PHYLOGENOMICS OF THE COCONUT (*COCOS NUCIFERA* L.)

BEE FONG GUNN

A thesis submitted for the degree of Doctor of Philosophy

of

The Australian National University, Canberra, Australia

MAY 2016

Division of Evolution, Ecology and Genetics

Research School of Biology

College of Medicine, Biology and Environment

The Australian National University

Canberra, ACT, 2601, Australia

Declaration

I, Bee Fong Gunn, certify that this thesis is the result of research undertaken while a student in the Research School of Biology at the Australian National University. The work described is original and my own, except as otherwise stated or referenced in the text. I also certify that all assistance and resources that contributed to the production of the thesis has been duly acknowledged. This thesis has not been submitted to another university or similar institution.

Bee Fong Gunn

28 April 2016

Acknowledgements

Institutions

I wish to thank the following institutions for funding support to make this study possible, for fieldwork assistance and for attendance of conferences to present my results.

The Australian National University, Australian Postgraduate Scholarship

The Australian National University, Research Sch. of Biology, Div. of Evolution, Ecology and Genetics, HDR fieldwork grant (2010)

I was fortunate to receive fieldwork support for collecting samples in Flores and the Bogor Botanical Gardens.

The Australian National University Vice Chancellor's Travel grant (2011)

It was a privilege to receive this grant, which enabled me to present a talk at the European Network of Palm Scientists (EUNOPS) conference in Montpellier, France and spend three weeks at the laboratories of the French Agricultural Research Centre for International Development (CIRAD), Montpellier working on coconut microsatellites with coconut geneticist Dr. Luc Baudouin.

National Geographic Society, CRE grant 8178-07

(Co-PIs: Kenneth Olsen, Washington University in St. Louis, USA and Bee F. Gunn). This important grant enabled me to carry out fieldwork in Madagascar, Comoros, Seychelles, Lizard Island, Flores and Taveuni.

The Australasian Systematic Botany Society travel grant (2011)

I'm very grateful to the ASBS for enabling me to attend and present a poster and lightning talk at the XVIII International Botanical Congress in Melbourne.

The Australian French Association for Science and Technology Travel Fellowship (2013)

It was an honour to receive the AFAS travel fellowship to Montpellier for collaborative project with CIRAD/IRD, France on genome size of coconut with Drs. Alain Rival and Luc Baudouin.

The Australian Museum, Lizard Island Research Station (2009)

For hosting me during fieldwork on Lizard Island and special thanks to Dr. Anne Hoggett and staff at LIRS for the logistics while on the island.

The Australian National Herbarium, Canberra

The herbarium has been a great resource for my research and essential for depositing vouchered specimens without which publications will not be possible. Specials thanks for Brendan Lepschi, Jo Palmer, Maggie Nightingale, Serkan Alasya and all the staff and volunteers at ANH.

The Australian National Botanic Garden, Canberra

For acquiring the Lizard Island coconut as part of Australia's native living collection for research and permissions for collecting in National Parks.

The Royal Botanic Gardens and the Domain, Sydney

For permission and assistance for collecting fresh leaf material for Cocoseae palms and for use of the herbarium. Simon Goodwin, Collections Manager for palms, Karen Wilson and Leonie Stanberg.

Bogor Botanical Garden, Bogor, Indonesia

For sponsoring my fieldwork permit application from LIPI, local travel arrangements and contacts.

Queensland Parks and Wildlife Service

Permit for collecting in the Lizard Island National Park.

Centre for Biodiversity Analysis, ANU

For organizing many important workshops in Beast analysis, Landscape Genomics and Computational Macro-evolution and Next Generation Sequencing analyses.

Many people have to be thanked for providing me with much help during my PhD. Special acknowledgements must be given to my PhD committee, Mike Crisp, Rod Peakall, Matthew Prebble and Joe Miller for their generous support, guidance and patience throughout the years.

I am deeply grateful to my principal supervisor Mike Crisp who has given me the opportunity to complete my PhD at an outstanding research institution with world-class academics, postdocs and students as colleagues. I thank Mike for placing me at the cutting edge of science and technology, launching me into high throughput sequencing, an experience that will be unforgettable. I was fortunate to participate in Mike's course on Australian Plant Diversity as a teaching assistant and appreciate the opportunity to mentor and interact with students.

I am incredibly thankful to my co-advisor Rod Peakall who has been a supportive mentor and taken much time to guide me in structuring my thesis. I'm very grateful for the many discussions and the time spent on clarifying ideas and analyses. Rod has taught me how to carry out rigorous science, to turn things on its head and to ask critical questions.

I am privileged to have Matthew Prebble as an advisor from the ANU College of Asia and the Pacific, who has contributed to my understanding of archaeology and human influence on Holocene environments in the Pacific. I thank him for collecting coconut samples for DNA from the outlier islands in the Pacific and recording local names and uses of each sample. I was fortunate to be on two field trips in the Pacific with archaeologists from ANU and USA. I am grateful for his interest in using ancient DNA for dating phylogenies and is currently work in progress but time does not permit inclusion into the thesis.

I was very fortunate to have Joe Miller as an advisor and his support throughout the course of my PhD, the many helpful discussions and affiliation with the Australian National Herbarium, Canberra.

Members of the Crisp lab present and past, thanks to Bokyung Choi, Meredith Crosgrove, Melita Milner, Andrew Thornhill, Richard Carter, Lindsay Popple and David Morris for their friendship and as colleagues.

I would like to thank Carsten Kulheim for his advice on the next generation sequencing laboratory experiments and the bioinformatics to extract the sequences. Celeste Linde for her invaluable help with part of the next generation sequencing expenses and for advice on the migration and gene flow analyses. Aaron Chuah from the Genome Discovery Unit, John Curtin School of Medical Research, ANU for his bioinformatics input with the next generation sequencing pipeline. Geoff Hope, ANU College of Asia and the Pacific for collecting coconut leaf samples for DNA from Aneityum, Vanuatu.

Many thanks to Justin Borevitz, for giving me the unique opportunity to carry out Genotyping By Sequencing in his lab and understanding landscape genomics. Thanks to Penny Gullan and Peter Cranston for their collegiate support, kind generosity and encouragement during the years and imparting a wealth of knowledge.

Thanks to Scott Keogh, Head of EEG and former convener of the HDR students for his general mentoring and keeping the department running well. Adrienne Nicotra for looking after HDR students and to Karen Scholte for keeping me on track with milestones and assistance for HDR matters. Patricia Blackwell, Michael Jennions and Dave Rowell for keeping me in great humour and catering for students at the ANU Kiola Field Station. Our wonderful staff at Evolution, Ecology

and Genetics, Division, IT, Teaching and Learning Centre, especially Jan Elliott, Audra Johnstone, Wes Keys, Denis Hawkins for all their help.

Coopers and Cladistics discussion group over beer at University House are memorable times and I would like to thank the group and especially Mike Crisp and Peter Cranston for keeping the latest publications and Coopers flowing.

Bootstrap camps were valuable times to gather with members of the Cook and Bruhl labs to learn from colleagues and be excited to show and tell what was accomplished the past year in the surrounds of native plants and fresh air. I would like to specially thank Lyn Cook for the discussion sessions on phylogenetics and the thought provoking questions that ensued.

Collaborators:

I'm very lucky to have international collaborators and specialists who have been very helpful and inspiring throughout my PHD.

Luc Baudouin, Geneticist, CIRAD, Montpellier who has been a fountain of knowledge on coconuts and has enlightened me on the genetics and breeding of coconuts.

Kenneth Olsen, Associate Professor, Washington University, St. Louis, USA who has inspired me to understand plant domestications and his enthusiasm whenever we discuss about the evolutionary biology of coconuts.

Alain Rival, CIRAD, Regional Director in Asia, for enabling the genome size project to come to fruition and keeping the tradition of French/Australian Scientific and Technological collaborations alive.

I wish to thank the examiners for their reviews of the thesis.

Last but not least, my very special thanks to my family, Michael Gunn for his patience as I toiled for years and kept us on breadline and Bonnie and Jonathan who have been proud children. I would like to dedicate this thesis to them.

Prologue

Format

The main layout of the thesis follows the Style Guide of the Australian National University. This thesis consists of six chapters, one of which has been published and one has been submitted and under review. The formatting and bibliographic style of the papers follow the editorial style of the target journals.

Thesis layout

The current status of each of the papers are presented below and the percentage of my contribution. Chapter 1 (Introduction) provides the project background and rationale and links for each of the chapters. Chapter 6 (General Discussions and conclusions) summarizes the findings and conclusions for each of the chapters and provides link between each of them.

Chapter 1

Introduction

Chapter 2

Independent origins of cultivated coconuts (Cocos nucifera L.) in the Old World Tropics

Authors: **Bee F. Gunn**, Luc Baudouin and Kenneth M. Olsen

PLoS ONE 6(6):e21143; DOI: 10.137/journal.pone.0021143; 2011

Contribution: 75%

Chapter 3

Phylogenomics and population genomics of the coconut: integrating phylogeography, phylogeny and population genomics

Chapter 4

Patterns of gene flow and dispersal of coconut

Chapter 5

Genome size variation in Attaleinae (Arecaceae) with emphasis on coconuts (*Cocos nucifera* L.)

Authors: **Bee F. Gunn**, T. Beule, L. Baudouin, P. Ilbert, C. Duperray, M. Crisp, A. Issali, J-L. Konan and A. Rival

Published: *Ploidy and domestication are associated with genome size variation in Palms in the American Journal of Botany 2015: 102(10): 1625-1633 and is included in the Appendix.*

Contribution: 75%

Chapter 6

General discussion and conclusions

Appendix 1

The presence of coconut in southern Panama in pre-Columbian times: clearing up the confusion

Authors: Luc Baudouin, **Bee F. Gunn** and Kenneth M. Olsen

Annals of Botany 113: 1- 5, 2014. DOI: 10.1093/aob/mct244.

Contribution: 30%

Appendix 2

Ploidy and domestication are associated with genome size variation in Palms in the American Journal of Botany 2015: 102(10): 1625-1633

Authors: **Bee F. Gunn**, L. Baudouin, T. Beule, P. Ilbert, C. Duperray, M. Crisp, A. Issali, J-L. Konan and A. Rival

Contribution: 75%

Abstract

The coconut palm (*Cocos nucifera* L.) is a monotypic member of the Cocoseae tribe (subtribe Attaleinae) and its evolutionary history is profoundly intertwined with that of human civilization. It is well adapted to drift-dispersal by oceanic currents, colonizing coastal ecosystems and islands. Both today and in the past, humans have exploited it as a potable source of water, nutritious food, fibre and shelter during their prehistoric voyages of civilization across the Pacific and Indo-Atlantic Oceans. This long-term human interaction and dissemination has altered its phenotype and the lack of a universal domestication trait has obscured the putative wild phenotype and its original geographical location. The main objectives of this phylogenomic study of the coconut are: 1) to determine the centre of coconut domestication, 2) elucidate the geographical origin of the coconut, 3) identify hotspots of genetic diversity, 4) understand migration and gene flow patterns and 5) the impacts of domestication on coconut genome size. Bayesian analysis of population genetic structure was applied to multi-locus microsatellites generated from 1,322 coconut accessions from across the species range. Results strongly suggest that coconuts are differentiated into two genetic populations corresponding to the Indo-Atlantic and Pacific oceanic basins. This pattern suggests independent regions of domestication in these two regions and proposed two centres: island Southeast Asia and the southern margins of the Indian subcontinent. I uncovered evidence for admixtures between these populations consistent with Austronesian trade routes from Southeast Asia to Madagascar and Arab trading along east African coast. To address the overarching objective of the geographical origin of the coconut, I integrated the sub-disciplines of phylogeography, phylogenetics and population genetics to evaluate four criteria: i) ancestral haplotype location, ii) phylogeny and divergence times, iii) coalescence

and ancestral reconstruction and iv) genetic diversity. I applied high throughput sequencing technology from chloroplast (14 loci) and nuclear (4 loci) genomes from 118 coconuts across 19 subpopulations representing the species' distribution. Evaluation of criteria using genomic-scale sequence data, taken together with fossil evidence, suggest that the ancestral geographical origin of the extant coconut is likely in Australasia encompassing Australia, Indonesian Archipelago and Papua New Guinea. The Indo-Atlantic is a hotspot for genetic diversity and a sink population. Migration patterns and gene flow directions were inferred by testing hypotheses of migration models based on geographical and genetic *a priori* implementing Bayesian coalescent framework and Log Bayes Factors (LBF). For first set of models, LBF indicated that the coconut is not panmictic. The network model showed migration trend from out of Southeast Asia into Oceania consistent with Austronesian migrations. For the second set, bi-directional gene flow model between the Indo-Atlantic and Pacific showed best support. The impact of domestication on genome size and ploidy levels was investigated by flow cytometry technique. Quantifications of genome size of 23 cultivars including Talls, Dwarfs, hybrids and wild-sown coconuts indicate variation. My findings demonstrated that highly domesticated Dwarf types expressed significantly less genome size variation than the Tall types. Ancestral reconstruction of genome sizes amongst Attaleinae show that polyploidy evolved independently at least four times.

Table of Contents

List of Figures

Chapter 2: Independent origins of cultivated coconut (*Cocos nucifera* L.) in the Old World Tropics

NB: The page numbering for Chapter 2 refers to the publication.

Chapter 3: Phylogenomics and population genomics of the coconut: Integrating phylogeography, phylogeny and population genomics

Note: This chapter has been removed.

