Euphytica (2020) 216:45
https://doi.org/10.1007/s10681-020-02582-1

Genotyping-by-sequencing technology reveals directions for coconut (*Cocos nucifera* L.) breeding strategies for water production

**Pedro Henrique Araújo Diniz Santos · Thiago Motta Venâncio ·
Pedro Henrique Dias dos Santos · Helaine Christine Cancela Ramos ·
Fernanda Abreu Santana Arêdes · Alinne Oliveira Nunes Azevedo ·
Marcela Santana Bastos Boechat · Gonçalo Apolinário de Souza Filho ·
Semiramis Rabelo Ramalho Ramos · Luiz Angelo Mirisola · Wilson Menezes Aragão ·
Messias Gonzaga Pereira**

Received: 14 September 2019 / Accepted: 18 February 2020 / Published online: 24 February 2020

Abstract The dwarf coconut is widely used in the coconut hybrids production (tall × dwarf) because it presents higher precocity and higher productivity of the number of fruits in relation to the other varieties. However, these hybrids originated from Intervarietal crossing, produce fruits with low market acceptance for water quality. Intravarietal crosses (dwarf × dwarf) could act as an alternative, but little is known about the diversity within and between the dwarf sub-varieties, which can lead to a misdirection to breeding programs. In this study, we report the level of genetic variability between dwarf coconut accessions belonging to three sub-varieties (green, yellow and red), as well as the analysis of population structure. We used the RAD-sequencing methodology for analyzing 39 genotypes belonging to dwarf coconut populations collected in different producing regions of Brazil. Our results show that the SNP markers increased the power of detection of genetic variability, facilitating the decision making regarding future crosses in order to explore the heterosis in dwarf coconut breeding programs. Inter-sub-varietal crosses of dwarf coconut are highly promising to reduce the time and optimize the process of obtaining new cultivars for water production requiring field assessments to confirm its agronomic potential.

Keywords Genetic diversity · Population structure · SNP markers · RAD-sequencing

Introduction

The coconut palm (*Cocos nucifera* L.) is commonly grown in tropical regions around the world, being known and appreciated for its wide potential of use in natural consumption or industrialization of fruits and other organs, such as root, inflorescence, strain, leaves and palm heart, originating more than 100 products and by-products with significant economic value, being an important option in the generation of income and local development for producing countries

P. H. A. D. Santos · T. M. Venâncio ·
P. H. D. dos Santos · H. C. C. Ramos · F. A. S. Arêdes ·
A. O. N. Azevedo · M. S. B. Boechat ·
G. A. de Souza Filho · M. G. Pereira (✉)
Laboratório de Melhoramento Genético Vegetal,
Universidade Estadual do Norte Fluminense Darcy
Ribeiro, Campos dos Goytacazes, RJ, Brazil
e-mail: messiasgpereira@gmail.com

S. R. R. Ramos · W. M. Aragão
Embrapa Tabuleiros Costeiros, CPATC, Aracaju,
SE, Brazil

L. A. Mirisola
Commercial Regon LTDA, Belo Horizonte,
MG, Brazil

(Marina et al. 2009). Among all these products and by-products, the commercialization of coconut water is the activity most explored by producers, comprising most of the revenues generated by coconut crop (Monteiro et al. 2013).

The global coconut water market was valued at USD 961.9 million in 2017 and is expected to expand around 25.0% from 2017 to 2025 (HexaResearch 2019). Due to the growing world demand and the expansion of the market, it is also important offer to the farmers, hybrids more productive which produce water with good acceptance by consumers.

The plant, which is not native to Brazil, had its introduction and adaptation in the coastal sandy soils of Brazil, allowing the beginning of coconut production in the country, even with a lack of resources (Siqueira et al. 2002). The first citations of coconut plantations in Brazil date from 1587 and were exclusive of the tall variety (Bondar 1955). The first dwarf variety plants arrived in Brazil in 1925 and the plantations started in lands of the Brazilian northeastern coastal regions (Siqueira et al. 2002).

Currently, there are three sub-varieties of dwarf coconut: Red, Green, and Yellow. Green Dwarf coconut is the commercial variety most widely used for this purpose since it shows a higher acceptance by consumers and producers due to characteristics such as good palatability (Kwiatkowski et al. 2008), high productivity and precocity (Monteiro et al. 2013). This variety begins to produce early, between three and four years, while the tall variety takes about seven years to start the production. In addition, it produces a greater number of fruits/plant/year and presents a small plant height, facilitating harvest (Castro et al. 2009).

