Isolation and characterization of microsatellite markers for the threatened African endemic tree species *Pterocarpus erinaceus* Poir.

Benziwa Nathalie Johnson[1], Marie Luce Akossiwoa Quashie[2], Gilles Chaix[3], Letizia Camus-Kulandaivelu[3], Kossi Adjonou[1], Kossi Novinyo Segla[1], Adzo Dzifa Kokutse[1], Kouami Kokou[1], and Hélène Vignes[3]

[1]Université du Lomé
[2]Universite du Lome
[3]CIRAD Departement Systemes biologiques

May 6, 2020

Abstract

To study the genetic diversity and structure of the forest species *Pterocarpus erinaceus* Poir., seventeen polymorphic nuclear microsatellite markers were isolated and characterized, using Illumina MiSeq sequencing technology. Three hundred and sixty five (365) individuals were analysed within fifteen (15) West Africa populations. The alleles' number for these loci varied from 4 to 30 and 0.23 to 0.82 for the heterozygosity. The seventeen primers designed here will be useful to analyse ecology population and mechanisms of population differentiation of this threaten species.

Isolation and characterization of microsatellite markers for the threatened African endemic tree species *Pterocarpus erinaceus*Poir.

Benziwa Nathalie Johnson*[1], Marie Luce Akossiwoa Quashie[1], Gilles Chaix[2,3], Letizia Camus-Kulandaivelu[2,3], Kossi Adjonou[1], Kossi Novinyo Segla[1], Adzo Dzifa Kokutse[1],Kouami Kokou[1], Hélène Vignes[2,3]

[1]*Laboratoire de Recherche Forestière (LRF), Université de Lomé, 01BP 1515, Lomé (Togo)*

[2]*CIRAD, UMR AGAP, F-34395 Montpellier, France*

[3]*AGAP, Univ Montpellier, CIRAD, INRAe, Institut Agro, Montpellier, France*

*Corresponding author:*benziwa.johnson@gmail.com*

Abstract

To study the genetic diversity and structure of the forest species*Pterocarpus erinaceus* Poir., seventeen polymorphic nuclear microsatellite markers were isolated and characterized, using Illumina MiSeq sequencing technology. Three hundred and sixty five (365) individuals were analysed within fifteen (15) West Africa populations. The alleles' number for these loci varied from 4 to 30 and 0.23 to 0.82 for the heterozygosity. The seventeen primers designed here will be useful to analyse ecology population and mechanisms of population differentiation of this threaten species.

Keywords: African tree species, genetic diversity, next-generation sequencing, nuclear microsatellites, *Pterocarpus erinaceus* .

Introduction

Pterocarpus erinaceus Poir. (1823) commonly known as African rosewood is an important tree belonging to the Fabaceae family and native to the Guinean forest-savannah mosaic ecoregion and reported from Senegal to Cameroun (Adjonou et al., 2019; Arbonnier, 2004; Giffard, 1974). Stands of this Leguminous species are especially targeted for timber, service wood fuelwood production and have several medicinal uses (Fontodji et al., 2011; Kokou et al., 2009; Segla et al., 2015). Recent scientific investigations on *P. erinaceus* international trade have highlighted a considerable increase in export volume for its wood furniture from West Africa countries for Asia, particularly China (Dumenu, 2019; Lawson, 2015). As a consequence of its overexploitation, the Convention on International Trade in Endangered Species of Wild Fauna and Flora classified the species as threatened (CITES, 2016) and it has become the focus of conservation concern in African countries. Because of its high-quality wood, but also as a drought and fire-resistant plant species with traditional medicine uses in sub Saharan Africa (Duvall, 2008; Karou et al., 2003; Ouedraogo et al., 2012), *P. erinaceus* is a good model species for the study of a sustainable genetic diversity in *Pterocarpus* genus.

To understand the evolution dynamic of population *P. erinaceus*populations in West Africa in order to establish appropriate and efficient production and conservation strategies, it is necessary to study genetic diversity and structure of its natural stands dynamics. The study also wanted at analyzing the characteristics of mating system existing among individuals and the gene flow. Among various molecular tools used for assess plant genetic diversity, microsatellite simple sequence repeats (SSR) markers are the most widely employed because they are codominant and they possess high levels of polymorphism and stability (De et al., 2017; Morgante and Olivieri, 1993).