List of Tables

CHAPTER 1: INTRODUCTION

The coconut (*Cocos nucifera* L.) encapsulates the history of humanity in the humid tropics because its usefulness has enabled mankind to colonize islands and create trade routes across the Pacific, Atlantic and Indian Oceans. Both historically and today, this palm has a myriad of uses as a source of food, drink, and fuel (Burkill, 1966; Purseglove, 1972). Every part of the plant is useful to man, and recently coconut oil has been manufactured into bio-diesel in the Pacific. The history of dispersal and domestication of this species is thus fundamentally intertwined with human history in the tropics. Currently, little is known about the domestication of the coconut, location of its geographical origin, dispersal history and the impacts of domestication on its genome size. Understanding the phylogeography, phylogenetics and population genomics of this species would provide profound insights into our own history on this planet.

Cocos nucifera L., a monotypic genus in the Cocoseae tribe (Arecaceae), is monoecious and reproduces entirely by seed. Coconuts are adapted to drift-dispersal by ocean currents (Edmondson, 1941) and the fossil records indicate that the species underwent an ancient (mid-Tertiary) dispersal event long before being exploited by humans (Sauer, 1971; Gunn, 2004). Importantly, this early dispersal is expected to have created a population genetic signature in this species, so that the dispersal route could be traced by examining the phylogeographic structure of plants sampled across the species range.

Superimposed on this ancient phylogeographic structure is the more recent history of dispersal, cultivation and domestication by humans. It is widely accepted that the cultivated coconut exists in two main forms, *niu kafa* and *niu vai*,

which are distinguished by the nut-to-husk ratio and fruit shape (Harries, 1978). Plant breeders distinguish the mostly cross-pollinating Tall type from the mostly self-pollinating Dwarf type. "Dwarfs" are short-stemmed, mostly autogamous and presumed to be the more highly domesticated form due to their habit, low genetic variation, fruit color and occurrence near human habitation. "Talls" have long stems, are later bearing and mostly allogamous. "Talls" can bear fruits that are *niu kafa* or *niu vai* types depending on the cultivar whereas "Dwarfs" only bear *niu vai* type fruits. "Tall" varieties have higher genetic variability and are preferred for plantations because their endosperms produce higher quality copra. Current data, while limited, suggest that the "dwarfs" worldwide are closest genetically to the "Talls" in the Pacific, tentatively suggesting a single domestication origin of the "Dwarfs" (Lebrun, Grivet et al., 1998). However, due to its long history of dispersal, first by water alone and then by human activity, the identity and location of the origin of domestication is still unknown.

1.1 **Ancient distribution and dispersal of coconut**

The natural range of the coconut species, predating humans, is most likely in the Indo-Pacific (Dransfield, Uhl et al., 2008). The earliest coconut endocarp fragments and roots, similar to the *niu kafa* type, were documented from Aneityum Island (Vanuatu) and radiocarbon dated to 5,040 BP; these coconuts are thought to have arrived by natural dispersal (Spriggs, 1984). Other early coconut remains (4,555 BP) were discovered on Aitape (northern Papua New Guinea) in association with human skeletal remains (Hossfeld, 1965). On Pagan (Marianas), Fosberg and Corwin (1958) identified a fossil coconut seedling and attributed it to human dispersal in Quaternary tuff (4,000 BP), although Sauer (Sauer, 1971) argued that it was pre-human based on the geology of the region.

1.2 Human-mediated dispersal

The waves of Austronesian voyagers during the Holocene, most likely from island Southeast Asia (Soares, Rito et al., 2011) were responsible for the spread of the coconut's range in Oceania, and the presence of wild coconuts aided their colonization of these islands. Coconuts were critical for survival on these islands as well as during their long sea journeys (Bellwood, 1978; Massal and Barrau, 1980). The Seychelles are ancient oceanic islands and among the last to be discovered by humans. Abundance of "coker nutts" was reported by two separate chroniclers of the *Ascension* captained by Alexander Sharpeigh, which happened onto the Seychelles islands in 1609 (Sauer, 1967). In the Seychelles, the cultivated coconuts have been proposed to represent independent domestication of the native coconuts, which had extremely thick husks and small nuts, rather than introduction of domesticated varieties from other regions (Sauer, 1967).

Pre-Columbian records by Oviedo (1851) documented the presence of coconuts and cultural uses by the indigenous Indians on the Pacific coast of Panama, Costa Rica and Colombia (Stone, 1966; Zizumbo-Villarreal and Quero, 1998). In Panama, coconuts were grown but not used for fibers, as cotton and agave provided this need (Stone, 1966). Early Spanish settlers established coconut plantations in the Central and South American coasts, most likely from stocks from the Philippines and Panama.

Coconuts were not recorded as growing in the Atlantic-Caribbean region until their introduction by Portuguese colonizers. The first record of coconut introduction to the West Indies was in 1582 in Puerto Rico from Portuguese plantations in Cape Verde. The Portuguese started coconut plantations in West Africa, Cabo and Brazil during the 16th Century, after Vasco da Gama's expedition in 1498 to the Indian Ocean (Sauer, 1967).

1.3 Present-day wild populations

Gruezo (1984) described a coconut population from eastern Samar Island (Philippines) showing no evidence of domestication and growing in an area that has had minimal human influence, both historically and today. Buckley (1984) similarly reported putative wild populations on Lizard Island (Australia); both of these Pacific populations are "Talls" characterized by *niu kafa*-like fruits. It has been suggested that at the eastern edge of the Pacific, there is occurrence of a natural coconut population on the Pacific coast of Colombia. Fossil cocosoid fruit from the upper Paleocene has been discovered in northern Colombia by Gomez-Navarro et al. (2009), providing further evidence for the long-term occurrence of the species in this region. Hill (1929) proposed that the Cocos Keeling Islands coconuts originated from ocean borne nuts from the eastern Archipelago of the Pacific Ocean. Sauer (1967) suggested that wild coconuts became established without human intervention on the oceanic islands of Seychelles in the Indian Ocean.

1.4 Overview of objectives and Thesis Structure

1.4.1: What is the population structure of the coconuts worldwide and where is its centre(s) of domestication?

Independent origins of cultivated coconut (Cocos nucifera L.) in the Old World Tropics (Chapter 2)

Authors: **Bee F. Gunn**, Luc Baudouin and Kenneth M. Olsen

PLoS ONE 6(6): e21143 doi:10.1371/journal.pone.0021143

The coconut lacks a single universal domestication trait such as shattering of stalk in wild relatives of rice or branching in maize. Only the Dwarf form coconuts are considered highly domesticated, having traits associated with

domestication such as the dwarf habit, autogamy and *niu vai* fruits. The Tall forms have also been domesticated but are highly variable in their domestication status. The first objective of this study is to understand the domestication history of the coconut and to determine its centre(s) of domestication. In this study (Gunn, Baudouin et al., 2011), we used ten microsatellite loci from a sample of 1,322 coconut accessions from across the globe to estimate the population genetic structure as it relates to the history of human migrations. Our findings suggested that: i) despite the widespread movement of coconuts by humans, the species has retained clear population structure on the global scale, one corresponding to the Indo-Atlantic ocean basin and the other in the Pacific oceanic basin; ii) present-day cultivated coconuts arose through independent domestications in the Indian and Pacific and Indian Ocean basins and iii) geographical locations of genetically admixed populations are consistent with human introductions of Pacific germplasm along ancient trading routes connecting Asia to Africa. We proposed two geographical origins of coconut cultivation: island Southeast Asia and the southern margins of the Indian subcontinent.

Although, the centres of domestication of the coconut have been proposed, the phylogeography, phylogenetic history, dispersal and the gene flow patterns between the populations have not been investigated and the geographical location of the origin of *Cocos nucifera* is still unknown. Microsatellite data do not allow us to reconstruct the deeper phylogeographic or phylogenetic history providing estimations of lineage ages and divergence times. Genomic data offers a new era of phylogenetics and phylogeography – that of phylogenomics to understand evolutionary relationships. To address the questions in the following two chapters of this thesis, I used next generation sequence data from targeted loci within the Large Single Copy region of the chloroplast genome and from the nuclear genome.

1.4.2 Where is the geographical location of the origin of the coconut and the hotspots of genetic diversity?

Phylogenomics and population genomics of the coconut: integrating phylogeography, phylogeny and population genetics (Chapter 3)

Note: This chapter has been removed.

The phylogeography of the coconut is intriguing because of the complexity of the natural dispersal by oceanic currents and its long history of dissemination by humans has obscured the location of the wild populations. During the past decade we have the empirical capabilities to generate genome-scale data from high throughput sequencing (HTS) which may be exploited to integrate the micro-evolutionary to macro-evolutionary scales for understanding biodiversity patterns. The second major aim of this thesis was to pinpoint the likely geographical location of the origin of *Cocos nucifera* L. and to elucidate the hotspots of coconut genetic diversity to provide insights into future conservation of untapped coconut germplasm and landraces (traditional varieties) which may well carry disease resistance genes or traits advantageous for crop improvement. To address these overarching objectives, I used phylogenomics to tease apart the natural and human-mediated dispersal patterns of the coconut by i) examining the coconut's phylogeography, and using using haplotype networks to infer their dispersal patterns; ii) determining the phylogenetic relationships and divergence time of coconut lineages and iii) investigating the genetic diversity hotspots of the coconut.

In this chapter and the following chapters, I applied genomic scale data from the chloroplast and four low copy nuclear genes for 118 coconuts from putative wild and cultivated populations (19) sampled from across the globe

integrating approaches provided by phylogeography, phylogenetics and population genomics disciplines.

1.4.3 Is the coconut a panmictic population?

Patterns of gene flow and dispersal of coconut (Chapter 4)

Note: This chapter has been removed.

Long distance dispersals (LLD) have been invoked to explain the biogeographical distributions of many terrestrial flora and fauna (Gillespie, Baldwin et al., 2012; Miryeganeh, Takayama et al., 2014). Plants capable of transoceanic dispersal, such as coconuts, are expected to show high gene flow across vast areas and are often panmictic. The third major objective of this thesis was to assess migration models using the coalescent framework implemented in the software Migrate (Beerli and Palczewski, 2010) based on sequences of multiple genes and individuals from both chloroplast and nuclear genomes. The first set of migration models tested the hypothesis of panmixia versus a network of symmetrical migrations among regions and the second set of migration models tested the directionality of gene flow between the Indo-Atlantic vs. Pacific gene pools. Dispersal patterns and estimated relative migration rates will provide insights into the direction of migration of genes to tease apart human-mediated from current influenced dispersals.

1.4.4 What are the impacts of domestication on genome size? (Chapter 5)

Ploidy and domestication are associated with genome size variation in Palms

Authors: **Bee F. Gunn**, L. Baudouin, T. Beulé, P. Ilbert, C. Duperray, M. Crisp, A. Issali, J-L. Konan and A. Rival

American Journal of Botany 102(10): 1625 – 1633. 2015

The impacts of domestication on the genome size of annual crops such as *Zea mays* (Laurie and Bennett, 1985), *Triticum sp.*, (Dvořák, Terlizzi et al., 1993),

Poa annua (Grime, 1983), *Panicum virgatum* (Riley and Vogel, 1982) and *Solanum tuberosum* (Spooner, Rodríguez et al., 2008) have been well studied but very few genome size evolution studies have been carried out for long-lived tree crops (Miller and Gross, 2011). Polyploidy and gene duplications are likely to increase the genome size and have also been associated with domestication traits such as phenology in sunflowers (Blackman, Rasmussen et al., 2011). Genome size variation and ploidy levels among the Tall cultivars and domesticated Dwarf cultivars have not been examined. We applied flow cytometry method to estimate the genome sizes of 23 coconut cultivars worldwide including wild-sown coconuts. The main objectives were: 1) to determine the actual genome size of coconut for which contradictory values were published; 2) to identify and study intraspecific variation, and the impact of domestication on genome size; 3) to test whether genome size is less variable in Dwarf than Tall coconut types and 4) to reconstruct ancestral genome sizes across the subtribe Attaleinae.

The determination of the ploidy levels and absolute genome size is a prerequisite for the future of genome sequencing of the coconut, optimizing depth of reads and accuracy of annotations of its whole genome. A fully annotated coconut genome sequence will provide immeasurable resources for genome wide association studies and Quantitative Trait Loci mapping vital for the future of crop improvement and understanding of disease resistance.

1.4.5 General discussion, conclusions and future directions (Chapter 6)

Note: This chapter has been removed.

In this chapter I discuss the main results and implications of findings from each of the chapters, the main conclusions and future directions.

1.5 Significance of the study

This study on the phylogenomics of the coconut is highly significant for understanding the history of human civilization in the tropics and human impacts on the landscape through their long-term interactions with the coconut palm. This project will provide key information on the genetic diversity of putative wild coconut populations, which may be exploited by global coconut breeding programs (International Coconut Genetic Resources) to enhance germplasm collections. It is critical to identify regions with high genetic diversity as island countries are vulnerable to climate change and sea level rises due to global warming leading to permanent loss of heterogeneous coconut germplasm. Breeding programs in coconuts have depended on a narrow gene pool and genetically heterogeneous wild populations were excluded. Unfortunately, lethal yellowing disease is threatening to devastate coconut populations globally and the need to extend the genetic diversity and identify disease resistance genes is vital.

This research will provide a wealth of information on the population genetic structure, dispersal and gene flow patterns of cultivated and non-cultivated coconuts worldwide which can be used for identification and conservation of germplasm from source-sink populations, characterizing desirable traits and high-yielding products in crop breeding programs, assessment of the interactions between genomes and the environment and for the future of whole genome sequencing and annotations of the coconut.