The growing demand of coconut water in large urban centers and the financial profitability of coconut crop make it extremely necessary to obtain new cultivars with a focus on water production, giving producers the power to choose highly productive cultivars more adapted to specific regions. Studies aiming at developing varieties adapted to specific climate conditions remain scarce, focuses only on Intervarietal levels (tall × dwarf) to obtain hybrids (Ninan and Satyabalan 1964; Azevedo et al. 2018). However, these hybrids produce fruits with low market acceptance for water quality.

An alternative for this could be the use of Intravarietal crosses in Dwarf coconut as reported by Azevedo et al. (2018). Some breeding programs aiming at coconut water quality have focused on Intra-sub-varietal selections in Green Dwarf Brazilian populations. Still now, little is known about the diversity in sub-varieties, since the results reported by Azevedo et al. (2018) only detect at the intra-varietal level, requiring more robust techniques, generating a gap in the best direction to take (Inter-sub or Intra-sub crosses?).

That gap could be more effectively covered by a combination of classical breeding with molecular biology techniques, genomics and bioinformatics. This will not only help generate more productive cultivars in a short-term but will also fuel the development of a molecular database to support future studies worldwide.

Molecular markers, such as microsatellite or simple sequence repeats (SSRs), have been widely applied in studies on coconut genetic diversity (Ribeiro et al. 2013; Geethanjali et al. 2018) According to Perera et al. (2000), molecular markers could also be used in the analysis of the structure of coconut populations, being suitable for assessing and characterizing coconut germplasm. The growing development of new technologies in molecular biology, especially those involving DNA sequencing, has dramatically changed the scenario for detecting and monitoring genome-wide polymorphism (Schuster 2008; Craig et al. 2008). Currently, single nucleotide polymorphisms (SNPs) are rapidly replacing SSRs in plant breeding and genetics applications, in particular, because they are more abundant, stable, efficient, and increasingly cost-effective (Duran et al. 2009; Edwards and Batley 2010; Desalegne et al. 2017).

Genotyping by sequencing (GBS) techniques, particularly the restriction-site associated DNA-sequencing (RAD-seq) (Davey et al. 2011), allow the genome analysis at a reduced complexity level. RAD-seq can be applied in any species, without prior knowledge of SNPs, whose quantitative value depends on the enzyme chosen for cleavage and on the size and complexity of the genome. This methodology has been proven very efficient in the discovery of thousands of single nucleotide polymorphisms (Rowe et al. 2011).

Rowe et al. (2011) emphasize that this technology can be used for different purposes, such as genetic maps, genome assembly, QTL identification and structural population studies, aiming to understand interspecific organizations such as migration and speciation patterns.

In this sense, we raise the hypothesis that the GBS technique can understand deeper relationships from sub-varietal dwarf coconut plants could direct ideal crosses for the generation of new hybrids. In this perspective, our study aims to investigate and describe the variability and genetic structure of dwarf coconut populations available in Brazil via SNPs identified by the RAD-seq methodology in order to select divergent populations and obtain, in the future, dwarf coconut hybrids, reducing the time and optimizing the process of obtaining new cultivars for water production.

Materials and methods

Plant material

Thirteen dwarf coconut populations from different producing regions of Brazil were analyzed, being potential parents for the development of hybrids in these regions. Population choice was based on the results obtained by Azevedo et al. (2018). Leaf samples were collected in the states of Pará, Pernambuco, Ceará, Sergipe, Paraíba, and Bahia (Table 1).

DNA extraction

The genomic DNA of three individuals from each of the 13 populations previously described in Table 1 was extracted by the DNeasy Plant Mini kit from QIAGEN. DNA concentrations were estimated through the use of the NanoDrop 2000C spectrophotometry (Thermo Scientific).

Preparation of libraries

Genomic DNA samples were digested by the restriction enzyme AseI. The resulting fragments of the digestion were subjected to analysis in Bioanalyzer 2100 from Agilent Technologies® and fragments of approximately 500 bp were selected for the preparation of the RAD libraries with the Illumina™ TruSeq® DNA Sample Prepare kit, following manufacturer's instructions. In order to identify and sequence all individuals in a single sequencing lane, each sample received a dual index barcode. Library quality was assessed using a Bioanalyzer instrument and sequencing was performed using a Hiseq 2500 platform (paired-end mode, 2 × 100 bp) at LaCTAD/UNICAMP.