In the context of assessing genetic diversity into *Pterocarpus*genus, Muller et al. (2006) have developed microsatellite markers for*Pterocarpus officinalis* Jacq. belonging to the Carribean wetland forest species, but especially for *P. erinaceus* no nuclear microsatellite markers have been identified so far. Owing to the laborious and expensive microsatellites development by conventional methods, next-generation sequencing (NGS) technologies were carried out here. The major advantage of this approach is in generating important volume of sequencing data allowing the identification of large numbers of microsatellite markers (Rico et al., 2013; Senan et al., 2014; Vieira et al., 2016).

We developed here microsatellite markers for *P. erinaceus* and evaluated their polymorphism in West Africa.

Material and Methods

We sampled nine to thirty adult trees in fifteen populations (Table 1) with a total of three hundred sixty-five trees in four countries of West Africa which are Benin, Burkina Faso, Niger and Togo.

Table 1 : Characteristics of sampled sites for *Pterocarpus erinaceus* : country location, name of sampling sites, number of tree sampled per site.

Country	**Sampling sites**	**Number of tree sampled**
Niger	Parc W**	27
	Tamou**	14
Bénin	Koussoukpa*	25
	Bassila*	24
	Pénessoulou**	27
	Parc W	24
	Adakplamè**	27
	Houin*	24
Togo	Oti-Kéran**	25
	Fazao-Malfakassa*	30
	Abdoulaye**	30
	Plateau Akposso	23

	Togodo*	30
Burkina-Faso	Sarya**	9
	Tiogo**	25
Total	**Total**	**365**

**and ** = Geographic origin of individuals selected for genomic library construction*

*** = Geographic origin of individuals selected for the first screening*

Freshly collected leaves were dried in a coffee filter containing 10 g Silica gel. Each filter containing sample was put in a plastic zip bag for transport to the laboratory. Our genomic library was constructed using DNA of *P. erinaceus* samples from twelve randomly selected individuals among populations (Table 1).

DNA Extraction

Total genomic DNA extraction was performed with a solution of alkyltrimethylammonium bromide (MATAB) using ttwenty milligrams of dried leaves from each tree sample. Extraction protocol used derived from Bousquet et al. (1990) methodology.

The quality of the genomic extracted DNA was controlled on a 1% agarose gel, and quantification done by Hoescht assay using fluoroskan (Fluoroskan Microplate Fluorometer).

Construction of the DNA library and validation

The Westburg NGS DNA Library PrepKit was used to prepare the DNA bank with extract mixed DNA of twelve individuals. We started with 1µg of*Pterocarpus* mixed DNA. This DNA was fragmented, and ligated with Illumina adapters. Purification on magnetic beads (Agencourt AMPure XP beads -A63881-, Beckman Coulter) was performed before and after PCR. PCR amplification was run with 7 cycles.

The quality of DNA library was controlled using an Agilent 4200 TapeStation with a screen tape D5000. The size of the fragments were expected between 100 and 600 pb. The library was quantified using the Takara kit (638324) on a qPCR machine (LightCycler® 480 Real-Time PCR System, Roche Life Science).

Sequencing the DNA library

MiSeq system Illumina sequencer DNA was used to perform DNA sequencing on the genotyping platform at CIRAD-Montpellier. A 500 cycles NANO V2 cartridge Illumina (2 x 250 pb) was used to sequence the library.

Design and choice of primers

A total of 800,000 reads was generated for *P. erinaceus* samples. Development of optimised and streamlined microsatellites was based on the bioinformatics Galaxy pipeline and its following tools: FASTQ Groomer tool, Filter FASTQ tool and ABySS parallel assembler (Simpson et al., 2009). The MISA MIcroSAtellite identification tool (Thiel, 2003) and primer modelling software Primer3 (Whitehead Institute) were used for the identification and design microsatellites primers in nucleotide sequences generated. A data matrix containing all the microsatellite primers was obtained as output.

