33	PORTO	CRIN Mambilla
34	T1997	CRIN Mambilla
35	TH-F1-5-1	CRIN Mambilla
36	TH-F1-18-1	CRIN Mambilla
37	TH-F1-34-1	CRIN Mambilla
38	T1999-6-1	CRIN Mambilla
39	T238	CRIN Mambilla
40	th-F1-32-2	CRIN Mambilla
41	th-F1-9-3	CRIN Mambilla
42	TH-F1-4-1	CRIN Mambilla
43	T2255	CRIN Mambilla
44	T2000	CRIN Mambilla
45	T990	CRIN Mambilla
46	Nicar	CRIN Mambilla
47	T971	CRIN Mambilla
48	T977	CRIN Mambilla
49	G202	CRIN Mambilla
50	T1996	CRIN Mambilla

CRIN = Cocoa Research Institute of Nigeria

2.2. DNA Extraction

DNA was isolated from Intermediate leaves using CTAB (Cetyl trimethyl ammonium bromide) method of extractionwith slight modification. The leaves were collected, put into lyophilized bags and placed on ices sequel to transportation to the lab, .the The samples were stored in a freezer at -80° C for 4 days and it was lyophilized at usingthe lyophilization machine for 5 days. The lyophilized samples were grinded using the genogrinder at 3500 rpm for 2 minutes. Similarly, *in vitro* plantlets grown in Murashige and Skoog's (MS) medium (1962) were carefully removed from the test tubes and the leaves of the plantlets of each accessionswere wereplaced in a mortar with a volume of 5ml of liquid nitrogen for grinding.1ml of Hepes buffer was added and centrifuged at

Comment [p6]: What do you mean by intermediate leaves? Did you utilize young healthy leaves for your study?

3500 rpm for 13 minutes and supernatant decanted. 400 µl of CethylTrimetyl Ammonium Bromide (CTAB) and 10 µl of proteinase K were added and incubated in the water bath for at 65°C for 1 hour, and at 10-minute intervals the samples were checked and shaken to ensure proper homogenization. Thereafter 600 µl of Chloroform Isoamyl Alcohol (CIA) 24:1 was added and centrifuged at 3500rpm for 13 minutes. The aqueous phase was carefully transferred into another set of well-labelled extraction tubes and 500 µl of CIA added and centrifuged at 3500rpm for 13minutes. Then 300 µl of cold isopropanol and 50 µl of Nacl were added and shaken properly and incubated at -80°C for 1hour. The samples were centrifuged at 3500 rpm for 13 minutes and supernatant was decanted to obtain the pellet of DNA. Afterwards 500 µl of 70% ethanol was added and centrifuged for 10 minutes at 3500 rpm. The DNA pellets were dried for 30 minutes and 50 µl of low salt TE and 10 µl of RNase were added to suspend the DNA, and incubated at 4°C for 1 hour and later stored at -20°C.

2.3. Polymerase Chain Reaction

PCR was conducted using 10 µl volume in a 96 well micro titer plate with an automated thermal cycler. The reaction volume was 3 µl of DNA template, 2.54 µl of autoclaved distilled water, 1µl of 10x reaction buffer, 0.6 µl of 2.5 mM dNtps, 0.8 µl of 50 mM $MgCl_2$, 1µl of primer forward, 1 µl of primer reverse, 0.06 µl of taq polymerase enzymes. The PCR cycles consisted of denaturation at 94.0°C for 3 minutes, followed by 10 cycles of 94.0°C for 0.30 seconds at 60.0°C for 0.30 seconds and 72.0°C for 1.0 minutes and followed by 25 cycles at 94.0°C for 0.30 seconds, and at 50°C for 0.30 seconds the final extension step at 72.0°C for 15 minutes.

Formatted: Subscript

2.4. Data acquisition and analysis

The alleles were scored in binary codes; clearly visible bands were assigned 1 for presence, 0 for absence and m for missing data. Percentage of polymorphism was calculated as the proportion of polymorphic bands over the total number of bands [18]. The Polymorphic Information Content (PIC), a measure of variability for each locus was calculated across the assay of units by the formula of [19]:

PIC*i*= 2*fi*(1 - *fi*)

where*fi* is the frequency of the amplified allele (band present), and (1 – *fi*) is the frequency of the null allele (band absent) of marker *i*.

To estimate the level of genetic diversity, genetic similarities were evaluated using Nei and Li/Dice similarity index [20] with the aid of the NTSYSpc software, version 2.11 [21]. A dendrogram was generated from the similarity matrix using the UPGMA (Unweighted Pair Group Method of Analysis using arithmetic averages) in NTSYSpc program.

Comments

How many species studied? How many genotypes per species studied? Any GIS info on collection sites? No information on germplasm status utilized in your study.

Were the collected germplasm established at Cocoa Research Institute and leaf samples collected from them or leaf samples were collected on coffee farmers' fields?