Data analyses

RAD-Seq reads were processed using Stacks (Catchen et al. 2011), an efficient package to analyze RAD-Seq data. Stacks processes RAD-tags generated by restriction enzymes and uses a statistical maximum likelihood model to identify SNPs and remove sequencing errors. *Process_Radtags* was used to de-multiplex and filter sequencing reads. The criterion for the presence of three or more identical sequences to constitute a RAD-tag was chosen. Tags were recognized by sequences present at the beginning of each fragment. Subsequently, we used *Ustacks*, which constructs the *loci*, inferring SNPs, alleles, and tags present in each sample. *Cstacks* assembled the catalog of loci from SNPs and *Sstacks* compared each sample against the catalog. Finally, the function *populations* were used, providing the genotype matrices and population parameters for diversity analysis.

The observed heterozygosities (estimated by the proportion of heterozygotes observed at a given locus), observed homozygosities (estimated by the proportion of homozygotes observed at a given locus), the number of private alleles in each population, the number of individuals sharing the same allele within each population, the number and percentage of polymorphic sites, and the average frequency of the most frequent allele at each locus (P) were obtained by the package *populations*. In addition, at least 8 of the 13 populations were required to share the same locus to be considered in the analysis.

The data were submitted to four sequential filters by PLINK software (Purcell et al. 2007). The SNP with more than 50% of missing data, and SNP with minimum allele frequency (MAF) less than 0.05 have also been removed. Finally, highly correlated ($r^2 \sim 1$) markers were eliminated, that is, redundant markers to explain the genetic distance between individuals. The distance matrix was obtained using the identity by state complement (i.e. 1-IBS) for further analysis.

Phylogenetic reconstruction was performed using Bayesian Inference (BI) through the Monte Carlo Markov chain (MCMC) method and Maximum Likelihood (ML) estimation. The software MrMODELTEST (Posada and Buckley 2004) and Modeltest 3.7

Table 1 Description of the coconut populations for the study of variability through sequencing

Dwarf types	Population	Sampling site
BGDJ	Brazilian Green Dwarf-Jiqui	COHIBRA-CE
BGD-PRPB	Brazilian Green Dwarf-Paraipaba	COHIBRA-CE
BGD-BA	Brazilian Green Dwarf-Bahia	Ribeira do Pombal-BA
BGD-SZ 1	Brazilian Green Dwarf-Souza 1	Figueiredo's Farm Souza-PB
BGD-SZ 2	Brazilian Green Dwarf-Souza 2	Carlos Dantas's Farm Souza-PB
BGD-PA 1	Brazilian Green Dwarf-Pará 1	Moju-PA
BGD-PA 2	Brazilian Green Dwarf-Pará 2	Moju-PA
BGD-TR	Brazilian Green Dwarf-Trairi	COHIBRA-CE
MRD	Malayan Red Dwarf	EMBRAPA -SE
MYD	Malayan Yellow Dwarf	EMBRAPA -SE
BRDG	Brazilian Red Dwarf-Gramame	COHIBRA-CE
CRD	Cameroon Red Dwarf	COHIBRA-CE
BYDG	Brazilian Yellow Dwarf-Gramame	COHIBRA-CE

(Posada and Crandall 1998) were used to select the nucleotide substitution models of BI and ML analyses, respectively. Likelihood values were calculated and the model selected according to the Akaike Information Criterion (AIC). The evolution model selected for each analysis was HKY+ G for both analyses.

The reconstruction of Bayesian inference was completed with MrBayes v.3.1.1 (Ronquist and Huelsenbeck 2003). The four MCMC chains were conducted simultaneously, randomly starting trees up to 10^7 generations. Trees were sampled every 1000 generations, resulting in 10,000 trees. The first 2500 trees were discarded from the analysis. The posterior probability (PP) values (Rannala and Yang 1996) were determined from the consensus tree through the remaining 7500 trees. The convergence of the log-likelihood was analyzed with the software TRACER v.1.7 (Rambaut et al. 2018).

The maximum likelihood estimation was performed in the software PAUP v. 4.0 (Swofford 2001) and the heuristic search was carried out using the TBR algorithm. The bootstrap calculation was performed with 1000 replications. Heat map analysis was conducted in R (R Development Core Team 2015), using the heatmap2 function.

The software Structure v. 2.3.4 (Pritchard 2010) was used to investigate the genetic structure of the population. The admixture model and correlated allele frequencies were used with a *Burnin Period* = 5000, followed by the extension of 50,000 repetitions during the analysis, in addition to the identification of the optimal K following the method of Evanno et al. (2005).

Results and discussion

Genetic parameters of diversity

Thirteen populations containing three individuals from different coconut producing regions were analyzed, totaling 39 individuals. We detected, on average, 164,013 variant sites (SNPs), i.e. sites in which the reference genome differed from individuals regardless of the variation between them. From this total of SNPs, on average 27,873 polymorphic sites were found, allowing differentiating the populations and individuals belonging to them (Table 2).