REFERENCES:

BEERLI, P., AND M. PALCZEWSKI. 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185: 313-326.

BELLWOOD, P. 1978. Mans Conquest of the Pacific: The Prehistory of Southeast Asia and Oceana.

BLACKMAN, B. K., D. A. RASMUSSEN, J. L. STRASBURG, A. R. RADUSKI, J. M. BURKE, S. J. KNAPP, S. D. MICHAELS, AND L. H. RIESEBERG. 2011. Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* 187: 271-287.

BUCKLEY, R., AND H. HARRIES. 1984. Self-Sown Wild-Type Coconuts From Australia. *Biotropica* 16: 148-151.

BURKILL, I. H. 1966. A dictionary of the economic products of the Malay Peninsula. Second ed. Ministry of Agriculture and Cooperatives, Kuala Lumpur.

DRANSFIELD, J., N. W. UHL, C. B. ASMUSSEN, W. J. BAKER, M. HARLEY, AND C. E. LEWIS. 2008. Genera palmarum: The Evolution and Classification of Palms. Kew Publishing, Royal Botanic Gardens.

DVOŘÁK, J., P. D. TERLIZZI, H.-B. ZHANG, AND P. RESTA. 1993. The evolution of polyploid wheats: identification of the A genome donor species. *Genome* 36: 21-31.

EDMONDSON, C. H. 1941. Viability of coconut seeds after floating in sea. Bernice P. Bishop Museum.

FOSBERG, F. R., AND C. G. 1958. A fossil flora from Pagan, Mariana Islands. *Pacific Science* 12: 2-26.

GILLESPIE, R. G., B. G. BALDWIN, J. M. WATERS, C. FRASER, R. NIKULA, AND G. K. RODERICK. 2012. Long-distance dispersal: a framework for hypothesis testing. *TREE* 27: 47 - 56.

GOMEZ-NAVARRO, C., C. JARAMILLO, F. HERRERA, S. L. WING, AND R. CALLEJAS. 2009. Palms (Arecaceae) from a Paleocene rainforest of northern Colombia. *American Journal of Botany* 96: 1300–1312.

GRIME, J. P. 1983. Prediction of weed and crop response to climate based upon measurements of nuclear DNA content, 4. National Vegetable Research Station: The Association of Applied Biologists, Warwick.

GRUEZO, W. S., AND H. C. HARRIES. 1984. Self-sown, wild-type coconuts in the Philippines. *Biotropica*: 140-147.

GUNN, B. F. 2004. The phylogeny of the Cocoeae (Arecaceae) with emphasis on *Cocos nucifera*. *Annals of the Missouri Botanical Garden* 91: 505–522.

GUNN, B. F., L. BAUDOUIN, AND K. M. OLSEN. 2011. Independent origins of cultivated coconut (*Cocos nucifera* L.) in the Old World Tropics. *PloS One* 6: e21143.

HARRIES, H. 1978. The evolution, dissemination and classification of *Cocos nucifera* L. *Botanical Review* 44: 265-319.

HILL, A. W. 1929. The original home and mode of dispersal of the coconut. *Nature* 124: 133-134.

HOSSFELD, P. S. 1965. Radiocarbon dating and paleoecology of the Aitape fossil human remains. *Proceedings of the Royal Society of Victoria* 78: 161–165.

LAURIE, D. A., AND M. D. BENNETT. 1985. Nuclear DNA content in the genera Zea and Sorghum. Intergeneric, interspecific and intraspecific variation. *Heredity* 55: 307—313.

LEBRUN, P., L. GRIVET, AND L. BAUDOUIN. 1998. The spread and domestication of the coconut palm in the light of RFLP markers. *Soiviviaire* 5: 235-245.

MASSAL, E., AND J. BARRAU. 1980. Food plants of the South Sea islands, South Pacific Community, 28–30, Noumea, New Caledonia.

MILLER, A. J., AND B. L. GROSS. 2011. From forest to field: Perennial fruit crop domestication. *American Journal of Botany* 98: 1389-1414.

MIRYEGANEH, M., K. TAKAYAMA, Y. TATEISHI, AND T. KAJITA. 2014. Long-distance dispersal by sea-drifted seeds has maintained the global distribution of *Ipomea pes-caprae* subsp. *brasiliensis* (Convolvulaceae). *PloS One* 9: e91836.

OVIEDO Y VALDEZ, G. F. D. 1851. 1535. *Historia general y natural de las Indias, islas y Tierra Firme del Mar Oceano. Pa raqua*.

PURSEGLOVE, J. 1972. Tropical crops. Monocotyledons. Longman Group Limited, London.

RILEY, R. D., AND K. P. VOGEL. 1982. Chromosome numbers of released cultivars of switchgrass, indiangrass, big bluestem, and sand blue-stem. *Crop Science* 22: 1082–1083.

SAUER, J. D. 1967. Plants and man on the Seychelles coast, 14. The University of Wisconsin Press, Madison.

SAUER, J. D. 1971. A re-evaluation of the coconut as an indicator of human dispersal. *In* C. Riley [ed.], Man across the sea, 309-319. University of Texas Press, Austin.

SOARES, P., T. RITO, J. TREJAUT, M. MORMINA, C. HILL, E. TINKLER-HUNDAL, M. BRAID, et al. 2011. Ancient Voyaging and Polynesian Origins. *AJHG* 88: 239-247.

SPOONER, D. M., F. RODRÍGUEZ, Z. POLGÁR, H. E. BALLARD, AND S. H. JANSKY. 2008. Genomic origins of potato polyploids: GBSSI gene sequencing data. *Crop Science* 48: S-27-S-36.

SPRIGGS, M. 1984. Early coconut remains from the south Pacific. *Journal of Polynesian Society* 93: 71-76.

STONE, D. 1966. Synthesis of lower Central America Ethnohistory, 309-319. Univ. of Texas Press, Austin.

ZIZUMBO-VILLARREAL, D., AND H. J. QUERO. 1998. Re-evaluation of early observations on coconut in the New World. *Economic Botany* 52: 68-77.

OPEN ACCESS Freely available online

PLoS one

Independent Origins of Cultivated Coconut (*Cocos nucifera* L.) in the Old World Tropics

Bee F. Gunn[1], Luc Baudouin[2], Kenneth M. Olsen[3]*

1 Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, Australia, **2** Centre International de Recherches en Agronomie pour le Développement (CIRAD), Montpellier, France, **3** Biology Department, Washington University, St. Louis, Missouri, United States of America

Abstract

As a portable source of food, water, fuel, and construction materials, the coconut (*Cocos nucifera* L.) played a fundamental role in human migrations and the development of civilization across the humid tropics. Here we investigated the coconut's domestication history and its population genetic structure as it relates to human dispersal patterns. A sample of 1,322 coconut accessions, representing the geographical and phenotypic diversity of the species, was examined using ten microsatellite loci. Bayesian analyses reveal two highly genetically differentiated subpopulations that correspond to the Pacific and Indo-Atlantic oceanic basins. This pattern suggests independent origins of coconut cultivation in these two world regions, with persistent population structure on a global scale despite long-term human cultivation and dispersal. Pacific coconuts show additional genetic substructure corresponding to phenotypic and geographical subgroups; moreover, the traits that are most clearly associated with selection under human cultivation (dwarf habit, self-pollination, and "*niu vai*" fruit morphology) arose only in the Pacific. Coconuts that show evidence of genetic admixture between the Pacific and Indo-Atlantic groups occur primarily in the southwestern Indian Ocean. This pattern is consistent with human introductions of Pacific coconuts along the ancient Austronesian trade route connecting Madagascar to Southeast Asia. Admixture in coastal east Africa may also reflect later historic Arab trading along the Indian Ocean coastline. We propose two geographical origins of coconut cultivation: island Southeast Asia and southern margins of the Indian subcontinent.

Citation: Gunn BF, Baudouin L, Olsen KM (2011) Independent Origins of Cultivated Coconut (*Cocos nucifera* L.) in the Old World Tropics. PLoS ONE 6(6): e21143. doi:10.1371/journal.pone.0021143

Editor: Pär K. Ingvarsson, University of Umeå, Sweden

Received March 26, 2011; **Accepted** May 20, 2011; **Published** June 22, 2011

Copyright: © 2011 Gunn et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The study was funded by a National Geographic Society (CRE 8178-07) grant to KMO and BFG, http://www.nationalgeographic.com/field/, and COGENT network and the CGIAR Generation Challenge Progamme to LB, http://www.inibap.org/cogent/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kolsen@wustl.edu

Introduction

The impact of the coconut palm (*Cocos nucifera* L.) on the history of human dispersal in the humid tropics is unparalleled in the plant kingdom. As a portable source of both food and water, the coconut played a critical role in the ability of humans to voyage, establish trade routes, and colonize lands in the Pacific Rim and regions throughout the Old World tropics [1,2]. This species continues to have hundreds of uses as a source of food, drink, fiber, construction material, charcoal, and oil (used in cooking, pharmaceuticals, industrial applications, and biofuels); over 12 million hectares of coconut are currently planted across 89 tropical countries [3]. The history of dispersal and cultivation of this species is thus fundamentally intertwined with human history in the tropics.

The long-term interaction between humans and coconuts has shaped both the geographical distribution of *C. nucifera* and its phenotypic diversity. While the coconut fruit is naturally adapted for dispersal by sea currents [4], its pantropical dissemination was achieved with the help of humans [5,6]. A native of the Old World tropics, the species was spread to eastern Polynesia and subsequently introduced to the Pacific coasts of Latin America, most likely by pre-Columbian Austronesian seafarers from the Philippines [7]. In the Indian Ocean, the composition of coconut populations was likely influenced by Austronesian expansions westward to Madagascar. Later, coconuts were introduced by Europeans from India to the Atlantic coasts of Africa and South America and to the Caribbean [8]. The species is typically found in areas of present or past human activity, and all or nearly all coconut populations worldwide have likely been influenced by human cultivation and dispersal.

Phenotypically, coconuts vary widely in the degree to which they show evidence of selection under human cultivation. Classic analyses of coconut fruit morphology revealed two predominant fruit types, named after traditional Polynesian varieties: the '*niu kafa*' form, characterized by oblong, triangular fruits with a large proportion of fibrous husk; and the '*niu vai*' form, whose fruits are rounded and often brightly colored, with a large proportion of liquid endosperm [9,10]. The '*niu kafa*' form has been interpreted as the more ancestral morphology, reflecting natural selection for ocean dispersal, and the '*niu vai*' form as reflecting selection under human cultivation [1]. Coconuts have also been traditionally classified into 'Dwarf' and 'Tall' varieties based on tree habit. 'Dwarfs' represent about 5% of coconut palms and are cultivated worldwide; they are typically found near human habitation and show traits closely associated with human selection: slow trunk growth, self-pollination, and the production of *niu vai* fruits [11]. The more common 'Tall' coconuts are outcrossing and grow faster than 'Dwarfs,' resulting in greater height at reproductive maturity. Many 'Talls' are grown for the production of copra for oil

extraction and coir for fiber; while actively cultivated, these varieties lack the obvious domestication traits of the self-pollinating Dwarfs.

The lack of universal domestication traits among coconut varieties, combined with the long history of human interaction with this species, have made it difficult to trace the coconut's cultivation origins. However, applications of molecular markers for purposes of crop germplasm characterization have provided some insights into the coconut's evolutionary history, genetic diversity and population structure (e.g., [12,13]). Analyses using RFLPs (e.g., [14]), microsatellites [15,16] and AFLP markers [17] have suggested the presence of two genetically distinct groups, corresponding broadly to the Pacific Ocean basin on one side and the Indian and Atlantic Oceans on the other (see also [18,19]).

In the last decade, a worldwide coconut germplasm collection, coordinated through the International Coconut Genetic Resources Network (COGENT) and the French Agricultural Research Centre for International Development (CIRAD), with further support through the Generation Challenge Programme (GCP: http://gcpcr.grinfo.net/index.php), has served as the primary source of materials for genetic characterizations. Together with a polymorphic microsatellite marker kit [20], the GCP/CIRAD coconut collection has been used to characterize genetic diversity in regional coconut collections (e.g., [21,22]), infer origins of specific cultivars [7], and assess planting material for trueness to type [23]. Importantly, this worldwide collection has not been used previously to examine the coconut's cultivation history. Moreover, while global in scope, the GCP/CIRAD collection has left some geographical regions under-represented. Most notably, it contains few coconuts from the western Indian Ocean, which would be key to elucidating any influence of ancient Austronesian expansions in this region.

In the present study, we have employed ten polymorphic loci from the GCP/CIRAD microsatellite kit to examine genetic variation in a worldwide collection of >1300 coconuts, representing GCP/CIRAD germplasm plus collections from key under-sampled regions of the western Indian Ocean: Madagascar, Comoros, and Seychelles islands. We use population structure analyses, together with ethnographic and archaeobotanical evidence, to examine the impacts of human-mediated dispersal and domestication on this important tree crop. Our analyses suggest the following: 1) Despite the widespread movement of coconuts by humans, both historically and today, the species has retained clear population structure on a global scale; 2) Present-day cultivated coconuts arose through independent domestications in the Indian and Pacific Ocean basins; however, the definitive domestication traits dwarf habit, self-pollination, and *niu vai* fruits arose only with the Pacific domestication event; and 3) Geographical locations of genetically admixed populations are consistent with human introductions of Pacific germplasm along the ancient trading routes connecting Asia to Africa.