Among the 38,715 single sequence repeats identified, primers were designed for 11,718 sequence repeats of which 3,530 were dinucleotide repeats, 2,970 trinucleotide repeats, 2,847 tetranucleotide repeats, 1,001 pentanucleotide repeats, 525 hexanucleotide repeats, and 844 contained complex SSR motifs.

Dinucleotide and trinucleotide microsatellites motifs were selected for SSR-PCR amplification screening. These primers exhibited a minimum of five repetitions of the repeated motif and amplified fragments between 100 and 400 bp in length. Thirty suitable microsatellite were identified and selected for initial screening. This first test was performed on an ABI 3500XL sequencer (Life technologies, Carlsbad, California, USA) using genomic DNA extract from eight individuals selected from different countries (Table 1).

An M13 tailed primer (5'-CACGACGTTGTAAAACGAC-3'), allowing detection of fluorescence was added to the forward primers. Each PCR amplification was performed in a 96-well plate using 10-µL volume reaction containing 20 ng of DNA, 1X PCR buffer (without $MgCl_2$), 0.08 µM of the M13-labeled primer, 0.1 µM of each forward fluorescent primer (FAM, NED, PET, and VIC) and the reverse primer, 0.1 µM of M13 fluorescent primer, 2 mM of $MgCl_2$, 200 µM dNTPs, 0.4X Solution "S" (additive solution that facilitates amplification of difficult templates) and 0.05 U/µL of *Taq* DNA polymerase. PCR running conditions were: initial denaturation at 94°C for 4 min followed by 36 cycles each at 92°C in 30 s, 1 min at 52°C, 45 s at 72°C and with a final extension step at 72°C for 5 min.

Electropherograms were analysed and allele sizes were determined with GeneMapper® software version 4.1 using GeneScan 600 LIZ as a size standard (Applied Biosystems). Among the 30 primer pairs tested, 17 were selected. Indeed, we eliminated primers with profiles that were difficult to read on GeneMapper®, or with no or little polymorphism. The 17 selected primers are shown in Table 2 and they were used for screening of remaining individuals in order to calculate genetic parameters.

Genetic parameters including alleles' number per locus, observed (Ho) and expected (He) heterozygosity were computed using GenAIEx software version 6.0 (Peakall and Smouse, 2012). Deviation from the Hardy–Weinberg equilibrium (HWE) was measured for each locus by $\chi 2$ tests, and with a Bonferroni correction procedure (Rice, 1989), p-value significance was assessed in the context of multiple testing. Significant linkage disequilibrium was rated among these loci by using GENETIX software version 4.05 (Belkhir et al., 1996). MICRO-CHECKER software version 2.2.3 (Van Oosterhout et al., 2004) was used to check for the null alleles in microsatellite data.

Results and Discussion

A total of 237 alleles were identified for the 17 locus on the 365 characterized trees, with each locus having from 4 to 30 alleles (mean of 13.9 alleles per locus). Mean values for the expected heterozygosity (He) varied from 0.307 to 0.781 (0.571 ± 0.176) and from 0.234 to 0.821 (0.551 ± 0.190) for observed heterozygosity (Ho) for the individuals screened. Evidence of significant linkage disequilibrium, was found for 12 for 136 possible SSR pairwise combinations after Bonferroni corrections. Significant departures from Hardy–Weinberg equilibrium for 14 of the 17 loci was recorded after Bonferroni corrections and presence of null alleles was suggested for all loci excepted for mPeCIR_D2 and mPeCIR_T3. This set of 17 pairs of specific primers of*Pterocarpus erinaceus* would serve to study the genetic diversity of this species in West Africa.

Table 2 : Characteristics of 17 microsatellite primers designed for *Pterocarpus erinaceus* Poir.