The average frequency of the most frequent allele at each locus (P) was 0.93, ranging from 0.89 (BYDG) to 0.95 (BGD-TR and MRD). Because this estimate represents the degree of fixation based on allelic frequencies found in each population, we can infer that the alleles of these populations are fixed. When considering the average of sub-varieties for this estimate, we noticed that green dwarfs, red dwarfs, and yellow dwarfs presented values of 0.94, 0.94, and 0.91, respectively. These data corroborate the high estimates of homozygosity observed for each population (> 0.90), confirming the fixation of alleles.

The high homozygosity of populations is perfectly expected, as well as the high value of the most frequent allele because it is an autogamous plant with a self-

Table 2 Estimates of genetic parameters of diversity in 13 dwarf coconut populations

Access	Polymorphic sites	Polymorphic sites (%)	P*	Observed homozygosity
BGDJ	12,689	0.06	0.94	0.94
BGD-PRPB	26,342	0.08	0.94	0.93
BGD-SZ 2	29,122	0.10	0.93	0.92
BGD-SZ 1	25,155	0.09	0.93	0.92
BGD-BA	32,193	0.13	0.92	0.90
BGD-PA 1	26,449	0.08	0.94	0.93
BGD-PA 2	25,229	0.08	0.94	0.93
BGD-TR	24,859	0.07	0.95	0.94
MRD	22,990	0.07	0.95	0.94
MYD	25,547	0.09	0.94	0.93
BRDG	29,160	0.09	0.94	0.93
CRD	25,137	0.08	0.94	0.93
BYDG	57,483	0.19	0.89	0.87
Mean	27,873.46	0.093	0.93	0.93

*Average frequency of the most frequent allele at each locus

pollination rate of 94.3% (Bourdeix 1988). This indicates that these populations can be used as lines in the production of simple hybrids of dwarf coconut, increasing the heterotic effect in crossings directed based on molecular descriptors, such as the crossing between genetically distant unrelated individuals, and morpho-agronomic descriptors, such as populations with a high production of coconut water.

The average value obtained for the number of private alleles was 8390, ranging from 2773 (BGDJ) to 36,237 (BYDG). Private alleles are those that refer to the counting of alleles present in only one of the sub-varieties (Marshall and Brown 1975), being their presence directly related to the sample size.

Phylogenetic analyses

From the analysis of 164,013 SNPs in the 39 analyzed individuals, we observed a high number of redundant markers that would not be useful to explain the genetic distance among the assessed populations. Thus, the markers that showed a high correlation were eliminated, leading to a final analysis of 538 SNPs, corresponding to 37 individuals. We excluded two individuals from the BGDJ population due to filter applied to lost data per individual. With the exception of Table 2, the remainder of the analyzes were performed using the filtered dataset, containing 538 SNPs.

The sequence alignment resulted in 1614 characters, of which 163 were parsimony informative, 372 variable, and 766 conserved, and the combined phylogenetic reconstruction of Bayesian inference and Maximum-likelihood estimation was performed with 13 populations. We identified two well supported and very distinct clades. The first one was formed exclusively by populations of green dwarf and the second one was formed by populations of red and yellow dwarf (Fig. 1).

Our results, as well as those reported by Gunn et al. (2011), suggest that an evolutionary divergence occurred in the long-term among populations and no within. We also observed that populations tended to group not by their locality but by their sub-variety. This fact associated with the characteristics of dwarf coconut trees, such as slow growth, self-pollination, and production of the *niu vai* form of fruits, is consistent with domestication bottlenecks during the evolution period of these sub-varieties (Gunn et al. 2011).

The group of green dwarfs was divided into two sub-clades, being the first one formed only by the BGD-BA population and the second group by the other seven populations. One of the main challenges to be overcome in the search for superior green dwarf cultivars is the high evidence of inbreeding that probably result from the domestication of this sub-variety (Gunn et al. 2011). Our results show that the BGD-BA population came from an older ancestor in relation to the others and could be a starting point for crosses of more distinct populations (Fig. 1).