If the genotypes were established at the experimental site, which experimental design was utilized?

What were the environmental (climatic) conditions that prevail in the experimental area.

When were the genotypes established in the field?

The authors could not well articulate the molecular markers in the materials and methods of of their study. They should recast the Polymerase chain reactions sub title as 'Molecular markers and polymerase chain reactions.'

Formatted: Font: (Default) Arial, 12 pt

3.RESULTS

Twelve SSR markers were used to assess genetic diversity of 50 coffee genotypes from the germplasms. A total of 83 alleles were recorded for the markers. An average of 6.92 alleles was observed per marker, which varied from 2 to 24 alleles (Table 2). The average allele frequency was 0.532 and varied from 0.260 to 0.740. Genetic diversity of the primer used ranged from 0.385 to 0.889, where DL456 revealed the highest genetic diversity and an average of 0.602 was observed for gene diversity (Table 2). Average polymorphic information content (PIC) value was 0.545 and rangedfrom 0.311(DL355) to 0.882(DL456)(Table 2).

UPGMA dendrogram generated from pooled data (Fig. 1) grouped the genotypes into four distinct clusters at 74% similarity level. The predominant cluster, IV contains 52% (26) of the genotypes (A81, C36, TG107, TG181, A118, C90, C96, C105, C107, C108, TG126, E106, M10, M53, T1049, TG149, T204, T921, TG375, TG216, C111, H139, D57, T24, G202, W109). Cluster I had five genotypes (T1997, TH-F1-34-1, TH-F1-5-1, T971, TH-F1-18-1), cluster II had eleven genotypes (E1, TH-F1-12-2, T1999-6-1, T992, PORTO, CPR, T238, th-F1-32-2, TG405, TG468, TG211) and the remaining eight genotypes formed cluster III (T977, TH-F1-4-1, T2255, th-F1-9-3, T2000, T990, Nicar, T1996) (Fig. 1).

Comment [p7]: Did the grouping or clustering also reflect species and collection site?

Table 2: Characteristics of polymorphic SSR markers used in molecular analysis of 50 *Coffee* genotypes

Markers	Major Allele	No of	Gene	PIC

Comment [p8]: All tables should be self-explanatory and drawn based on the guidelines of this journal house.

	Frequency	alleles	Diversity	
DL020	0.520	7	0.632	0.578
DL305	0.420	11	0.762	0.738
DL257	0.660	3	0.482	0.410
DL003	0.360	10	0.774	0.744
DL753	0.280	11	0.832	0.812
DL334	0.540	4	0.582	0.507
DL329	0.580	4	0.582	0.525
DL456	0.260	24	0.889	0.882
DL095	0.720	2	0.403	0.322
DL355	0.740	2	0.385	0.311
DL445	0.720	3	0.414	0.343
DL477	0.580	2	0.487	0.369
Mean	**0.532**	**6.92**	**0.602**	**0.545**

PIC=polymorphic information content

Formatted: Font: 12 pt

Comments

The key phenotypic and genetic attributes of genotypes in each cluster have not been mensioned.

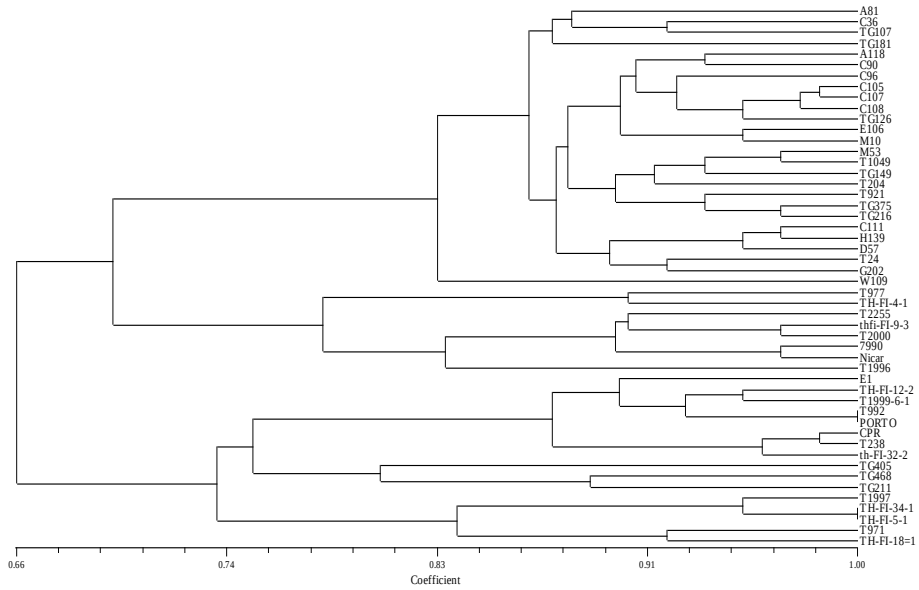

Figure 1: Molecular ~~Dendrogram~~ dendrogram showing genetic dissimilarity among the fifty genotypes of ~~*Coffee*~~coffeebased on SSR markers

Formatted: Font: Not Italic

.

