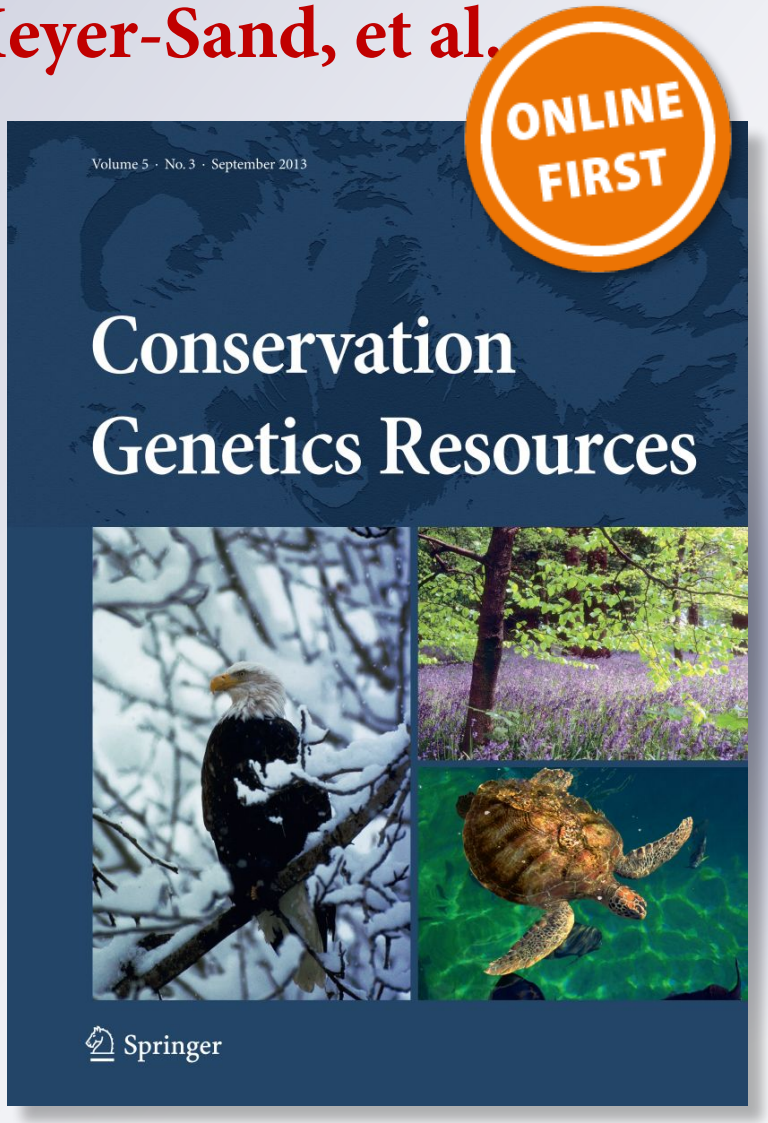
Development of nuclear and plastid SNP markers for genetic studies of Dipteryx *tree species in Amazonia*

Eurídice N. Honorio Coronado, Céline Blanc-Jolivet, Malte Mader, Carmen R. García-Dávila, Alexandre M. Sebbenn, Barbara R. V. Meyer-Sand, et al.

Conservation Genetics Resources

ISSN 1877-7252

Conservation Genet Resour
DOI 10.1007/s12686-019-01081-3



Springer

Conservation Genetics Resources
https://doi.org/10.1007/s12686-019-01081-3

TECHNICAL NOTE

Development of nuclear and plastid SNP markers for genetic studies of *Dipteryx* tree species in Amazonia

Eurídice N. Honorio Coronado[1] · Céline Blanc-Jolivet[2] · Malte Mader[2] · Carmen R. García-Dávila[1] · Alexandre M. Sebbenn[3] · Barbara R. V. Meyer-Sand[4] · Kathelyn Paredes-Villanueva[5] · Niklas Tysklind[6] · Valerie Troispoux[6] · Marie Massot[7] · Bernd Degen[2]