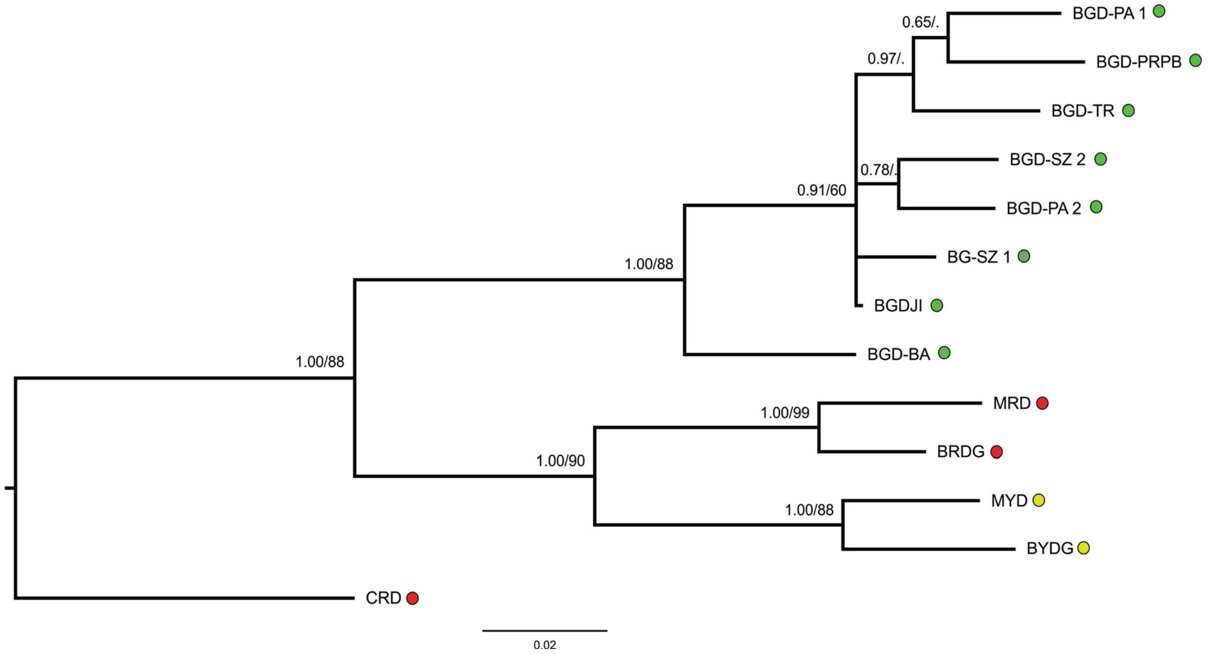

Fig. 1 Bayesian and maximum-likelihood topology depicting the evolutionary relationships of the 13 studied dwarf coconut populations. Support values at nodes represent Bayesian posterior probabilities (PP) and bootstrap percentages (ML) (on the right), respectively. The tree was rooted in CRD

The two populations of yellow dwarf, MYD, and BYDG, which presented a high degree of genetic similarity, were already evidenced in the literature (Ribeiro et al. 2013).

The red dwarf population group was grouped in a clade containing MRD and BRDG populations and shared the same ancestor that gave origin to the red and yellow clades, making it possible for us to infer that yellow dwarf populations were originated from red dwarf ancestors. Azevedo et al. (2018) identified specific alleles in red dwarf populations capable to differentiate from other red dwarf populations. Therefore, the CRD population has alleles that can be differentiated from other dwarf populations, not only by color but also by their origin, evidencing that this population may have been subjected to a founding effect process in which the phenomenon of evolution is the result of the establishment of a new population by few original founders (Mayr 1940; Futuyma 2006).

Population structure and differentiation

The first group was formed by accessions of Brazilian green dwarf and the second group by populations of red and yellow dwarf (Fig. 2). The group of green dwarfs is in accordance with the genetic diversity classification proposed by the Generation Challenge Program of the French Agricultural Research Centre for International Development (GCP/CIRAD), which states that there is no variability between the Brazilian dwarf coconut (BGD).

Heat map plot

The data were submitted to the heat map analysis (Fig. 3) in the software R by using the *heat map* package. This analysis was also concordant with the dendrogram, as a high genetic similarity was observed between the dwarf coconut populations green × green, yellow × yellow, and red × red.

This analysis allows identifying crosses based on the genetic distance between populations, and the clearer the color in the matrix, the higher the distance between populations and, consequently, the higher the probability of exploring the heterosis in crossing, as is the case of crosses (CRD × BGD-PA 1, CRD × BGD-PA 2, and BRDG × BGD-PA 2).