4. DISCUSSION

Simple sequence repeat (SSR) technique is a PCR based method;SSR markers are highly polymorphic and are useful in studies on genetic diversity, phylogeny, gene tagging, genome mapping and evolutionary biology. SSRs have been successfully used to estimate the extent of genetic diversity level in a wide range of crop species including *Saccharum* spp[22], *Theobroma cacao* L[23],and*Solanum elaeagnifolium*[24]. The technique provided useful information for the exploitation of available genetic variability. The number of alleles and the high gene diversity (0.889) observed in this study proved that significant genetic variability occurs among the coffee genotypes. Botsteinet al. [25] reported that the Polymorphic information content (PIC) marker values more than 0.5 is considered as highly informative. The average PIC value of 0.545 obtained in the current study is highly informative suggesting that the SSR marker employed in the study was very useful for diversity study in ~~*Coffee*~~coffee.All the twelve polymorphic SSRs primers used in this study were found useful for the delineation of accessions showing high allelic variation DNA of the coffee genotypes. These observations are

Formatted: Font: Not Italic

Formatted: Font: Not Italic

Formatted: Font: Not Italic

similar to those reported by [22]where twelve primers were used to distinguish eight genotypes of *Saccharum* spplikewise a report by [23]where twenty primers were used to distinguish twenty seven accessions of *Theobroma cacao.* Also, a report by [26] shows that 18 primers were used to distinguish 20 wheat genotypes studied. A dendrogram produced by Unweighted Pair Group Method with Arithmetic Mean (UPGMA) grouped the germplasm into four distinct clusters. Based on the information from this dendrogram, genotypes that are far from each other by virtue of genetic origin and diversity index are strongly recommended to be used as parent for crossing. This will bring about greater genetic diversity, thus resulting into increase in selection gain.

Formatted: Font: Not Italic

Comment [p9]: A better understanding of genetic and phenotypic trait values is critical for selection of parents with complimentary traits for crossing.

5. CONCLUSION

The microsatellite (SSR) results obtained for the coffee genotype showed valuable diversity among the coffee genotypes studied. The methods divided the fifty coffee genotypes into several clusters and showed diversity within the grouping. Hence, this findings offer a foundation for further assessment and utilization of the available genetic diversity utilizing different marker system.

Comments
The authors should conclude on their objectives and demonstrate the mechanistic links among their findings.

REFERENCES

Comment [p10]: The authors should thoroughly revise and update the references based on the journal requirement.

[1] Iscaro, J., 2014. The Impact of climate change on coffee production in Colombia and Ethiopia. Global Majority E-Journal. 2014; 399 5(1):33-43.

[2] Mustopha, F. B., Adesanya, K. A., Aremu-Dele O., 2022. Production trend of coffee in Nigeria: A review. World Journal of Advanced Research and Reviews, 2022, 13(03), 137–146.

[3] Wachamo, H.L., 2017. Review on Health Benefit and Risk of Coffee Consumption. Med Aromat Plants 6: 301. doi:10.4172/2167-0412.1000301

[4] Oñatibia-astibia, A., Franco, R., Martínez-pinilla, E., 2017. Health benefits of methylxanthines in neurodegenerative diseases. Molecular Nutrition & Food Research.

[5] Harpaz, E., Tamir, S., Weinstein, A., Weinstein, Y., 2017. The effect of caffeine on energy balance. Journal of Basic and Clinical Physiology and Pharmacology 28: 1-10.

[6] Gan, Y., Wu, J., Zhang, S., Li, L., Cao S et al.2017. Association of coffee consumption with risk of colorectal cancer: a meta-analysis of prospective cohort studies. Oncotarget 8: 18699.

[7] Kelleni, M.T., 2017. Benefits of green coffee beans extract in health and obesity. Food ~~chemistry~~ Chemistry 64: 9663-9674.

[8] Nakagawa, S.H., Ito, H., Hosono, S., Oze, I., Tanaka H, et al. 2017. Coffee consumption and the risk of colorectal cancer by anatomical subsite in japan: results from the herpacc studies. International Journal of Cancer141: 298-308.

[9] Saab, S., Mallam, D., Cox, G.A., Tong, M.J., 2014. The impact of coffee on liver diseases: a systematic review. Liver International 34: 495-504.