Primer name	Primer sequences (5'-3')	Repeat motif	Allele size	TA°C	Na	Ho	He	p-value
mPeCIR_D1	F : TTTCTTCTACTTTCCTTTCCC R : AAGCAGGCTCAAGAGAAGA	$(CT)_{15}$	109-124	54.4	16	0.698	0.735	0.0000**
mPeCIR_D2	F : AACATGCAAGCAAAGCA R : AAGGTGGAGCTAAAGAAGGT	$(AG)_{13}$	107-123	54.6	12	0.736	0.672	0.0000**
mPeCIR_D4	F : TCGGTTTTGGTCTTTGTG R : CAGACCGTTGGGAAGAA	$(TC)_{14}$	152-167	55.4	16	0.821	0.781	0.0000**
mPeCIR_D5	F : TGTCCCGTGAAGAAAGG R : AAGCAGGCTCAAGAGAAGA	$(GA)_{10}$	102-159	55.3	10	0.428	0.385	0.0000**
mPeCIR_D7	F : CGTCAGCCTCCAATCTC R : CGCTTGATTTGGTCCTC	$(GA)_{14}$	189-203	54.9	20	0.686	0.676	0.0017**
mPeCIR_D8	F : CTCATGGGCACAGAACAA R : GATGGGCTTCACAGCAA	$(TA)_{11}$	177-205	56.4	30	0.706	0.749	0.0085**
mPeCIR_D9	F : TTTCCCGGTGTCAAGAA R : GACACACGCACATACAGAGA	$(TC)_{16}$	188-208	55.8	20	0.706	0.685	0.0017**
mPeCIR_D10	F : TCACCAAAACATGCACAA R : GCTCATGCTTAGCCCC	$(TG)_{14}$	214-230	55.1	11	0.464	0.52	0.0000**
mPeCIR_D11	F : GGGTTAGAGTTTGAATGGG	$(AG)_{17}$	221-239	54.5	22	0.753	0.784	0.0000**

Primer name	Primer sequences (5'-3')	Repeat motif	Allele size	TA°C	Na	Ho	He	p-value
	R : GCCTTCCTCAGCACTATTT							
mPeCIR_D12	F : AACCTGCCCATCCATTT	$(TC)_{16}$	238-253	56.1	10	0.522	0.538	0.0000**
	R : TACACTGGGTCGTTGGG							
mPeCIR_D14	F : CAGCACTGGCACCAAC	$(AG)_{13}$	280-307	55.1	29	0.758	0.778	0.0000**
	R : CACCACACCGCTTAATGT							
mPeCIR_T1	F : TCCATTGGGGTATCTATGTG	$(ATC)_6$	115-121	55.7	4	0.234	0.34	0.0000**
	R : CCTCAAGGGTGTTTTGTGT							
mPeCIR_T2	F : ATCACGGGCTCTTCCTC	$(TCT)_8$	121-130	56.0	9	0.379	0.432	0.0000**
	R : TCATTGTTTCTGCAAATCCT							
mPeCIR_T3	F : GGCCATTCTTCATGTGTTT	$(CTT)_8$	99-146	55.9	9	0.395	0.387	0.1139ns
	GGAGATGGGTGAGAGTGAA							
mPeCIR_T4	F : CAGGAGGGGTGGTGG	$(GAA)_6$	146-152	56.3	4	0.315	0.307	0.3077ns
	R : GCATCCTAGCCCGATTT							
mPeCIR_T5	F : AGACCCGAACTTGTCCC	$(TTA)_{11}$	145-167	55.7	11	0.478	0.61	0.4590ns
	R : TGCCAGTGTGTGATGGA							
mPeCIR_T15	F : CCCTCATCAAGAAGAACCA	$(ACA)_7$	277-295	56.0	4	0.295	0.332	0.0000**
	R : CTTGCATCACCACCCTC							

***Na**: number of individuals; **Ho**: observed heterozygosity; **He**: expected heterozygosity under Hardy–Weinberg equilibrium; **TA°C**: annealing temperature. p-values for the Hardy–Weinberg Equilibrium test, significance threshold adjusted using sequential Bonferroni correction: *p [?] 0.05, ***p [?] 0.001, ns = not significant .*

Conclusion

Illumina MiSeq sequencing technology enables us to develop 30 pairs of microsatellites primers based on *P. erinaceus* populations. Seventeen among them were polymorphic in fifteen (15) West African locations and they provide the first set of microsatellite markers dedicated for *P. erinaceus.* These microsatellite markers will be useful for the characterization of genetic diversity and analysis of genetic structure for *P. erinaceus* populations.