In a study based on morpho-agronomic data in order to study the genetic variability in dwarf coconut populations, Cambuí (2007) verified the division of six

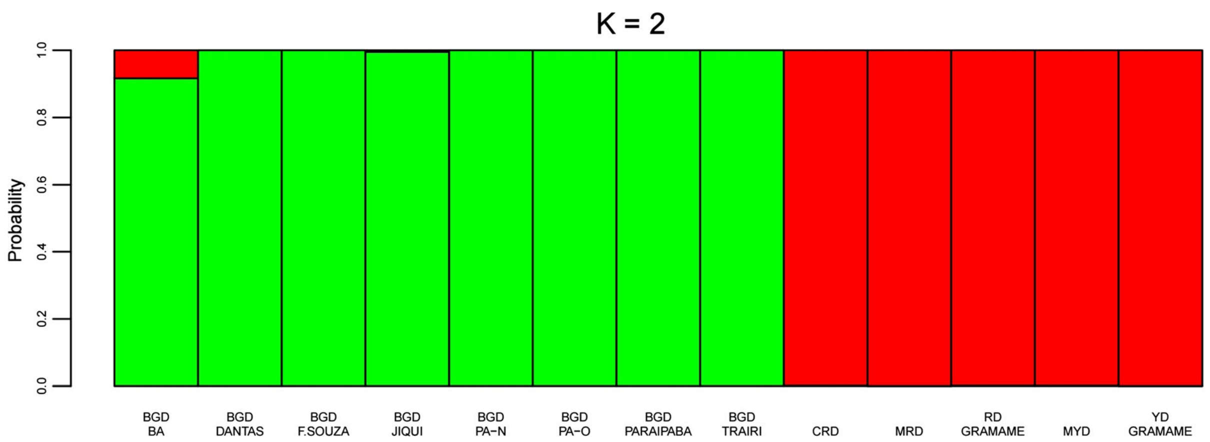

Fig. 2 Result of the analysis of the Structure of 13 populations of dwarf coconut. Assignments for each population are shown for k = 2 (primary K)

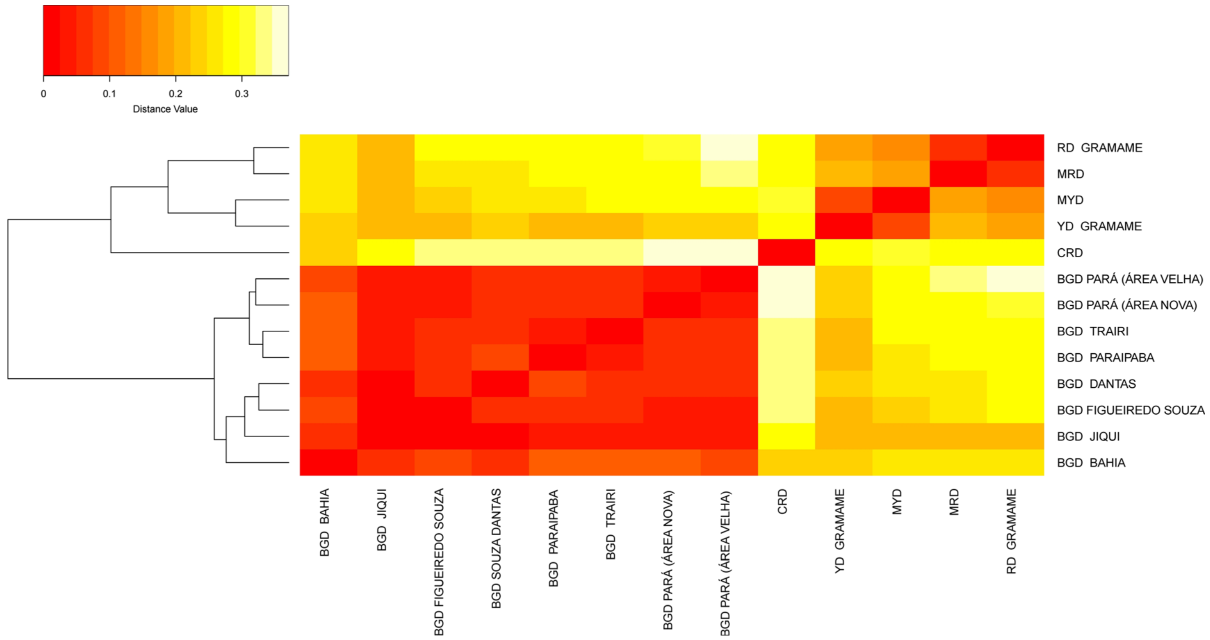

Fig. 3 Heat map considering the 13 coconut populations

cultivars in three groups. The first group was formed by the accessions BRDG, MYG, and BYDG, the second group by the accessions MRD and BGDJ, and the third group by the accession CRD, which is in accordance with the results reported here, in particular in relation to CRD, which remained isolated in the grouping.

Azevedo et al. (2018) conducted a study that aimed at characterizing cultivars by the fingerprint technique using microsatellite markers. The results were consonant with those found in our study. Regarding the genetic variability, GBS was more effective in detecting tags that contributed to better discriminate populations.