[10] Banavath, J. N., Seong-Cheol, K., Ragula, S., Pinjari, A. B., Eun, Y.S. 2021 Coffee cultivation techniques, impact of climate change on coffee production, role of nanoparticles and molecular markers in coffee crop improvement, and challenges. J Plant Biotechnol (2021) 48:207–222 DOI:https://doi.org/10.5010/JPB.2021.48.4.207

[11] Banavath J.N, Seong-Cheol, K., Min Ju, S, Chun Whan, K., Chan Kyu, L., Hyun Joo, A., 2019. Responses to biotic and abiotic stresses and transgenic approaches in the coffee Plant J Korean Soc Int Agric 31:359-377

[12] Jae-HoonBae, J., Jae-Hyung, P., Seung, S., Dae-Kyu, S., 2014. Coffee and Health. Integr Med Res 3:189-191

[13] Patay, É. B., Bencsik, T., Papp, N., 2016. Phytochemical overview and medicinal importance of Coffea species from the past until now. Asian Pac J Trop Med 9:1127-1135

[14] Juliana, D.P., Adriana, F., 2019. Caffeine consumption through coffee: content in the beverage, metabolism, health benefits and risks. Beverages 5:1-51

[15] De Paiva Barbosa, I., da Costa, W. G., Nascimento, M., Cruz, C. D., and de Oliveira, A. C. B. (2019a). Recommendation of *Coffea arabica* genotypes by factor analysis. Euphytica 215:178. doi: 10.1007/s10681-019-2499-x

{16] Ahmed, HGM-D., Rizwan, M., Naeem, M., Khan, M. A., Baloch F.S., Sun S, et al., 2022 Molecular characterization and validation of sunflower (*Helianthus annuus* L.) hybrids through SSR markers. PLoS ONE 17(5): e0267383. https://doi.org/10.1371/journal.pone.0267383

[17] Khatab, I. A., Morsy, A. R., Fares, W. M., Kumamaru, T., 2016. Genetic ~~Diversity~~ diversity of ~~Soybean~~ soybean ~~Genotypes~~ genotypes revealed by ~~Agro~~agro–

morphological and SSR markers. J. Fac. Agr., Kyushu Univ., 61 (1), 79–87 (2016)

[18] Martos, V., Royo C., Rharrabti Y., Garcia del Moral, L.F., 2005 Using AFLPs to ~~Determine~~ determine ~~Phylogenetic~~ phylogenetic ~~Relationships~~ relationships and ~~Genetic~~ genetic ~~Erosion~~ erosion in ~~Durum~~ durum ~~Wheat~~ wheat ~~Cultivars~~ cultivars ~~Released~~ released in Italy and Spain throughout the 20th Century. Field Crop Res. 2005;91:107-116.

[19] Roldán-Ruiz, I., Dendauw J., Van Bockstaele, E., Depicker, A., De Loose M., 2000. AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Molecular Breeding. 2000; 6:125–134

[20] Nei M., Li W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academic of Science, USA.1979;76:3269-3273.

[21] Rohlf, F.J., 1998. NTSYSpc Numerical Taxonomy and Multivariate Analysis System.
Exeter Software, Setauket, New York.1998; 2.

[22] Khaled, K. A. M., Sheren, E. E., Abdelhadi A. A., 2017. Simple ~~Sequence~~ sequence ~~Repeats~~ repeats (SSRs) and ~~Morphological~~ morphological ~~Parameters~~ parameters associated with ~~Drought~~ drought ~~Tolerance~~ tolerance in ~~Sugarcane~~ sugarcane (*Saccharum*spp.)*J. Agric. Chem. and Biotechn., Mansoura Univ. Vol. 8(2): 85 - 89, 2017*

[23] Thondaiman, V, Rajamani, K., Senthil, N., Shoba, N., John A., 2013. Genetic diversity in cocoa (*Theobroma cacao* L.) plus trees in Tamil Nadu by simple sequence repeat (SSR) markers. Vol. 12(30), pp. 4747-4753, 24 July, 2013. DOI: 10.5897/AJB2013.12423

[24] Zhu, X. C., Wu, H. W., Raman, H., Lemerle, D., Stanton, R., Burrows, G. E., 2012. Evaluation of simple sequence repeat (SSR) markers from Solanum crop species for *Solanum elaeagnifolium*. Weed Research. DOI: 10.1111/j.1365-3180.2012.00908.x

[25] Botstein, D., White, R. L., Skolnick, M., Davis, R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet.1980;32:314-331.

[26] Singh, M., Vaishali, K., Lalit K., 2020. Diversity Analysis of Wheat Genotypes ~~Using~~ using SSR ~~Molecular~~ molecular ~~Markers~~markers. *Int.J.Curr.Microbiol.App.Sci*(2020)*9*(10): 1413-1423

Formatted: Font color: Red

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Formatted: Font: Not Italic

Formatted: Font color: Red

Comment [p11]: Which journal??

Formatted: Highlight

Comment [p12]: Lacks volume, issue and page numbers.

Formatted: Font color: Red

Formatted: Highlight