Data Accessibility

The data have been deposited with links to BioProject accession number PRJNA604893 in the NCBI BioProject database (*https://www.ncbi.nlm.nih.gov/bioproject/*)

17 Primers designed and evaluated: end of text (Table 2).

ORCID

Benziwa Nathalie Johnson ID:*https://orcid.org/0000-0003-0018-8085*

Conflicts of Interest

The authors have no conflicts of interest to declare.

Authors contributions

All authors contribute significantly to the present study and to the revision of the manuscript. Manuscript writing: B.N.J., with support from M.L.A.Q., G.C., L.C-K., H.V., A.D.K and K.K. DNA extractions, PCR running, DNA sequencing, microsatellites primers identification and selection: B.N.J. and H.V.

Genetic analyses and statistical treatments: B.N.J., G.C., L.C-K.

Coordination of data sampling: H.V. M.L.A.Q., K.A., A.D.K. and K.K.

Trees sampling in the field: B.N.J. and K.N.S.

Leading research project: M.L.A.Q.

Acknowledgements

This research was financially supported by: Sud Expert Plantes Developpement Durable (SEP2D) program [AAP1-45], Agence Universitaire de la Francophonie (AUF) [DRAO-1448-2019-02], Programme d'Appui a la Lutte contre le Changement Climatique : PALCC –AMCC+ [DCI-ENV/2016/39289 Devis programme multi annuel ndegDCI-ENV-2017/387-416], Service de Cooperation et d'Action Culturelle (SCAC) and Campus France, CIRAD for the "Actions Incitatives-Soutien aux doctorants" Program, and genotyping platform of Agap unit managed by Pierre Mournet.

References

Adjonou K., Houetchegnon T., Rabiou H., Kossi S.N., Abotsi E.K., Johnson B.N., Alaba P., Ouinsavi C.A.N., Quashie A.M.L., and Kokutse A.D. (2019). Challenges of Conservation and Sustainable Management of African Rosewood (*Pterocarpus erinaceus*) in West Africa. In *Natural Resources Management and Biological Sciences* , (Intech Open), 29 p.

Arbonnier M. (2004). *Arbres, arbustes et lianes des zones seches d'Afrique de l'Ouest* (Editions CIRAD, MARGRAF PUBLISHERS GMBH, MNHN). 579 p.

Belkhir K. Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. (1996). GENETIX 4.05, logiciel sous Windows TM pour la genetique des populations. http://www. Genetix.univ-montp2.fr/genetix/genetix.htm.

Bousquet J., Simon L., and Lalonde M. (1990). DNA amplification from vegetative and sexual tissues of trees using polymerase chain reaction.*Canadian Journal of Forest Research* **20** , 254–257.

CITES (2016). Convention sur le commerce international des especes de faune et de flore sauvages menacees d'extinction. *Dix-septieme session de la Conference des Parties Johannesburg* (Afrique du Sud).

De J., Zhu W., Liu T., Wang Z., and Zhong Y. (2017). Development of microsatellite markers using Illumina MiSeq sequencing to characterize*Ephedra gerardiana* (Ephedraceae) 1. *Appl Plant Sci* 5. 4 p.

Dumenu W.K. (2019). Assessing the impact of felling/export ban and CITES designation on exploitation of African rosewood (*Pterocarpus erinaceus*). *Biological Conservation* **236** , 124–133.

Duvall, C.S. (2008). *Pterocarpus erinaceus* Poir. [Internet] Record from PROTA4U. Louppe, D., Oteng-Amoako, AA & Brink, M. (Editors). PROTA (*Plant Resources of Tropical Africa/Ressources vegetales de l'Afrique tropicale*), Wageningen, Netherlands.