Results

With new sample collections that fill an important gap in an already extensive worldwide data set, we have examined variation at ten microsatellite loci in a global collection of coconut germplasm. Genotypes were successfully obtained for 1322 samples, representing 1210 individuals from the GCP/CIRAD collection and 112 samples from the western Indian Ocean (Table S1). For germplasm characterization purposes, the GCP/CIRAD collection has previously been categorized into a hierarchical classification scheme based on a combination of criteria, including phenotypes, molecular markers, geographic distribution, and known introduction history [7]. Compositions of the 16 GCP/CIRAD groups and three additionally sampled Indian Ocean regions are shown in Table 1. The highest level in the GCP/CIRAD classification divides coconuts into two groups, A and B. Group A coconuts occur primarily in the region spanning Southeast Asia to the Pacific coast of America. Group B coconuts occur across coastal S. Asia, W. Africa, the New World Atlantic, and the Caribbean [2,14]. Subgroups correspond to geographical and/or phenotypic subsets within each group (Table 1); the greater number of subgroups for Group A coconuts reflects this group's higher phenotypic diversity.

Within-group genetic diversity

Genetic diversity for Dwarf coconut varieties (populations 1 3; Table 1) is on average less than half that of Talls, with mean unbiased gene diversity values of $H_e = 0.271$ and 0.579 for the two growth forms, respectively. Dwarfs also show greater evidence of inbreeding (mean observed heterozygosity, $H_o = 0.060$ and 0.480 for Dwarfs and Talls, respectively), consistent with the low within-cultivar genetic heterogeneity characterizing these self-pollinating varieties, most of which are pure-breeding lines. This overall pattern of reduced genetic variability in Dwarfs has been reported previously (e.g., [18]) and is consistent with domestication bottlenecks during the evolution of these highly selected cultivars. Among Talls, genetic diversity is lowest for the Pacific coast Latin American collections ('Panama Talls') (population 14; He = 0.324; Table 1), concordant with a founder event in their prehistoric introduction from Southeast Asia [7].

Global genetic differentiation and independent origins of domestication

Consistent with earlier molecular marker studies (e.g., [14 18]), our population structure analysis using a worldwide sample set indicates that coconuts are differentiated into two major subpopulations. We performed Bayesian analyses using *Structure* 2.3 [24], with K (the number of putative genetic subpopulations) ranging from 1 to 10, and assessed rates of change in log likelihood values. The optimal value, as determined by the *ad hoc* criterion ΔK [25], was K = 2 (Fig. 1; see also Fig. S1). A secondary ΔK peak at K = 5 suggests further substructure within the major subpopulations (discussed below). An analysis of molecular variance (AMOVA) indicates that 33% of the total genetic variation is partitioned between the two genetic subpopulations (Table S2). This very high level of differentiation suggests long-term evolutionary divergence between the two subpopulations, with independent origins of cultivated coconuts from within each lineage. Moreover, the two genetic subpopulations are structured geographically and are broadly concordant with the 'A' and 'B' groups in the GCP/CIRAD classification scheme (Table 1; Fig. 1). Nearly identical patterns to those observed in the *Structure* analysis are found using *InStruct* [26], a similar Bayesian analysis that relaxes assumptions of random mating within subpopulations (Fig. S2). Taken together, these patterns strongly suggest independent domestication events in the Pacific and Indian Ocean basins.

Human migration and coconut admixture in the Indian Ocean. Historical records suggest that 14 16 centuries ago, Austronesians and Arabs were trading along the oceanic route connecting Southeast Asia to southern coastal east Africa [27]. This route spanned both Pacific and Indian Ocean coconut subpopulations and therefore could have served as an avenue of introgression of Pacific coconuts into the Indian Ocean. The trade route included Comoros and Madagascar, but not the Seychelles, which were among the last islands in the Indian Ocean to be inhabited [8]. Population membership coefficients in our *Structure*

Table 1. Genetic diversity and population structure in a worldwide sample of coconuts[a].

Population (Group)	N (cvs)	Growth Form	Primary Region	H_e	H_o	Q_1	Q_2
1 (A1a)	16 (9)	Dwarf	worldwide	0.270	0.081	**0.966**	0.034
2 (A1b)	32 (7)	Dwarf	SE Asia	0.239	0.099	**0.994**	0.006
3 (A2)	6 (4)	Dwarf	worldwide	0.303	0.000	**0.985**	0.015
4 (A3a)	66 (9)	Tall	SE Asia	0.612	0.532	**0.927**	0.073
5 (A3b)	25 (5)	Tall	SE Asia	0.556	0.428	**0.976**	0.024
6 (A3c)	89 (10)	Tall	SE Asia	0.583	0.447	**0.988**	0.012
7 (A4a)	38 (8)	Tall	PNG[c]	0.607	0.499	**0.990**	0.010
8 (A4b)	34 (8)	Tall	PNG	0.596	0.522	**0.990**	0.010
9 (A4c)	48 (10)	Tall	PNG	0.564	0.484	**0.986**	0.014
10 (A4d)	21 (3)	Tall	PNG	0.610	0.586	**0.991**	0.009
11 (A4e)	360 (10)	Tall	Melanesia	0.624	0.547	**0.980**	0.020
12 (A5)	43 (11)	Tall	Micronesia	0.644	0.508	**0.881**	0.119
13 (A6)	30 (6)	Tall[b]	Polynesia	0.644	0.529	**0.944**	0.056
14 (A7)	105 (5)	Tall	Panama	0.324	0.230	**0.950**	0.050
15 (B1)	150 (18)	Tall	S. Asia+Atlantic	0.483	0.364	0.030	**0.970**
16 (B2)	147 (14)	Tall	E. Africa	0.640	0.570	0.150	**0.850**
17 —	13 (—)	Tall	Comoros	0.672	0.544	0.426	0.574
18 —	44 (—)	Tall	Madagascar	0.691	0.546	0.333	0.667
19 —	55 (—)	Tall	Seychelles	0.413	0.351	0.018	**0.982**

[a]Group labels correspond to GCP/CIRAD designations. N = sample sizes, cvs = number of named cultivars. H_e = mean unbiased gene diversity, H_o = mean observed heterozygosity, and Q_1 and Q_2 indicate subpopulation membership coefficients in *Structure* analyses at K = 2 subpopulations. Bold font indicates membership coefficients of Q≥80%.

[b]includes '*Niu Leka*,' an outcrossing compact-growth variety that is phenotypically distinct from other 'Dwarfs.'

[c]Papua New Guinea.

doi:10.1371/journal.pone.0021143.t001

analysis support the hypothesis of Pacific coconut introgression specifically along the ancient trade route. For coconuts outside of this region (populations 1–15, 19; Table 1), evidence of admixture between the two subpopulations is minimal; >96% of accessions can be assigned unambiguously to either the Pacific or Indian Ocean subpopulation with membership coefficient values of Q≥80% (Fig. 1; Table S1). In contrast, for coconuts from the Comoros and Madagascar (populations 17–18), fewer than one-third of accessions are assigned to the Pacific or Indian Ocean subpopulation at Q≥80%. Similarly, in nearby East Africa (population 16), 23% of accessions show ambiguous assignment (Q<80%). Membership coefficient values assigned at the level of population groupings are also consistent with these patterns of admixture (Table 1).

Introgression from Pacific coconuts into the western Indian Ocean is further reflected in the distributions of individual microsatellite alleles whose frequencies differ between the two major subpopulations and which can therefore serve as subpopulation-diagnostic markers. We identified six such alleles using Shannon's mutual information index (see Methods). Their distributions are very similar across the Indian Ocean, with high coefficients of determination that corroborate the scenario of Pacific coconut admixture (mean $R^2 = 0.866$). To explicitly evaluate the relative contributions of the two subpopulations to the genomes of the putative admixed populations, we calculated a composite introgression index ($\mathcal{I}_i$; Table 2; see Methods). This measure suggests that for Madagascar and Comoros, Southeast Asian admixture accounts for approximately one-half of the genetic variation present in these regions ($\mathcal{I}_i = 0.407$ and 0.509 for Madagascar and Comoros, respectively; Table 2). For East African collections, the level of inferred introgression falls to approximately one-quarter of the total genetic variation ($\mathcal{I}_i = 0.254$). In the Seychelles, outside the Austronesian trade route, no evidence of introgression is observed ($\mathcal{I}_i = -0.065 \approx 0$).

Regional population structure

The presence of a secondary peak of the ΔK *ad hoc* statistic (Fig. S1) prompted us to perform an analysis with K = 5. It revealed substructure that preserves the integrity of the Indo-Atlantic lineage but divides the Pacific group into four components, referred to here as Panama, Dwarf, Papua New Guinea (PNG) and South Pacific (Fig. 2). These names refer to the region (or

Figure 1. Results of *Structure* analysis for a worldwide sample of 1322 coconuts. Population assignments for each accession are shown at K = 2 subpopulations. Numbers along the x-axis correspond to group designations in Table 1. Vertical black lines distinguish the population groups.

doi:10.1371/journal.pone.0021143.g001

Table 2. Assessments of introgression from Southeast Asian coconuts into western Indian Ocean populations[a].

		Allele frequency						
Allele	**Sh**	**A3**	**B1**	**B2**	**COM**	**MAD**	**SEY**	**R^2**
$CnCirA3_{228}$	0.715	0.072	0.97	0.68	0.35	0.424	0.75	0.848
$CnCirC12_{167}$	0.631	0.006	0.834	0.614	0.375	0.465	0.771	0.971
$CnCirE12_{174}$	0.604	0.023	0.85	0.541	0.545	0.394	0.856	0.741
$CnCirF2_{193}$	0.390	0.025	0.67	0.674	0.654	0.625	0.95	0.863
$CnCirE10_{244}$	0.389	0.081	0.767	0.514	0.375	0.512	0.922	0.934
$CnCirC7_{157}$	0.378	0.662	0.027	0.155	0.563	0.279	0	0.839
Mean introgression index (T_I)		**1.000**	**0.000**	**0.254**	**0.509**	**0.407**	**−0.065**	**0.866**

[a]Shannon's mutual information index (Sh), frequencies of six subpopulation-diagnostic microsatellite alleles by population grouping, coefficients of determination (R^2), and mean introgression index values (T_I). Population groups correspond to Table 1. The introgression model assumes admixture between group A3 (Southeast Asia, populations 4–6) and group B1 (Indo-Atlantic, population 15).
doi:10.1371/journal.pone.0021143.t002

coconut type) where they predominate, although most components span multiple regions, as described below.

Table 3a presents pairwise distances calculated in *Structure* (above diagonal) and Jost's [28] relative differentiation (D) (below diagonal) for these five subpopulations. Both measures highlight the genetic isolation of the Indian Ocean from the Pacific populations, consistent with long-term evolutionary divergence between the two lineages. The main interest of Jost's measure is that differentiation and diversity represent structurally independent between- and within-population diversity components. As a

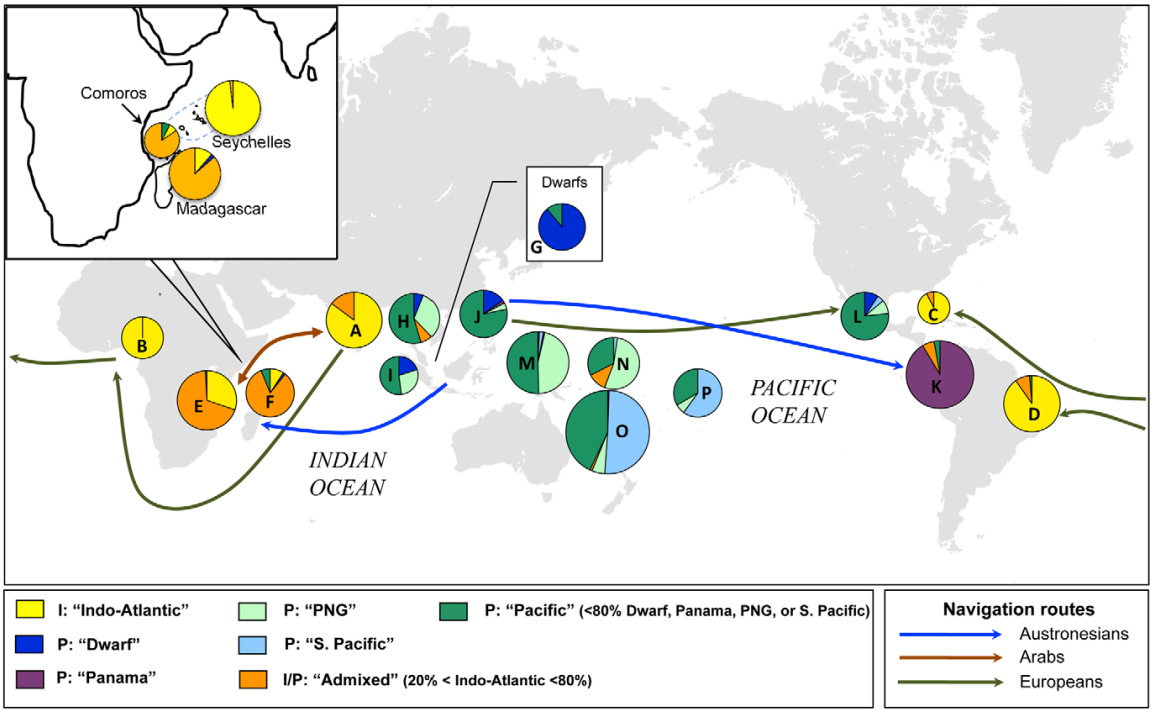


Figure 2. Geographical distributions of Indo-Atlantic and Pacific coconut subpopulations. Subpopulation designations correspond to assignments at Q≥80% membership in *Structure* analyses at K = 5. 'I' and 'P' prefixes in the legend indicate 'Indo-Atlantic' and 'Pacific' population assignments at K = 2 assumed populations (≥80% membership; see Fig. 1). Lines indicate proposed coconut dispersal routes by humans. Pie chart labels correspond to the following countries (ISO abbreviations) and sample sizes: A = IND, LKA, SEY (114); B = BEN, CIV, CMR, GHA (29); C = JAM, MEX (Atlantic) (13); D = BRA (72); E = KEN, MOZ, TZA (116); F = MAD, COM (65); G = Dwarf (54); H = CHN, KHM, MYS, THD, VNM (66); I = IDN (25); J = PHL (46); K = PAN (105); L = MEX (Pacific) (43); M = PNG (141); N = KIT, MHL, TUV (43); O = NCL, SLB, VUT (360); P = COK, FJI, PYF (30). Inset: subpopulation compositions for Madagascar, Comoros, and Seychelles. Pie chart composition is selected to reflect geographical population structure and does not correspond directly to GPC/CIRAD designations in Table 1.
doi:10.1371/journal.pone.0021143.g002

Table 3. Distances ($D_{A,B}$), differentiation (*D*) and diversity parameters for populations identified by *Structure*[a,b].