Received: 24 November 2018 / Accepted: 6 January 2019

Abstract

We developed nuclear and plastid single nucleotide polymorphism (SNP) and insertion/deletion (INDEL) markers for *Dipteryx* species using a combination of restriction associated DNA sequencing (RADSeq) and low coverage MiSeq genome sequencing. Of the total 315 loci genotyped using a MassARRAY platform, 292 loci were variable and polymorphic among the 73 sampled individuals from French Guiana, Brasil, Peru, and Bolivia. A final set of 56 nuclear SNPs, 26 chloroplast SNPs, 2 chloroplast INDELs, and 32 mitochondrial SNPs identifying significant population structure was developed. This set of loci will be useful for studies on population genetics of *Dipteryx* species in Amazonia.

Keywords Cumaru · Timber species · SNP · MassARRAY

Dipteryx Schreb. (Fabaceae) is a Neotropical genus of nine accepted tree species distributed across South America, Central America, and the Caribbean. *Dipteryx* timber is heavy, dense, and resistant to moisture and pest attack, and has been exploited heavily over the last decade in response to the increasing international demand of hardwood (Putzel et al. 2011). The seeds, known as tonka beans, are fragrant and valuable in the perfume industry. Two species, *D. odorata* (Aubl.) Willd and *D. charapilla* J.F. Macbr. Ducke are classified as vulnerable in the IUCN Red List of Threatened Species (IUCN 2018). Nuclear microsatellites (nSSRs) have been developed to study gene flow of *D. odorata* (Vinson et al. 2009) and *D. alata* Vog. (Soares et al. 2012). However, the amplification and scoring of microsatellite loci from timber material is problematic due to the low quality and quantity of DNA which can be recovered. New sets of single nucleotide polymorphisms (SNP) have been recently developed for valuable tropical timber species and have been successfully amplified in DNA of timber in South America (Chaves et al. 2018; Meyer-Sand et al. 2018) and Africa (Blanc-Jolivet et al. 2017; Jardine et al. 2016; Pakull et al. 2016). Here we present a new set of 315 nuclear and plastid SNPs tested on *Dipteryx* species from which we selected 100 loci to be used for the genetic characterization of species and population genetic analyses.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12686-019-01081-3) contains supplementary material, which is available to authorized users.

✉ Eurídice N. Honorio Coronado
eurihc@yahoo.com

[1] Instituto de Investigaciones de la Amazonía Peruana (IIAP), Av. José A. Quiñones km 2.5, Iquitos, Peru

[2] Thünen Institute of Forest Genetics, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany

[3] Instituto Florestal de São Paulo, CP 1322, São Paulo, SP 01059-970, Brazil

[4] Departamento de Fitotecnia, Faculdade de Engenharia de Ilha Solteira, Universidade Estadual Paulista, Ilha Solteira, SP, Brazil

[5] Carrera de Ingeniería Forestal, Laboratorio de Dendrocronología, Facultad de Ciencias Agrícolas, Universidad Autónoma Gabriel René Moreno, Km 9 carretera al Norte, El Vallecito, Santa Cruz, Bolivia

[6] INRA, UMR0745 EcoFoG, AgroParisTech, Cirad, CNRS, Université des Antilles, Université de Guyane, Campus Agronomique, Avenue de France, BP97387 Kourou Cedex, France

[7] Plateforme Génome Transcriptome de Bordeaux, INRA Site de Pierroton Bâtiment Artiga, 69 route d'Arcachon, 33610 Cestas, France

Published online: 10 January 2019

Dipteryx samples were collected from 73 individual trees in 35 localities in French Guiana, Brazil, Peru, and Bolivia (Table 1). Species determination was carried out using 16 herbarium vouchers and 25 photos of terminal branches taken in the field. Sampled material also included dried cambium or leaves preserved in silica gel. Genomic DNA was isolated from all reference samples according to Dumolin et al. (1995) at the Thünen Institute (Germany), at the São Paulo State University (UNESP), or at the Peruvian Amazon Research Institute (LBGM-IIAP). Three samples

Table 1 *Dipteryx* samples used for the initial SNP selection for MassARRAY genotyping, including information of tree species, geographic location, number of individuals (*n*) and structure genetic clusters