Fontodji, J.K., Atsri H., Adjonou K., Radji A.R., Kokutse A.D., Nuto Y., and Kokou K. (2011). Impact of charcoal production on biodiversity in Togo (West Africa). In *The Importance of Biological Interactions in the Study of Biodiversity* , (InTech), 215-230.

Giffard, P.L. (1974). L'Arbre dans le paysage senegalais: sylviculture en zone tropicale seche (GERDAT-Centre technique forestier tropical).431 p.

Karou D., Dicko M.H., Sanon S., Simpore J., and Traore A.S. (2003). Antimalarial activity of *Sida acuta* Burm. f. (Malvaceae) and*Pterocarpus erinaceus* Poir. (Fabaceae). *Journal of Ethnopharmacology* **89** , 291–294.

Kokou K., Nuto Y., and Honan A. (2009). Impact of charcoal production on woody plant species in West Africa: A case study in Togo.*Scientific Research and Essays* **4** , 881–893.

Lamarck, de Monnet de J.B.A.P., and Poiret J.L.M. (1823). Tableau encyclopedique et methodique des trois regnes de la nature.*Botanique* . (Paris : Panckoucke, 1791-1823.).

Lawson S. (2015). The illegal rosewood boom in West Africa; How Chinese demand is driving conflict, corruption and human rights abuses. In*Presentation to Chatham House Illegal Logging Stakeholder Update Meeting* , 25th June 2015.

Morgante M., and Olivieri A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. *The Plant Journal* **3** , 175–182.

Muller F., Vaillant A., Ba A., and Bouvet, J-M. (2006). Isolation and characterization of microsatellite markers in *Pterocarpus officinalis* Jacq. *Molecular Ecology Resources* **6** , 462–464.

Ouedraogo, N., Sawadogo, W., Tibiri, A., Bayet, C., Lompo, L., Hay E.A., Koudou J., Dijoux M-G., and Guissoux P.I. (2012). Pharmacological properties and related constituents of stem bark of *Pterocarpus erinaceus* Poir. (Fabaceae). *Asian Pacific Journal of Tropical Medicine* **5** , 46–51.

Peakall R., and Smouse P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28** , 2537–2539.

Rice W.R. (1989). Analyzing tables of statistical tests.*Evolution* **43** , 223–225.

Rico C., Normandeau E., Dion-Cote A.-M., Rico M.I., Cote G., and Bernatchez L. (2013). Combining next-generation sequencing and online databases for microsatellite development in non-model organisms.*Scientific Reports* **3** , 8 p.

Segla K.N., Adjonou K., Radji A.R., Kokutse A.D., Kokou K., Rabiou H., Kamana P., Bationo B.A., and Ali M. (2015). Importance socio-economique de *Pterocarpus erinaceus* Poir. au Togo. *European Scientific Journal* , ESJ **11** (23). 199-217.

Senan S., Kizhakayil D., Sasikumar B., and Sheeja T.E. (2014). Methods for Development of Microsatellite Markers: An Overview. *Notulae Scientia Biologicae* , **6** , 1–13.

Simpson J.T., Wong K., Jackman S.D., Schein J.E., Jones S.J., and Birol I. (2009). ABySS: a parallel assembler for short read sequence data.*Genome Research 19* , 1117–1123.

Thiel T. (2003). MISA—Microsatellite identification tool. Website Http://Pgrc. Ipk-Gatersleben. de/Misa/[Accessed 17 June 2016].

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P., and Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes***4** (3), 535–538.

Vieira M.L.C., Santini L., Diniz A.L., and Munhoz C. de F. (2016). Microsatellite markers: what they mean and why they are so useful.*Genetics and Molecular Biology* **39** , 312–328.

Hosted file

Johnson et al_Table 1.docx available at https://authorea.com/users/317262/articles/447391-isolation-and-characterization-of-microsatellite-markers-for-the-threatened-african-endemic-tree-species-pterocarpus-erinaceus-poir

Hosted file

Johnson et al_Table 2.docx available at https://authorea.com/users/317262/articles/447391-isolation-and-characterization-of-microsatellite-markers-for-the-threatened-african-endemic-tree-species-pterocarpus-erinaceus-poir