	Indo-Atlantic (IA)	Panama (PAN)	Dwarf (DW)	PNG	S. Pacific (SP)
a)					
IA	—	0.566	0.469	0.365	0.377
PAN	0.890	—	0.221	0.202	0.221
DW	0.878	0.348	—	0.101	0.129
PNG	0.800	0.363	0.221	—	0.032
SP	0.824	0.396	0.283	0.085	—
b)					
H	0.464	0.264	0.468	0.623	0.620
J	0.536	0.736	0.532	0.377	0.380
Δ	1.866	1.358	1.878	2.655	2.635

[a]pairwise distances (above diagonal) and differentiation measures (*D*, below diagonal) between populations;
[b]expected proportions of homozygotes (*J*), heterozygotes (*H*), and diversity (Δ).
doi:10.1371/journal.pone.0021143.t003

result, the range of variation of *D* between the Indian and Pacific populations (0.800 0890) is much narrower than in the distances (0.365 0.566), which are, by construction, correlated with heterozygosity (see Table 3b). Jost's *D* is also related to Nei's distance measure ($D_{Nei} = -\ln(1-D)$ [29]), which yields values ranging from 1.60 to 2.21 between Indo-Atlantic and Pacific populations. These values are 3.2 4.4 times greater than the largest value between Pacific components (0.504 between Panama and South Pacific), further illustrating that Indo-Atlantic and Pacific coconuts diverged from each other long before any divergence within the Pacific.

To assess the geographical distribution of the five population components, we assigned accessions to one of seven categories based on population membership coefficients at K = 5: accessions with membership coefficients of Q>80% were assigned to each of the five subpopulations (Indo-Atlantic, Dwarfs, Panama, Papua New Guinea, South Pacific); those with 20 80% Indo-Atlantic membership were defined as 'admixed'; and remaining accessions (i.e., those with <20% Indo-Atlantic membership and with <80% membership in any single Pacific subpopulation) were assigned to a generic 'Pacific' class. Figure 2 shows the worldwide geographical distributions of these seven categories. In the descriptions below, letters in parentheses correspond to pie chart labels in Figure 2.

South Asia, Africa and the Caribbean. As is observed at K = 2, the Indian Ocean component predominates in South Asia and the Seychelles (A), as well as in West Africa (B), the Caribbean (C) and Brazil (D) (Fig. 2). Historical records indicate that coconut was unknown in the Caribbean and Atlantic basins until after European colonization [8]; the low level of Pacific admixture in these regions shows that these introductions did not involve admixed populations such as those found today in East Africa (E) or in the western Indian Ocean (F) (Figs. 1, 2). In the admixed populations (E, F), approximately 75% of the Pacific contribution can be assigned to the 'Dwarf' and 'Pacific' population components, consistent with Austronesian introductions from island Southeast Asia (see above; Table S1).

Southeast Asia and Pacific Neotropics. Admixture from the Indo-Atlantic subpopulation is evident at a low frequency in the Pacific coconuts of continental Southeast Asia (H), especially in Thailand, Malaysia, and Cambodia (Fig. 2; Table S1). This pattern may reflect the geographical proximity of these regions to eastern Indian Ocean populations (e.g., Andamans), or longer-distance trading with South Asia (see, e.g., [30]). Interestingly, the 'Dwarf' population component, characteristic of self-pollinating Dwarf cultivars (G), is shared with Talls of Southeast Asia (H, I and J). Previous analyses have suggested that the Dwarf varieties originated the Pacific (e.g., [5]). The present data strongly suggest an origin for these varieties specifically in Southeast Asia.

Pacific coast 'Panama Tall' coconuts (K) are characterized predominantly by the 'Panama' population component. This component is absent elsewhere, except in the Philippines (J) where it occurs at a low frequency (Fig. 2; Table S1). This pattern is consistent with the previously proposed origin of these varieties through a prehistoric introduction from the Philippines [7]. In contrast, the Pacific coast of Mexico (L), which was also populated largely by Philippine coconuts but in post-colonial times and through multiple introductions [2] shows a genetic composition that more closely reflects the genetic heterogeneity of the Philippines (Fig. 2). The small contribution of the 'South Pacific' component in Mexico may reflect early Spanish importations from the Solomon Islands [2].

South Pacific. In Papua New Guinea (M) and in Micronesia (N), the 'PNG' population component predominates. The apparent presence of Indo-Atlantic admixture in Micronesia (N, Fig. 2) may reflect European introductions from South Asia during the period when both regions were under British administration; the shared occurrence of similar green-fruited Dwarf varieties in Sri Lanka and Micronesia (Table S1) is consistent with this hypothesis. To the south and east of Micronesia, the proportion of the 'South Pacific' population component increases. Coconuts in Melanesia (O) are of similar genetic composition to those from Polynesia (P). More than 50% of the individuals in these regions are predominantly of the 'South Pacific' component (Table S1). This includes an outcrossing, compact-growth variety, *'niu leka'* ('Fiji Dwarf'), which represents an independent origin of the dwarf habit, distinct from the widely-cultivated self-pollinating Dwarfs of Southeast Asian origin (Tables 1, S1).

Discussion

Independent domestications of Pacific and Indo-Atlantic coconuts

A striking observation from our worldwide analysis of coconuts is the high level of genetic differentiation between Pacific and Indian Ocean samples (Table 1, Fig. 1; Fig. S2); 33% of the total observed variation is partitioned between the two genetic subpopulations corresponding to the two ocean basins. This finding has several important implications for coconut domestication. First, it makes it clear that *Cocos nucifera* is a native species of both the Indian and Pacific Oceans, with a long-standing evolutionary presence in both ocean basins. Fossil data from the Palaeocene also support the long-term presence of coconuts (or coconut-like species) in both the Indian and Pacific basins [31,32].

In addition, the clear genetic differentiation between the Pacific and Indian Ocean lineages allows us to conclude definitively that coconuts were brought into cultivation independently in each of these regions. In the Pacific, the phenotypic diversity and population heterogeneity associated with a region extending from the Malay peninsula to New Guinea (Table 1, Fig. 2) point to that area as a likely center of domestication. This region ('Malesia') was earlier claimed as the center of domestication for coconut [33]. Island Southeast Asia has also recently been identified as one of several centers of domestication for swine [34], an indication that this was likely an active area of agricultural development. For

Indian Ocean coconuts, archaeological and archaeobotanical findings (coconut shells and sennit rope) from Arikamedu (near Pondicherry) [35], together with Proto-South Dravidian linguistic evidence [36] and ancient Ayurvedic texts [37] suggest that coconuts were already in cultivation in the southern Indian subcontinent around 2,500 3,000 years ago. Our genetic data, when taken together with these other lines of evidence (see also Supporting Information, Text S1; Table S4), suggest that the region encompassing the southern periphery of India, including Sri Lanka, Maldives, and Laccadives, represents a likely center of coconut domestication. These two proposed centers of origin are consistent with those proposed in the 1930s by Vavilov, who also envisioned two centers of origin, one in India and one in the region spanning Indo-China and the Malay archipelago [38].

Interestingly, these two domestication events are associated with markedly different patterns of phenotypic diversification and population substructure. The Indo-Atlantic group shows only moderate gene diversity (Table 1), it is adequately represented by a single genetic subpopulation (Fig. 2), it comprises only the Tall growth form, and its fruit is almost exclusively the elongated (and presumed ancestral) '*niu kafa*' type. This group also remained confined within the Indian Ocean basin until the European colonial era. In contrast, the Pacific group has higher levels of gene diversity (Table 1), it shows evidence of genetic heterogeneity and population substructure that are correlated with its wide geographical distribution (Fig. 2), and it is phenotypically diverse. Pacific coconuts include Talls but are also the source of the widely disseminated, self-pollinating Dwarfs, which our data suggest originated in Southeast Asia (Fig. 2). An additional compact-growth form, the outcrossing Polynesian '*niu leka*' ('Fiji Dwarf') variety, also arose in the Pacific group (Table 1; Table S1). While the Pacific coconut fruit is predominantly of the round '*niu vai*' type, the '*niu kafa*' form is also present, including in Samoa where these names originate. Moreover, unlike the geographically limited Indian Ocean coconuts, Pacific coconuts had become widely distributed throughout the Pacific basin, including the New World tropics, before any European contact. Thus, there is a fundamental asymmetry in the genetic heterogeneity, phenotypic diversity, and regional and global impacts of these two domestication events.

Genetic impacts of coconut dispersal by humans

The genetic distinctness of the Indo-Atlantic and Pacific coconut lineages facilitates our ability to track the genetic footprints of human introductions around the world. Most striking is the genetic admixture in the western Indian Ocean reflecting Pacific coconut introgression. Our analyses suggest that admixed coconuts predominate in the region corresponding to the ancient Austronesian trade route connecting Southeast Asia to Madagascar and coastal east Africa; in contrast, no admixture is evident in the more northerly Seychelles, which fall outside the trade route (Table 2; Fig. 2). The influence of Austronesians along this corridor is well documented [39], perhaps most notably in its lasting impact on human population structure (e.g., [40]). Interestingly, like coconut, a recent study of rice in Madagascar also indicates a shared role for crop varieties originating from Southeast Asia (*japonica* rice) and the Indian subcontinent (*indica* rice), with admixture in Madagascar [41].

Admixture between Pacific and the Indian Ocean coconuts was likely further promoted by the later presence of Arabo-Persian merchants who regularly visited East Africa, trading coconut and favoring its cultivation [42]. Archaeobotanical sources from Pemba [43] show the importance of coconuts from 700 1500 CE in the food culture influenced by Islamic traders in the Indian Ocean. This dual dissemination of the coconut in the Indian Ocean, first by Austronesians and later by South Asians and Arabs, has been well captured linguistically by Allibert [27]: "I have been able to follow the diffusion of the coconut palm from the East to the West, through the Austronesian terms *buahniu* (Bali)/*voanio* (Madagascar), not to mention *vanu* in the Loyalty Islands, but also from *narikela* (Sanskrit)/*nargil* (Arabic, Persian)/*mnazi* (Bantu), a double linguistic pathway for the same tree, the one directly across the Indian Ocean, the other via the north of the same ocean." Recent observations of genetically admixed coconut populations in Oman [44] further support this dissemination history.

Within the Pacific basin, human influence on coconut population structure is most readily detectable in the pre-historic introduction of Southeast Asian coconuts to the New World coast. This introduction is estimated to have occurred ~2,250 years ago, and our analyses are consistent with previous findings suggesting a Philippine origin (Fig. 2; ref [7]); the low genetic diversity in Panama Talls provides further evidence of establishment through a founder event (Table 1). Later European influences are apparent in the Spanish establishment of Mexican populations (see ref [2]); the clear Pacific composition of these coconuts stands in marked contrast to European introductions into the Caribbean and Atlantic basins, which appear to be of Indian origin (Figs. 1, 2; Fig. S2; Table S1). Historical records confirm that the Portuguese established coconut plantations in West Africa, Brazil, and later the Caribbean after Vasco da Gama's 1498 expedition to the Indian Ocean [8]. In the Old World portion of the Pacific basin, our analyses reveal geographical substructure in a pattern that could plausibly reflect human dispersal of coconuts out of the proposed Southeast Asian center of domestication (H, I, J; Fig. 2) and south and east towards Polynesia (M and N; Fig. 2) (see also discussion in ref [45]).

Conclusions

In the most extensive genetic analysis of coconuts to date, we find evidence for independent origins of coconut cultivation in the Pacific and Indian Ocean basins. Interestingly, despite the long-term, extensive movement of coconuts by humans both within and between these oceanic basins, most contemporary coconuts do not show evidence of substantial genetic admixture between the two major genetic subpopulations (Fig. 1; Fig. S2). Given the absence of any known reproductive isolating barriers, the high level of genetic differentiation between these subpopulations suggests a long period of isolation prior to human influence. In this light, the predominance of genetic admixture in the western Indian Ocean (Figs. 1, 2; Tables 1, 3) suggests that humans likely played a prominent role in the establishment and propagation of coconuts in that region.

Besides revealing basic insights into the cultivation and dispersal history of this iconic tropical species, our findings may also facilitate efforts to protect the viability of the coconut as a crop species. Coconut lethal yellowing, a phytoplasma infection, has reached epidemic levels in the Caribbean and other regions of the Neotropics; susceptible trees typically succumb within a year of infection. Knowledge of the worldwide genetic structure of the coconut, including regions where genetic admixture has generated augmented levels of genetic diversity (e.g., Madagascar; Table 1), may ultimately prove useful in targeting source populations for disease resistance and other crop improvement traits.