Noël et al. (2011), in a study using microsatellites to assess the genetic diversity of green dwarf coconut accessions in Côte d'Ivoire, also found a low genetic diversity among Philippine green dwarfs. Meerow et al. (2003) reported similar results obtained with microsatellite markers to show that the Malaysian red dwarf is distinguished from the Malaysian green and yellow dwarfs. The latter two are believed to be very similar genetically.

Perera et al. (2000) hypothesize that dwarf coconut palms might have evolved over a long domestication period, from a small population of tall coconut palms, indicating that this event may be the reason for the observed low genetic diversity levels in dwarfs. This indicates a smaller genetic distance between dwarf coconut accessions of Brazil and, consequently, a higher level of genetic relationship.

Other studies using SSR markers have revealed the existence of a much higher genetic diversity in tall accessions compared to dwarfs (Rivera et al. 1999; Perera et al. 2000, 2003) which is a reflection of their breeding habits: talls being allogamous, while dwarfs generally autogamus.

Due to the magnitude of the observed results from phylogenetic and heat map analysis, the production of intra-sub-varietal hybrids is not recommended when heterosis and gains in coconut water production are envisaged. Phenotypic differences among these populations found in the field are mainly due to environmental factors, such as management, age of plants, among others.

This study shows that even with the use of SNP markers, the detected variability and genetic distance were not enough to ensure the recommendation of intra-sub-varietal crosses between green dwarf populations to obtain hybrids. In contrast, in correspondence to the reproductive system of the green dwarf coconut as a plant framed as preferential autogamous, plants with a high proportion of fixed loci are observed. We can say that nature made available inbred lines that can be used directly in crossings with other populations from other dwarf sub-varieties in order to explore heterosis. Thus, based on our results, we recommend for coconut water production crossings between green dwarf × red dwarf and green dwarf × yellow dwarf or even inter-sub-varietal crossings of red dwarf and yellow dwarf. Therefore, Inter-sub-varietal crosses of dwarf coconut are highly promising to reduce the time and optimize the process of obtaining new cultivars for water production, requiring field assessments to confirm its agronomic potential.

Acknowledgements The authors kindly thank the North Fluminense State University (UENF), Commercial Regon and FAPERJ for the financial support and infrastructure to carry out this work. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES).

References

Azevedo AON, Azevedo CDO, Santos PHAD et al (2018) Selection of legitimate dwarf coconut hybrid seedlings using DNA fingerprinting. Crop Breed Appl Biotechnol. https://doi.org/10.1590/1984-70332018v18n4a60

Bondar G (1955) A Cultura do Coqueiro no Brasil. Tipografia Naval, Salvador

Bourdeix R (1988) Genetic determinism in dwarf coconut germ colour. Oleagineux (France) 43:371–374

Cambuí E (2007) Diversidade genética entre cultivares de coqueiro anão (*Cocos nucifera* L., Var. Nana). Universidade Federal de Sergipe, São Cristóvão

Castro CP, Passos EEM, Aragão WM (2009) Fenologia de cultivares de coqueiro-anão nos tabuleiros costeiros de sergipe. Revista Brasileira de Fruticultura 31:13–19. https://doi.org/10.1590/S0100-29452009000100004

Catchen JM, Amores A, Hohenlohe P et al (2011) Stacks: building and genotyping Loci de novo from short-read sequences. G3. https://doi.org/10.1534/g3.111.000240

Craig DW, Pearson JV, Szelinger S et al (2008) Identification of genetic variants using bar-coded multiplexed sequencing. Nat Methods 5:887–893. https://doi.org/10.1038/nmeth.1251

Davey JW, Hohenlohe PA, Etter PD et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510

Desalegne BA, Dagne K, Melaku G et al (2017) Efficiency of SNP and SSR-based analysis of genetic diversity, population structure, and relationships among cowpea (*Vigna unguiculata* (L.) Walp.) germplasm from East Africa and IITA inbred lines. J Crop Sci Biotechnol 20:107–128. https://doi.org/10.1007/s12892-016-0051-0

Duran C, Appleby N, Clark T et al (2009) AutoSNPdb: an annotated single nucleotide polymorphism database for crop plants. Nucleic Acids Res 37:D951–D953. https://doi.org/10.1093/nar/gkn650

Edwards D, Batley J (2010) Plant genome sequencing: applications for crop improvement. Plant Biotechnol J 8:2–9. https://doi.org/10.1111/j.1467-7652.2009.00459.x

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Futuyma DJ (2006) Ernst Mayr, genetics and speciation. Trends Ecol Evol 21:7–8. https://doi.org/10.1016/j.tree.2005.10.001