Country	Location	Species	*n*	Latitude	Longitude	Cluster/sub-cluster	Herbarium voucher
French Guiana	Paracou	*Dipteryx punctata*	1	5.2570	−52.9357	1/K2	NT1
French Guiana	Belizon	*Dipteryx odorata*	1	4.2131	−52.5206	1/K1	
		Dipteryx punctata[a]	1	4.2057	−52.5206	1/K2	
French Guiana	Saut_Maripa	*Dipteryx odorata*	2	3.8783	−51.8833	1/K1	
		Dipteryx punctata	1	3.8485	−51.8805	1/H	
Brazil	Jari	*Dipteryx odorata*	1	−0.4569	−52.8309	1/K1	
Brazil	Carajas	*Dipteryx odorata*	3	−6.2266	−49.0732	1/K1	
Brazil	Parauapebas	*Dipteryx odorata*	3	−6.0536	−50.0757	1/K1	
Brazil	Tapajos	*Dipteryx odorata*	4	−2.8547	−54.9677	1/K1	
Brazil	Arapins	*Dipteryx odorata*	2	−3.0964	−55.2526	1/K1	
Brazil	Anavilhanas	*Dipteryx odorata*	1	−2.5343	−60.8398	1/K1	
Brazil	Jau	*Dipteryx punctata*[a]	2	−1.8279	−61.5932	1/K2	
Brazil	Bom_Jesus	*Dipteryx odorata*	2	−3.5243	−64.9736	1/K1	
Brazil	Humaita	*Dipteryx odorata*	1	−7.5061	−63.0208	1/K1	
		Dipteryx punctata[a]	3	−7.5061	−63.0208	1/K2	
Brazil	Jamari	*Dipteryx odorata*	3	−9.3894	−62.9259	1/K1	
		Dipteryx punctata[a]	1	−9.3895	−62.9217	1/K2	
Brazil	Xapuri	*Dipteryx* cf. *ferrea*[a]	1	−10.5024	−68.5907	2/K4	
Brazil	Cumaru	*Dipteryx* cf. *ferrea*[a]	1	−10.7724	−69.6794	2/K4	
Peru	Huiririma	*Dipteryx micrantha*	1	−2.5018	−73.7840	2/K3	
Peru	Allpahuayo	*Dipteryx micrantha*	5	−3.7275	−73.4762	2/K3	GFL109, EHC1688
Peru	Jeberos	*Dipteryx micrantha*	2	−5.2296	−76.3177	2/K3	GFL55
Peru	Shucushuyacu	*Dipteryx micrantha*	1	−5.9946	−75.8245	2/K3	
Peru	Arica	*Dipteryx* cf. *ferrea*	2	−6.4391	−74.7870	2/K4	
Peru	Contamana	*Dipteryx* cf. *ferrea*	2	−7.2812	−74.9861	2/K4	
Peru	Macuya	*Dipteryx* cf. *ferrea*	2	−8.8742	−75.0110	2/K4	EML8, GHP3
Peru	Breu	*Dipteryx* cf. *ferrea*	1	−9.4864	−72.7066	2/K4	
Peru	Santa_Clara	*Dipteryx* cf. *ferrea*	1	−9.9639	−73.9966	2/K4	GFL24
Peru	Purus	*Dipteryx* cf. *ferrea*	2	−10.0755	−70.9802	2/K4	LHC130
Peru	Inuya	*Dipteryx* cf. *ferrea*	1	−10.5622	−73.0773	2/K4	
Peru	Iñapari	*Dipteryx* cf. *ferrea*	2	−11.2006	−69.6422	2/K4	GFL6, CBF64
Peru	Manu	*Dipteryx* cf. *ferrea*	1	−11.8779	−71.4012	2/K4	
Peru	Los_Amigos	*Dipteryx* cf. *ferrea*	1	−12.5682	−70.0812	2/K4	
Peru	Tambopata	*Dipteryx* cf. *ferrea*	1	−12.8297	−69.2848	2/K4	
Bolivia	Loma_Alta	*Dipteryx* cf. *ferrea*	1	−10.8823	−65.9622	2/K4	KPV619
Bolivia	Tahuamanu	*Dipteryx* cf. *ferrea*	3	−11.2181	−68.8868	2/K4	
Bolivia	Rurrenabaque	*Dipteryx* cf. *ferrea*	2	−14.2225	−67.7784	2/K4	
Bolivia	Altamira	*Dipteryx alata*	4	−16.1584	−61.9991	2/K5	KPV822, KPV831, KPV837
Bolivia	Quitunuquina	*Dipteryx alata*	4	−18.4183	−59.4409	2/K5	KPV917, KPV864