Materials and Methods

GCP/CIRAD accessions correspond to those in the GCP database (http://gcpcr.grinfo.net/index.php); growth form, vari-

ety name, source country, and germplasm group assignment are indicated in Table S1. An additional 112 coconut palms were sampled from populations occurring on the islands of Madagascar, Comoros and Seychelles. Portions of emerging leaf fronds were collected from the crowns of trees; tissue samples were dried in silica gel desiccant for DNA extraction. Voucher herbarium specimens for the Indian Ocean collections are housed at the Missouri Botanical Garden (MO). Sampled accessions represent 11 locations on Madagascar, 5 on Comoros, and 6 on the Seychelles (Table S1). Genomic DNA was extracted using DNeasy Kits (Qiagen, Valencia, CA) at Washington University.

Genetic analyses were performed using ten microsatellite loci (CnCrF2, CnCrC12, CnCrE10, CnCrA9, CnCrC7, CnCrB6, CnCrE12, CnCrA3, CnCrG11 and CnCrH7). Genotyping of the GCP/CIRAD collection is described in ref [20]. For Indian Ocean accessions, PCR amplifications were performed using similar conditions, and products were separated on an ABI Prism 3130 genetic analyzer at Washington University. Control DNAs with known allele lengths were amplified for all ten loci to standardize scoring of allele sizes. Data were collected and assembled with Genotyper 2.5 software (Perkin Elmer Biosystems).

Genetic Analyses

Analyses of genetic diversity and AMOVA were performed with *GENALEX 6* [46]. To investigate population structure we used Bayesian clustering methods as implemented in *Structure* 2.3 [24] and *InStruct* [26]. *InStruct* is similar to *Structure* but relaxes assumptions of Hardy-Weinberg equilibrium within subpopulations. For *Structure* analyses, the number of subpopulations, K, was set at values ranging from 1 10, with 20 replicate runs apiece (100,000 burnin, 1,000,000 runs). An admixture ancestry model was selected with allele frequencies correlated. For the optimal inferred K value (K = 2), we employed CLUMPP version 1.1.2 [47] to confirm the similarity of clustering memberships among multiple *Structure* runs (the maximum H′ value was >0.9995 at the optimal inferred K value). *InStruct* analyses were performed using the Cornell University BioHPC web portal (http://cbsuapps.tc.cornell.edu/InStruct.aspx). The program *DISTRUCT* [48] was used to visualize outputs from *CLUMPP* and *InStruct* analyses.

Because Dwarf accessions are highly homozygous and show little genetic diversity, clustering analyses were performed both with and without Dwarfs to test for potential artifacts created by their inclusion; excluding these accessions did not substantially alter inferences. In additional analyses, we applied explicit spatial clustering as implemented in *BAPS* [49] and *GENELAND* [50]. However, results were highly biased towards sampling location, a reflection of the pan-global distribution of our dataset, and were not included in further analysis.

Introgression index

To test for Pacific introgression into the Indian Ocean populations, we defined 'diagnostic alleles,' i.e., alleles that are differentially represented in GCP/CIRAD subgroup A3 (a representative Pacific subgroup) relative to subgroup B1 (representative Indo-Atlantic), and we selected them using Shannon's mutual information [51,52] (Table S3). We calculated the entropy of allele a in population A3 as a function of p_{aA}, its frequency in population A3: $h_A(a) = -p_{aA}\log\ p_{aA} - (1-p_{aA})\log(1-p_{aA})$. Likewise, we calculated $h_B(a)$ based on p_{aB} its frequency in population B1 and $h_T(a)$, based on $p_{aT} = \frac{1}{2}(p_{aA}+p_{aB})$. The mutual information quantity between a (the allele) and G (the group) is thus $I(a;G) = h_T(a) - \frac{1}{2}[h_A(a)+h_B(a)]$. Expressed in Shannon units (Sh, using base 2 logarithms), the mutual information quantity may range from 0 (same frequencies in A3 and B1) to 1 (the allele is specific to one population). We retained alleles corresponding to the six top values.

Based on the frequencies of these alleles in six groups (A3, B1, B2, Madagascar, Comoros, and Seychelles), we then calculated 'introgression indices' for each allele: $T_{ia} = (Z_{ia} - X_a)/(Y_a - X_a)$ where X_a, Y_a, and Z_{ia} are the respective frequencies in B1, A3, and the four other groups. Indices i and a refer to group and allele, respectively. The mean of the index over all alleles (T_i) is an estimation of the percentage of alleles from Southeast Asia in each group. Finally, we assessed the consistency of the introgression model by calculating the coefficient of determination R^2 of the regression of the frequencies of each allele on T_i (excluding groups B1 and A3).

Differentiation measures

Jost [29] shows that Nei's heterozygosity (H) and the associated G_{ST} are not adequate measures of diversity and differentiation, respectively. He suggests instead using the reciprocal of Nei's identity as a measure of diversity, and he derives absolute and relative measures of differentiation. These measures are, respectively, $\Delta_{ST} = \Delta_T/\Delta_S = J_S/J_T$ and $D = (J_T/J_S - 1)/[(1/n) - 1]$. In these formulae, $J = 1 - F$ refers to Nei's identity and is the expected proportion of homozygotes in a population. J_S is the average of Nei's identities in the sub-populations. The within-population component of diversity is $\Delta_S = 1/J_S$. The total diversity is $\Delta_T = 1/J_T$ where J_T is calculated based on the allele frequencies in the pooled population. We derived these parameters from the *Structure* outputs (heterozygosities and distances).

Supporting Information

Figure S1 Assessment of subpopulation number in *Structure* analyses.
(DOC)

Figure S2 *InStruct* output at K = 2 subpopulations.
(DOC)

Table S1 Information on coconut accessions used in analyses and assignment probabilities at K = 2 and K = 5 using *Structure* analysis.
(DOC)

Table S2 Allele frequencies for each locus for Pacific and western Indian Ocean populations.
(DOC)

Table S3 Analysis of Molecular Variance (AMOVA) for all coconut accessions (1322 individuals).
(DOC)

Table S4 Language roots associated with the coconut in proto-South-Dravidian and proto-Telugu.
(DOC)

Text S1 Early evidence of coconut use in the southern Indian subcontinent and neighboring islands.
(DOC)

Acknowledgments

We thank Patricia Lebrun for technical support; SCMRT-MPA-Seychelles, CNDRS-Comoros, MO and Tsimbazaza-Madagascar for fieldwork assistance; the staff of Cornell University's BioHPC web computing resource facility for assistance with CLUMPP and InStruct analyses; and members of the Olsen lab group and BFG's PhD advisory committee for helpful discussions.

Author Contributions

Conceived and designed the experiments: KMO LB BFG. Performed the experiments: BFG LB. Analyzed the data: KMO BFG LB. Contributed reagents/materials/analysis tools: KMO LB. Wrote the paper: KMO BFG LB.

References

1. Harries HC (1978) The evolution, dissemination and classification of *Cocos nucifera*. L. Bot Rev 44: 265 319.
2. Zizumbo-Villarreal D (1996) History of coconut (*Cocos nucifera* L.) in Mexico: 1539 1810. Genet Resour Crop Evol 43: 505 515.
3. Batugal P, Rao V, Oliver J, eds (2005) Coconut Genetic Resources. IPGRI - Regional Office for Asia, the Pacific and Oceania, Serdang, Malaysia.
4. Ward G, Brookfield M (1992) The dispersal of the coconut: did it float or was it carried to Panama? J Biogeogr 19: 467 480.
5. Lebrun P, N'Cho Y, Bourdeix R, Baudouin L (2003) Coconut. In: Hamon P, Seguin M, Perrier X, Glaszmann JC, eds. Genetic diversity of cultivated tropical plants. Paris, France: Science Publishers, Inc. and CIRAD, France. pp 219 238.
6. Harries HC, Baudouin L, Cardena R (2004) Floating, boating and introgression: Molecular techniques and ancestry of the coconut palm populations on Pacific islands. Ethnobot Res and App 2: 37 53.
7. Baudouin L, Lebrun P (2009) Coconut (*Cocos nucifera* L.) DNA studies support the hypothesis of an ancient Austronesian migration from Southeast Asia to America. Genet Resour Crop Evol 56: 257 262.
8. Sauer JD (1971) A re-evaluation of the coconut as an indicator of human dispersal. In: Riley C, ed. Man across the sea. Austin, TX: University of Texas Press. pp 309 319.
9. Whitehead RA (1966) Sample survey and collection of coconut germplasm in the Pacific islands (30 May 5 September 1964). London, UK: HMSO, Ministry of Overseas Development.
10. Harries H (1981) Germination and taxonomy of the coconut. Ann Bot 48: 873 883.
11. Bourdeix R, Baudouin L, Billotte N, Labouisse J, Noiret J (2001) Tropical plant breeding. In: André C, Michel J, Serge H, Dominique N, eds. Tropical plant breeding. Montpellier, France: CIRAD. pp 106 127.
12. Mauro-Herrera M, Meerow A, Perera L, Russell JR, Schnell RJ (2010) Ambiguous genetic relationships among coconut (*Cocos nucifera* L.) cultivars: the effects of outcrossing, sample source and size, and method of analysis. Genet Res and Crop Evol 57: 203 217.
13. Meerow A, Noblick L, Borrone JW, Couvreur TL, Mauro-Herrera M, et al. (2009) Phylogenetic analysis of seven WRKY genes across the palm subtribe Attaleinae (Arecaceae) identifies *Syagrus* as sister group of the coconut. PLoS ONE 4(10).
14. Lebrun P, N'Cho YP, Seguin M, Grivet L, Baudouin L (1998) Genetic diversity in the coconut (*Cocos nucifera* L.) revealed by restriction fragment length polymorphism (RFLP) markers. Euphytica 101: 103 108.
15. Perera L, Russell JR, Provan J, Powell W (1999) Identification and characterization of microsatellite loci in coconut (*Cocos nucifera* L.) and the analysis of coconut populations in Sri Lanka. Mol Ecol 8: 344 346.
16. Rivera R, Edwards KJ, Barker JH, Arnold GM, Ayad G, et al. (1999) Isolation and characterization of polymorphic microsatellites in *Cocos nucifera* L. Genome 42: 668 675.
17. Teulat B, Aldam C, Trehin R, Lebrun P, Barker JHA, et al. (2000) An analysis of genetic diversity of the coconut (*Cocos nucifera*) populations from across the geographic range using sequence-tagged microsatellites (SSRs) and AFLPs. Theor Appl Genet 100: 764 771.
18. Perera L, Russell JR, Provan J, Powell W (2003) Studying genetic relationships among coconut varieties/populations using microsatellite markers. Euphytica 132: 121 128.
19. Baudouin L, Lebrun P, Rognon F, Ritter E (2005) Use of molecular markers for coconut improvement. In: Batugal P, Rao VR, Oliver J, eds. Coconut Genetic Resources. Serdang, Malaysia: IPGRI-APO. pp 268 281.
20. Baudouin L, Lebrun P (2002) The development of a microsatellite kit for use with coconuts. Rome, Italy: IPGRI. 66 p.
21. Rajesh MK, Arunachalam V, Nagarajan P, Lebrun P, Samsudeen K, et al. (2008) Genetic survey of 10 Indian coconut landraces by simple sequence repeats (SSRs). Sci Hortic 118: 282 287.
22. Martinez R, Baudouin L, Berger A, Dollet M (2010) Characterization of the genetic diversity of the tall coconut (*Cocos nucifera* L.) population by molecular markers microsatellite (SSR) types in the Dominican Republic. Tree Genet Genom 6: 73 81.
23. Baudouin L, Lebrun P, Berger A, Myrie W, Been B, et al. (2008) The Panama Tall and the Maypan hybrid coconut in Jamaica: Did genetic contamination cause a loss of resistance to Lethal Yellowing? Euphytica 161: 353 360.
24. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945 959.
25. Evanno G, Renaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611 2620.
26. Gao H, Williamson H, Bustamante CD (2007) An MCMC approach for joint inference of population structure and inbreeding rates from multi-locus genotype data. Genetics 176: 1635 1651.
27. Allibert C (2008) Austronesian migration and the establishment of the Malagasy civilisation: Contrasted readings in linguistics, archaeology, genetics and cultural anthropology. Diogenes 55: 7 16.
28. Jost L (2008) G_{ST} and its relatives do not measure differentiation. Mol Ecol 17: 4015 4026.
29. Jost L (2009) D vs. G_{ST}: Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Mol Ecol 18: 2088 2091.
30. Cooper Z (2002) Archeology and history: early settlements in the Andaman Islands. Oxford, UK: Oxford Univ. Press. 256 p.
31. Rigby JF (1995) A fossil *Cocos nucifera* L. fruit from the latest Pliocene of Queensland, Australia. In: Pant DD, ed. Global environment and diversification of plants through geological time, Society of Indian Plant Taxonomists, Allahabad University, Allahabad, India: Birbal Sahni Institute, Centennial Volume. pp 379 381.
32. Tripathi RP, Mishra SN, Sharma BD (1999) *Cocos nucifera*-like petrified fruit from the Tertiary of Amarkantak, M.P., India. Paleobotanist 48: 251 255.
33. Harries HC (1990) Malesian origin for a domestic *Cocos nucifera*. In: Baas P, Kalkman K, Geesink R, eds. The Plant Diversity of Malesia. Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 351 357.
34. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, et al. (2005) Worldwide phyloeography of wild boar reveals multiple centers of pig domestication. Science 307: 1618 1621.
35. Wheeler REM, Ghosh A, Deva K (1946) Arikamedu: an Indo-Roman trading station on east coast of India. Ancient India: Bulletin of the Archaeological Survey of India 2: 17 124.
36. Fuller D (2007) Non-human genetics, agricultural origins and historical linguistics in South Asia. In: Petraglia MD, Allchin B, eds. The evolution and history of human populations in South Asia. Dordrecht, The Netherlands: Springer. pp 393 443.
37. Menon KP, Pandalai KM (1958) The coconut palm, a monograph. Kerala, South India: Indian Central Coconut Committee. 384 p.
38. Chester K (1951) The origin, variation, immunity and breeding of cultivated plants. Selected writings of N. I. Valvilov; trans Chester KS. New York, NY: The Ronald Press Company. 364 p.
39. Spriggs M (1997) The Island Melanesians. Oxford, UK: Wiley-Blackwell Press. 320 p.
40. Hurles M, Skyes B, Jobling M, Forster P (2005) The dual origin of the Malagasy in Island Southeast Asia and East Africa: Evidence from maternal and paternal lineages. Am J Hum Genet 76: 894 901.
41. Mather KA, Molina J, Flowers JM, Rubinstein S, Rauh BL, et al. (2010) Migration, isolation and hybridization in island crop populations: the case of Madagascar rice. Mol Ecol 19: 4892 4905.
42. Schuilling M, Harries HC (1994) The coconut in East Africa 1. East African Tall. Principes 38: 4 11.
43. Walshaw S (2010) Converting to rice: urbanization, Islamization and crops on Pemba Island, Tanzania, AD 700 1500. World Archaeol 42: 137 154.
44. Perera L, Baudouin L, Bourdeix R, Bait Fadil A, Hountoundji FCC, et al. (2011) Coconut palms on the edge of the desert: genetic diversity of *Cocos nucifera* L. in Oman. CORD 27: 9 19.
45. Hurles M, Matisoo-Smith E, Gray R, Penny D (2003) Untangling Oceanic settlement: the edge of the knowable. Trends in Ecology and Evolution 18: 531 540.
46. Peakall R, Smouse PE (2006) *GENALEX 6*: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288 295.
47. Jakobsson M, Rosenberg NA (2007) *CLUMPP*: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801 1806.
48. Rosenberg NA (2004) *DISTRUCT*: a program for the graphical display of population structure. Mol Ecol Notes 4: 137 38.
49. Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics 20: 2363 2369.
50. Guillot G, Estoup A, Mortier F, Cosso J (2005) A Spatial Statistical Model for Landscape Genetics. Genetics 170: 1261 1280.
51. Shannon C (1948) A mathematical theory of communication. Bell System Tech J 27: 379 423, 623 656.
52. Battail G (1997) Théorie de l'information, Application aux techniques de communication. Paris, France: Masson. 396 p.