Geethanjali S, Anitha Rukmani J, Rajakumar D et al (2018) Genetic diversity, population structure and association analysis in coconut (*Cocos nucifera* L.) germplasm using SSR markers. Plant Genet Resour Charact Util 16:156–168. https://doi.org/10.1017/S1479262117000119

Gunn BF, Baudouin L, Olsen KM (2011) Independent origins of cultivated coconut (*Cocos nucifera* L.) in the old world tropics. PLoS ONE 6:e21143. https://doi.org/10.1371/journal.pone.0021143

HexaResearch (2019) Coconut water market size and forecast, by form (liquid, powder), by packaging (tetra pack, plastic bottle, others), by distribution channel (offline, online), by region, and trend analysis, 2015–2025. https://www.hexaresearch.com/research-report/coconut-water-market. Accessed 9 May 2019

Kwiatkowski A, Clemente E, Scarcelli A, Vida JB (2008) Quality of coconut water "in natura" belonging to Green Dwarf fruit variety in different stages of development, in plantation on the northwest area of Paraná, Brazil. J Food Agric Environ 6:102–105

Marina AM, Che Man YB, Nazimah SAH, Amin I (2009) Antioxidant capacity and phenolic acids of virgin coconut oil. Int J Food Sci Nutr. https://doi.org/10.1080/09637480802549127

Marshall D, Brown A (1975) Optimum sampling strategies in genetic resources conservation. In: Frankel O, Hawkes J (eds) Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge, pp 55–80

Mayr E (1940) Speciation phenomena in birds. Am Nat 74:249–278. https://doi.org/10.1086/280892

Meerow AW, Wisser RJ, Steven Brown J et al (2003) Analysis of genetic diversity and population structure within Florida coconut (*Cocos nucifera* L.) germplasm using microsatellite DNA, with special emphasis on the Fiji Dwarf cultivar. Theor Appl Genet. https://doi.org/10.1007/s00122-002-1121-z

Monteiro CM, Caron ES, da Silveira SF et al (2013) Control of foliar diseases by the axillary application of systemic fungicides in Brazilian coconut palms. Crop Prot 52:78–83. https://doi.org/10.1016/j.cropro.2013.05.013

Ninan CA, Satyabalan K (1964) A study of natural, self and cross (dwarf × tall) progenies of dwarf coconuts of the West Coast of India and its bearing on the genetics of dwarfs and the putative hybridity of their off-type progenies. Caryologia 17:77–91. https://doi.org/10.1080/00087114.1964.10796118

Noël KKJ (2011) Microsatellite gene diversity within Philippines dwarf coconut palm (*Cocos nucifera* L.) resources at Port-Bouët, Côte d'Ivoire. Sci Res Essays. https://doi.org/10.5897/sre11.877

Perera L, Russell JR, Provan J, Powell W (2000) Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera* L.). Genome 43:15–21

Perera L, Russell JR, Provan J, Powell W (2003) Studying genetic relationships among coconut varieties/populations using microsatellite markers. Euphytica 132:121–128. https://doi.org/10.1023/A:1024696303261

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808. https://doi.org/10.1080/10635150490522304

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics. https://doi.org/10.1093/bioinformatics/14.9.817

Pritchard JK (2010) Documentation for structure software: version 2.3. In: Practice. https://doi.org/10.1002/spe.4380060305

Purcell S, Neale B, Todd-Brown K et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. https://doi.org/10.1086/519795

R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0

Rambaut A, Drummond AJ, Xie D et al (2018) Posterior summarization in Bayesian phylogenetics using tracer 1.7. Syst Biol. https://doi.org/10.1093/sysbio/syy032

Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J Mol Evol 43:304–311. https://doi.org/10.1007/BF02338839

Ribeiro FE, Baudouin L, Lebrun P et al (2013) Genetic diversity in Brazilian tall coconut populations by microsatellite markers. Crop Breed Appl Biotechnol 13:356–362. https://doi.org/10.1590/S1984-70332013000400006

Rivera R, Edwards KJ, Barker JH et al (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. Genome 42:668–675

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574. https://doi.org/10.1093/bioinformatics/btg180

Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. Mol Ecol. https://doi.org/10.1111/j.1365-294x.2011.05197.x

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods 5:16–18. https://doi.org/10.1038/nmeth1156

Siqueira LA, Aragão WM, Tupinambá EA (2002) A introdução do coqueiro no Brasil, importância histórica e agronômica. Documentos. Embrapa Tabuleiros Costeiros, Aracaju

Swofford DL (2001) Phylogenetic analysis using parsimony. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland. https://doi.org/10.1007/bf02198856