NT N. Tysklind, *GFL* G. Flores, *EHC* E. Honorio, *EML* E. Mejía, *GHP* G. Hidalgo, *LHC* L. Huaratapairo, *CBF* C. Ballón, *KPV* K. Paredes

[a]Samples originally collected as *D. odorata*

Table 2 Location and samples of *Dipteryx* used for SNP discovery from both RADSeq (a) and MiSeq (b) sequencing approaches

Sample code	Tree code	Country	Location	Species	Latitude	Longitude
DIODO_3[a,b]	GF38.NT1	French Guiana	Paracou	*D. punctata*	5.2570	−52.9357
DIODO_4[a,b]	GF39	French Guiana	Paracou	*D. odorata*	5.2728	−52.9253
DIMIC_297[b]	DIMIPER14026	Peru	Contamana	*D.* cf. *ferrea*	−7.2718	−74.9819
DIODO_285[a,b]	3DO2.KPV689	Bolivia	Loma_Alta	*D. odorata*	−10.7905	−65.9798

were selected for SNP maker development with restriction site associated DNA sequencing (RADSeq) for nuclear SNPs (Miller et al. 2007) and four individuals were used for genome skimming with Illumina MiSeq genome sequencing to obtain chloroplast and mitochondrial SNPs and INDELs (Straub et al. 2012; Table 2). We selected contigs based on the quality of sequencing and read coverage, the absence of other SNPs in the same contig, and contrasting patterns of allele frequency among individuals. Then, all individuals were genotyped (Table 1) using a MassARRAY ® iPLEX™ platform (Agena Bioscience™, San Diego, USA) at the Genome Transcriptome Facility of Bordeaux in France. Primers were designed with Assay Design Suite and allele calling was done with Typer Viewer v.4.0.24.71 (Agena Bioscience).

Genetic structure was assessed using Bayesian clustering method in Structure version 2.3.4 (Pritchard et al. 2000). Samples were grouped by the genetic clusters for further analysis. We selected SNP/INDEL loci based on their high amplification success, high average genetic differentiation *delta* (Gregorius 1987), and high correlation between genetic and geographic distances estimated in GDA-NT (Degen unpublished). Loci with allele frequencies fixed to specific genetic clusters were also selected.

RADSeq with three *Dipteryx* samples provided 10,934 putative loci, from which 1370 satisfied MassARRAY design criteria (no polymorphism within the 50 bp flaking regions). Genome skimming yielded 177 SNPs and 56 INDELS in the chloroplast genome, 51 SNPs and 53 INDELS in the mitochondrial genome, from which a total of 9 INDELS and 72 SNPs were considered for MassARRAY design. A set of 315 loci was used in the screening of 73 samples of five *Dipteryx* species using a MassARRAY assay design in ten multiplexes (Supplementary Material S1). Of the total screened loci, 15 loci showed no polymorphism and allele scoring was not possible for 8 loci. Considering the 292 remaining usable loci, the mean amplification rate per sample was 85% ranging from 78 to 90%, while the mean amplification rate per locus was 95% ranging from 36 to 100%. The samples were grouped in five genetic clusters (Supplementary Material S2) that matched each *Dipteryx* species (Table 1). Based on the high rate of amplification, genetic differentiation, and correlation between geographic and genetic distances, a final set of 100 loci was selected (Supplementary Material S3) and prepared for MassARRAY genotyping in three multiplexes (Supplementary Material S4). The final set of SNP and INDEL markers described here will be useful for studies on the genetic characterization and population genetics of *Dipteryx* species.