CHAPTER 5

GENOME SIZE VARIATION IN ATTALEINAE (ARECACEAE) WITH EMPHASIS ON COCONUTS (*COCOS NUCIFERA* L.) [1]

BEE F. GUNN[2], THIERRY BEULÉ[3], LUC BAUDOUIN[4], PASCAL ILBERT[3], CHRISTOPHE DUPERRAY[5], MICHAEL CRISP[2], AUGUSTE ISSALI[6], JEAN-LOUIS KONAN[6] AND ALAIN RIVAL[3,7]

[2] Research School of Biology, Division of Evolution, Ecology and Genetics, The Australian National University, Canberra, ACT 0200, Australia

[3] CIRAD, UMR DIADE, 34398 Montpellier, France

[4]CIRAD, UMR AGAP, 34398 Montpellier, France

[5]Montpellier Rio Imaging, IRB, INSERM U1040, CHU de Montpellier, 34295 Montpellier, France.

[6] Station de Recherche Marc Delorme, PortBouët. 07 BP 13 Abidjan, Côte d'Ivoire.

The authors thank the Centre National de Recherche Agronomique de Côte d'Ivoire for providing fresh coconut samples, Spencer Brown (CNRS, France) for providing the selected *Petunia* seeds and the IRB (Institut de Recherche en Biothérapie) in Montpellier for kindly providing access to the flow cytometry facilities at Montpellier Rio Imaging. The Australian National Botanic Gardens (Canberra) are gratefully thanked for providing the coconut leaf material from Lizard Island and the Royal Botanic Gardens (Sydney) for material from Mission Beach and Malayan Yellow Dwarf. The Authors acknowledge editorial support from Prof. Patricia Blackwell, Research Sch. of Biol., Div. of Evolution, Ecology and Genetics, The Australian National University.

[7] Corresponding author. Research School of Biology, Division of Evolution, Ecology and Genetics, The Australian National University, Canberra, ACT 0200, Australia, (bee.gunn@anu.edu.au)

5.1 ABSTRACT

- *Premise of the study:* The C-value or a species' nuclear DNA content has significant evolutionary associations with growth, development and adaptation to environmental changes. Angiosperm C-values range 1200-fold and intraspecific variations occur frequently in commonly cultivated plants, but little is known about domestication impacts on genome size. Here we examined genome sizes representing coconut genotypes worldwide and members of the Attaleinae (Arecaceae). Our objectives were to 1) estimate the coconut's genome size, 2) determine intraspecific DNA ploidy levels, 3) to test whether Dwarf genome size is less variable than Tall cultivars and 4) to reconstruct ancestral genome sizes of the Attaleinae.
- *Methods* We used flow cytometric analysis of isolated nuclei in order to estimate genome size from young palm leaf material. Ancestral genome size reconstruction was based on maximum likelihood phylogeny from sequences of seven *WRKY* loci.
- *Key Results* The coconut's genome size show intraspecific variation associated with domestication. Variation among Tall cultivars was significantly different compared to Dwarfs. Comparison of Attaleinae genomes showed moderate variation across genera, except for *Jubaeopsis caffra, Voanioala gerardii, Beccariophoenix alfredii* and *Polyandrococos caudescens* for which polyploidy led to increased genome sizes.
- *Conclusions* Results contribute to the understanding of domestication on genome size of long-lived tree crops, and have important implications for implementation of whole genome sequencing of the coconut and other domesticated plants. Polyploidy evolved independently in two clades within Attaleinae.

- **Key words:** Attaleinae; C-value; Cocos nucifera L.; domestication; flow cytometry, genomic evolution; genome size, nuclear DNA content; ploidy.

5.2 INTRODUCTION

The nuclear DNA content of a species has major effects on the growth, meiotic and mitotic cycles and expansion of cells. Therefore DNA content affects the individual's morphological and physiological development as well as adaptations to its environment (Price and Baranova, 1976; Bennett, 1998; Knight, Molinari, and Petrov, 2005). The C-value (holoploid genome size) of a species corresponds to the DNA amount in its unreplicated haploid or gametic nucleus (pollen or sperm), regardless of its ploidy level (Swift, 1950; Greilhuber et al., 2005) and is measured in picograms (pg) or base pairs (bp). Large variation in C-values may have consequences or costs to the organisms and several studies have shown that C-values are often associated with ecological constraints (Bennett, 1987; Knight, Molinari, and Petrov, 2005), temporal shifts in phenology (Grime and Mowforth, 1982), sensitivity to ionizing radiations and climatic changes (Sparrow and Miksche, 1961; Sparrow and Sparrow, 1965; Sparrow, Schwemmer, and Bottino, 1971). The minimum generation time (MGT), defined as the time from germination until the first production of mature seed is positively correlated with the C-value of the species, suggesting that species with smaller genomes have shorter generation times. In flow cytometry, C-values are estimated from the dominant G1 peak of fluorescence and ploidy levels may also be detected by the numbers of dominant peaks.

The C-value is equivalent to genome size in diploid species but is always greater than the genome size(s) in polyploids (Bennett, Bhandol, and Leitch, 2000). Indeed, a diploid plant has two genomes, after gametic fertilization, whereas a polyploid has more than two genomes as a result of either autopolyploidization, allopolyploidization or hybridization (Stebbins, 1959). Polyploidy is known to occur among 80% of angiosperms (Masterton, 1994). It is an important phenomenon in the evolution of higher plants (Leitch and Bennett, 1997) and a driving force in evolution (Rieseberg,

2003). Polyploidy is also common in domesticated plants where it is detectable in major crops such as cereals (wheat and rye), maize, cotton, potato, banana, sugar cane and coffee (Gaut and Doebley, 1997; Wendel and Cronn, 2003; Heslop-Harrison and Schwarzacher, 2007) and adds complexity to identifying the wild ancestors of the domesticate (Olsen and Wendel 2013). Understanding the impacts of ploidy levels on the genome size is informative since gene duplications can play an important role in epigenetic gene silencing or expression and provide protection against harmful viruses and transposons (Pichersky, 1990).

Chromosome numbers (2n), C-values and ploidy levels are tightly linked and remain constant for most species; nevertheless, there are exceptions where variations do occur. Intraspecific variation in C-values is not rare despite having no change in chromosome number. Domesticated crops such as *Zea mays* (all with 2n = 20) showed 37% variation among the cultivar lines (Laurie and Bennett, 1985) and *Poa annua* (2n = 28) showed 80% variation (Grime, 1983). The switchgrass, *Panicum virgatum* L. is a North American native perennial cultivated for pastures, rangelands and fuel biomass. Cytological studies reveal that it is a polyploid series from diploid (2n = 18) to dodecaploid (2n = 12C = 108) (Church, 1940; Riley and Vogel, 1982).

Angiosperm C-values range from 0.1 to 127.4 pg (Bennett, Bhandol, and Leitch, 2000), and each value is characteristic of a given species. The palm family (Arecaceae) is among the most diverse, with C-values ranging from 0.9 to 30 pg (Angiosperm 1C-values database (http://data.kew.org/cvalues/). Within the Cocoseae tribe *Voanioala gerardii* J. Dransf., a polyploid (1C-value = 30 pg; n= ca. 300) shows the highest C-value. *Syagrus* and *Attalea* sister clades of *Cocos nucifera* (Meerow et al. 2009, 2014) are Neotropical and highly speciose.

The coconut palm *Cocos nucifera* L. (Arecaceae) is cultivated globally on over 12 million hectares in the humid tropics. *C. nucifera* L. (2n = 32) (Nambiar and

Swaminathan, 1960; Abraham and Mathew, 1963) and is the only species of its genus. Coconuts are economically important for millions of people depending on this palm for their livelihoods (Batugal, Bourdeix, and Baudouin, 2009). *Cocos nucifera* is best regarded as a semi-domesticated species, a complex of local populations with all degrees of dependence upon man, from nil to complete (Sauer, 1971). Although Harries (1978) distinguishes "domesticated" and "wild" coconuts, this distinction refers to an ancient domestication event and he acknowledges that both types are indifferently cultivated nowadays. Wild populations do exist but only in a few locations (Foale, 2005) but some of them might be feral ie: formerly cultivated population surviving spontaneously.

At the other end of the range, Dwarf coconut can be regarded as the most completely domesticated type (Gunn, Baudouin, and Olsen, 2011). This coconut type is usually grown near human habitations and account for only 5% of coconuts globally (Bourdeix et al., 2009). Its self-pollinating habit makes it possible to propagate a desirable genotype true to type, and to screen rare off-types based on recognizable phenotypic markers such as fruit color and shape. The Dwarf coconut is precocious and becomes reproductive usually after four years. It is especially appreciated from the water of its immature nuts and its slow growth makes harvesting relatively easy for most of its relatively short lifespan (ca. 35 years). Finally, it is dependent on human cultivation because it is a poor competitor in natural stands or in mixed plantings due to its short lifespan and to its reduced vigor. The Tall coconut lacks most of the "domesticated" features found in the Dwarf. It is predominantly cross-pollinating and highly heterozygous. It is fast growing and becomes reproductively mature later, usually after seven years and lives for 70 years or more. In some cases, the influence of selection under cultivation besides Talls and Dwarfs, relatively rare types are observed: Semi-Talls are self-pollinating like Dwarfs but relatively more robust. The "compact

Dwarf" represented by the Niu Leka Dwarf from the South Pacific is not related to the other Dwarfs. It is cross-pollinating, and as vigorous as a Tall but and owes its small size to a marked reduction in internode length and in the distance between leaflets.

To date, genome size estimates exist only for 3% of palm species, principally based on Feulgen-microdensitometry methods. Flow cytometry has become the predominant method not only for ploidy studies and determination of absolute DNA contents of cells, due to its high sample throughput and relative ease of sample preparation (Dolezel and Bartos, 2005; Dolezel, Greilhuber, and Suda, 2007). Intraspecific genome size has been shown to vary between cultivars and wild progenitors in Angiosperms (Greilhuber, 2005), and such subtle changes may be detected only when using flow cytometry.

In coconut, genome size values have been recorded, ranging from 5.1 pg (Röser, Johnson, and Hanson, 1997); unknown coconut variety, root tips) to 5.6 ± 0.2 pg ((Sandoval, Hocher, and Verdeil, 2003); Malayan Yellow Dwarf, callus tissue).

Determination of the genome sizes of cultivated coconuts and ploidy level are essential prerequisites for sequencing the coconut genome. This will provide precise calculation for the optimal depth of reads required and accurate assembly and annotations of the coconut genome. Genome sequences have been recently generated and made publicly available for two palm species of major economic importance: the date palm (Al-Dous et al., 2011) and the oil palm (Singh et al., 2013). For the coconut palm, future genome sequencing will be integral to identifying genes responsible for disease resistance and many other genes of agro-ecological interest such as drought or salt tolerance. The integration of gene discovery and Marker Assisted Breeding will pave the way for the generation of new coconut cultivars, which will be better adapted to changing agro-climatic conditions.