Acknowledgements This research was supported by the German Federal Ministry of Food and Agriculture (BMEL) and the Programa Nacional de Innovación para la Competitividad y Productividad (Innóvate Perú—contract 381-PNICP-PIAP-2014). Genotyping was performed at the Genomic and Sequencing Facility of Bordeaux (Grants from the Conseil Regional d'Aquitaine Nos. 20030304002FA and 20040305003FA, the European Union, FEDER No. 2003227 and Investissements d'avenir, No. ANR-10-EQPX-16-01 and CEBA: ANR-10-LABX-25-01 to the UMR EcoFoG). Permits ICMBIO Nos. 4760, 49369, registration CGen No. A16E4DA (Brazil), MAYA/VMABCCGDF/DGBAP/MEG No. 0280/2016 (Bolivia), R.D. No. 001-2016-SERNANP-DGANP, R.D. No. 001A-2015-SERFOR-DGGSPFFS-DGSPF, and Contrato No. 001-2016-SERFOR-DGGSPFFS-DGSPF (Peru) were granted for the collection, transport and genetic analyses of samples. We are also grateful to colleagues from IIAP, INRA French Guiana, Cirad Paracou Station, Office National des Forêts, and Thünen Institute for their assistance during field work and laboratory work.

References

Blanc-Jolivet C, Kersten B, Daïnou K, Hardy O, Guichoux E, Delcamp A, Degen B (2017) Development of nuclear SNP markers for genetic tracking of Iroko, *Milicia excelsa* and *Milicia regia*. Conserv Genet Resour 9:531–533

Chaves CL, Degen B, Pakull B, Mader M, Honorio E, Ruas P, Tysklind N, Sebbenn AM (2018) Assessing the ability of chloroplast and nuclear DNA gene markers to verify the geographic origin of Jatoba (*Hymenaea courbaril* L.) timber. J Hered 109:543–552

Dumolin S, Demesure B, Pettit R (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. Theor Appl Genet 91:1253–1256

Gregorius HR (1987) The relationship between the concepts of genetic diversity and differentiation. Theor Appl Genet 74:397–401

IUCN (2018) The IUCN red list of threatened species version 2018-2. http://www.iucnredlist.org. Accessed 15 Nov 2018

Jardine DI, Blanc-Jolivet C, Dixon RRM, Dormontt EE, Dunker B, Gerlach J et al (2016) Development of SNP markers for Ayous (*Triplochiton scleroxylon* K. Schum) an economically important tree species from tropical West and Central Africa. Conserv Genet Resour 8:129–139

Meyer-Sand BRV, Blanc-Jolivet C, Mader M, Paredes-Villanueva K, Tysklind N, Sebbenn AM, Guichoux E, Degen B (2018) Development of a set of SNP markers for population genetics studies of Ipe (*Handroanthus* sp.), a valuable tree genus from Latin America. Conserv Genet Resour 10:779–781

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Pakull B, Mader M, Kersten B, Ekue MRM, Dipelet UGB, Paulini M et al (2016) Development of nuclear, chloroplast and mitochondrial SNP markers for *Khaya* sp. Conserv Genet Resour 8:283–297

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Putzel L, Peters CM, Romo M (2011) Post-logging regeneration and recruitment of shihuahuaco (*Dipteryx* spp.) in Peruvian Amazonia: implications for management. For Ecol Manag 261:1099–1105

Soares TN, Melo DB, Resende LV, Vianello RP, Chaves LJ, Collevatti RG, Telles MPDC (2012) Development of microsatellite markers for the neotropical tree species *Dipteryx alata* (Fabaceae). Am J Bot 99:e72–e73

Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot 99:349–364

Vinson CC, Ribeiro DO, Harris SA, Sampaio I, Ciampi AY (2009) Isolation of polymorphic microsatellite markers for the tetraploid *Dipteryx odorata*, an intensely exploited Amazonian tree species. Mol Ecol Resour 9:1542–1544