It is not known if the phenotypic differences such as dwarf habit and fruit morphology between the Dwarf and the Tall cultivars and their generation times (three vs seven years) would have impacted their genome size. In this study, we explored genome size variation using flow cytometry of 21 coconut cultivars including two wild-sown coconuts representing a total of 23 genotypes from across the globe. Our objectives were to: 1) estimate the coconut's genome size, 2) determine intraspecific DNA ploidy levels, 3) test whether Dwarf genome size is less variable than Tall and 4) reconstruct ancestral genome sizes of the Attaleinae.

5.3 MATERIALS AND METHODS

5.3.1 ***Plant Material***—We sampled immature leaves from 23 adult palms originating from 23 coconut populations chosen to cover the genetic diversity (Appendix 1). Two of them were self-sown, putatively wild, populations from Australia (Mission Beach, lat. -17.869121°, long. 146.106338° and Lizard Island, lat. -14.667717°, long. 145.446729°). The others were traditional and advanced cultivars from the collection maintained at the Marc Delorme Research Station (CNRA Côte d'Ivoire). They include seven self-pollinating Dwarf cultivars, 15 cross- pollinating Tall cultivars, one cross-pollinating "compact Dwarf" cultivars and three population hybrids (one Tall × Tall and two Dwarf × Tall).

Fresh leaf material was collected from the unopened spear leaf of the palm whenever possible. In addition, we sampled leaf material for 16 species across 9 genera of the Cocoseae: *Attalea*, *Bactris*, *Beccariophoenix*, *Butia*, *Elaies*, *Jubaeopsis*, *Lytocaryum, Polyandrococos* and *Sygarus* from the living collections of the Royal Botanic Gardens, Sydney, Australia. We wrapped approx. 4 cm length of each leaf in moistened tissue paper and placed it into an envelope kept at 4°C to preserve it during transportation to the IRB laboratory in Montpellier, France.

5.3.2 ***Estimation of 2C-value***—To determine genome size, we first used razor blades to chop coconut and *Petunia hybrida* E. Vilm. leaves in order to extract nuclei. The *P. hybrida* Px PC6 (Vilmorin), 2C = 2.85pg was grown in the greenhouse and used as calibration standard following Coba de la Peña and Brown (2001). Approximately 1 cm^2 of fresh leaves were chopped in 500 μL of Dolezel's lysis buffer (Dolezel, Binarova, and Lucretti, 1989) with the following modifications: no spermine was added and we replaced β-mercaptoethanol with 10 mM sodium metabisulphite which was added immediately before use (Rival et al., 1997). The lysate was then filtered through disposable filters using 20 μm nylon mesh (Partec CellTrics®) in order to isolate nuclei from cell debris and aggregates. Then 500 μL of the filtrate was pipetted into a new disposable tube and 20 μL of DAPI (4',6-diamidino-2-phenylindole, dihydrochloride) fluorochrome solution (0.1mg mL^{-1}) was added, for a final DAPI concentration of 4 μg mL^{-1}. After homogenizing and stabilizing for 5 minutes at room temperature, the stained nuclei suspensions were analysed.

We measured the relative fluorescence intensities from stained nuclei using a Beckman-Coulter CyAN^{TM} ADP flow cytometer (Beckman Coulter Inc., U.S.A.) with at least 500 nuclei analyzed per run. We repeated measurements of the G1 peaks (non-replicated phase of the cell cycle) for each coconut cultivar 3-5 times with internal standards and used the means (μ ± s.d.) in our assessment of the absolute value of the coconut's genome size, yielding graphical outputs such as illustrated in Figure 1.

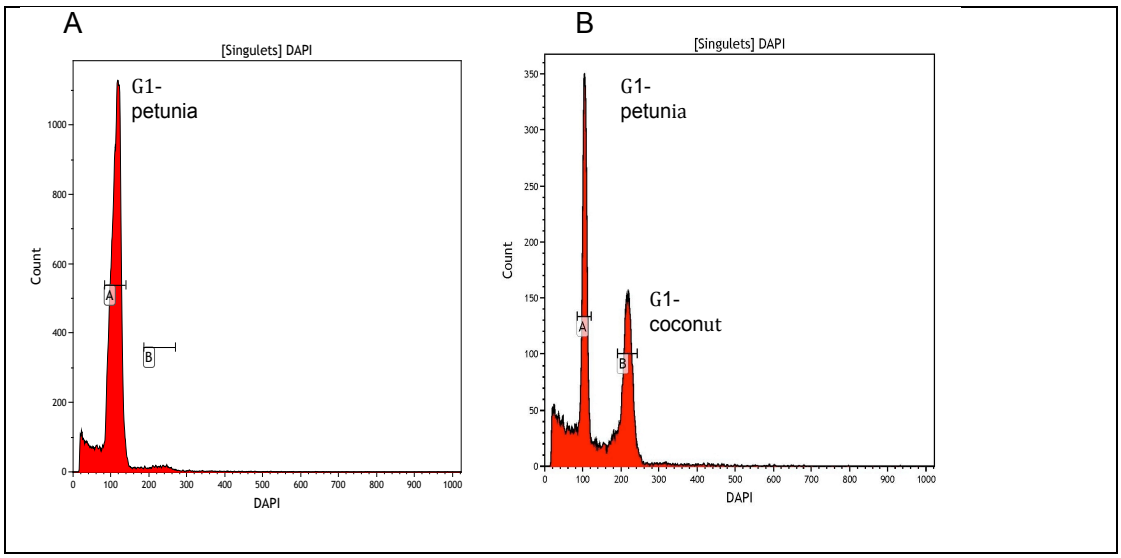


Fig. 1. Examples of flow cytometry histograms. A: Peak A: *Petunia* standard alone; B: Peak A: *Petunia* standard, Peak B: *Cocos nucifera* L. G1 represents the non-replicating cell phase.

5.3.3 ***Data Analysis***—The first step of data analysis consisted of a visual examination of the cytometer plots (Fig. 1) to exclude unreliable runs (i.e. the observations with a low signal to noise ratio, mainly due to insufficient quality of plant material). Calculations and graphical representation were carried out with R software (Chambers et al., 1983; R Development CoreTeam, 2011). The proportionality of the DAPI values between the coconut genotypes and the internal standard (*Petunia hybrida*) was checked through regression analyses to determine the correlation between the DAPI values of the internal standard and coconut genotypes. Genome size for each sample was estimated as $G_C = D_C/D_S * G_S$ where D_C is the DAPI value of coconut, D_S is the DAPI value of the standard, and G_S is the genome size of the standard (2.85 pg for *Petunia*). We examined variation in genome size among cultivar and species using ANOVA and we applied the F-test to determine the significance of the values. We tested for possible effects of domestication on genome size of *Cocos nucifera* by forming two groups: Tall, and Dwarf again using ANOVA. Finally, we used boxplots to visualize the variation in DNA amount in the Dwarf, and Tall coconut ecotypes.

5.3.4 ***Ploidy level***—Ploidy in flow cytometric assays equates to a constant DNA quantity (C value) of the complete chromosome complement with respect to a published reference standard of known ploidy. We determined the ploidy level of the coconut from the positions of the G1 peaks in cytometry histograms. The presence of polyploidy is reflected in the position of the dominant G1 peak and the appearance of more than one dominant peak apart from the internal standard.

5.3.5 ***Evolution of 2C value in Attaleinae***—We estimated the absolute genome size of the 13 species using flow cytometry and obtained C-values for an additional five species from the Angiosperm 1C-values database (Appendix 2). To determine the topology of the evolutionary tree of the *Attaleineae,* we used seven *WRKY* nuclear loci from Meerow et al. (2009), concatenated to sequence length of 5.648 kb for 56 taxa across the Attaleinae available from Genbank. We conducted maximum likelihood analyses using PHYML software (Guindon and Gasceul, 2003) implemented through Geneious 6.1.7 (Biomatters Dev. Team 2013) with the following criteria: initial BioNJ tree, NNI topology search, GTR substitution model, discrete Gamma model, 4 categories, random seed and 100 bootstrap replicates.

We applied the maximum likelihood approach as described in Pagel et al. (1999) for ancestral character reconstruction as implemented in Mesquite. The maximum likelihood trees (100) were imported into Mesquite Version 2.5 (Maddison and Maddison, 2008) and a character matrix of 2C values for 19 taxa were appended to the DNA sequences. We traced the 2C values sizes as continuous characters on to the ML tree in order to infer ancestral state likelihoods. We used *Bactris* and *Elaeis* as outgroups for the non-spiny Attaleinae.

5.4 RESULTS

5.4.1 ***Proportionality of DAPI values with internal standard***—The results from the regression analysis of the DAPI values for the coconuts against the internal standard

(*Petunia hybrida*) were highly correlated (corrected R^2 =0.9997 when the intercept was fixed to 0) confirming their proportionality. The proportionality coefficient was 2.0921 ± 0.0041 (mean ± s.e.). This enabled the use of the ratio of the coconut DAPI values to the internal standard to calculate the absolute genome size of the coconut ecotypes (see Appendix 1).

The observed value of DNA contents ranged between 5.720 and 6.250 pg, 50% of them being in the 5.915 – 6.020 pg range. We noted that the variation of genome sizes of the Dwarfs is about half that of the Tall ecotypes. The summary statistics for DNA contents in the two groups of coconuts are given in Tables 2A-C and illustrated in Fig. 2.

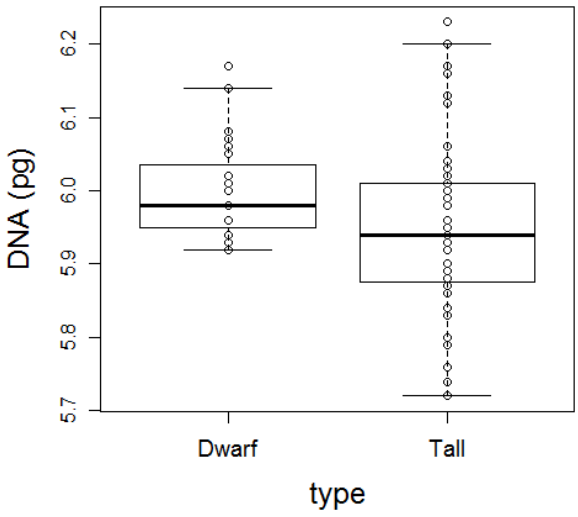


Fig. 2. Boxplot of estimated nucleus DNA content. The thick horizontal line corresponds to the median, the limits of the boxes are the first and the third quartiles. Individual observations are represented by dots.

5.4.2 ***Variation of genome size in coconut***— The overall mean of genome size was 5.963 pg. We performed an analysis of variance (ANOVA) based on 16 Tall and 7 Dwarf coconuts (Table 1). The residual standard deviation was 0.0641 pg. This represents the uncertainty due to the breadth of the peaks and to random fluctuations of

the experimental conditions. In average, Tall and Dwarf coconuts differed in genome size (F = 11.33, P value = 0.001). There were also significant differences among Talls (F=10.43, P value= 2.78 10^{-11} but Dwarfs were not significantly different (F=1.34, P value = 0.254). The estimated mean and confidence interval (α=0.05) of genome size are 6.00 [5.97 – 6.03] and 5.95 [5.74 – 6.16] in Dwarfs and Talls respectively. This takes into account both empirical errors and the estimated variance of genome size (in Talls). Although the genome size in Dwarf is superior to the *average* genome size of Talls, it remains within the range of Tall coconuts. It is also the case of the three additional individuals we sampled in population hybrids (one Tall × Tall 2C=6.13 pg and two Dwarf × Tall, 2C=5.90 and 5.92 respectively).

Our results reveal limited (CV=2%) but significant variation in genome size in coconut. These variations occur both in the Indo-Atlantic and in the Pacific genetic groups (respective $\alpha = 0.05$ confidence intervals [5.79 – 6.25] and [5.76 – 6.04]), but not among Dwarfs.

Table 1. ANOVA of estimated 2C-values (pg).

	Df	*Sum Sq*	*Mean Sq*	*F value*	*Pr(>F)*
Between types	1	0.04664	0.04664	11.33	0.00138
Within Dwarf type	6	0.03310	0.00552	1.34	0.25479
Within Tall type	15	0.64373	0.04292	10.43	2.775×10^{-11}
Residuals	56	0.23043	0.00412		

5.4.3 *Genome size of Attaleinae*

Within the Attaleinae subtribe, the holoploid genome sizes were as follows: *Voanioala gerardii* = 60 pg (Johnson et al. 1989), *Allagoptera caudescens* (Mart.) Kunze = 10.70 pg, *Attalea* sp. = 4.02 – 4.34 pg, *Butia* sp. = 3.06 – 3.42 pg, *Beccariophoenix sp*. = 3.6 – 7.47 pg, *Cocos nucifera* = 5.966 ± 0.111pg, *Jubaeopsis caffra* Becc. = 20.98 pg, *Lytocaryum weddellianum* (H. Wendl.) Toledo = 3.72 pg and

Syagrus sp. = 3.9 – 6.9 pg. The genome size of *Beccariophoenix madagascariensis* Jum. and H. Perrier was 3.6 pg whilst that of its sister taxon *Becc. alfredii* Rakotoarin et al. was almost twice (7.47 pg) suggesting that the latter is a tetraploid.

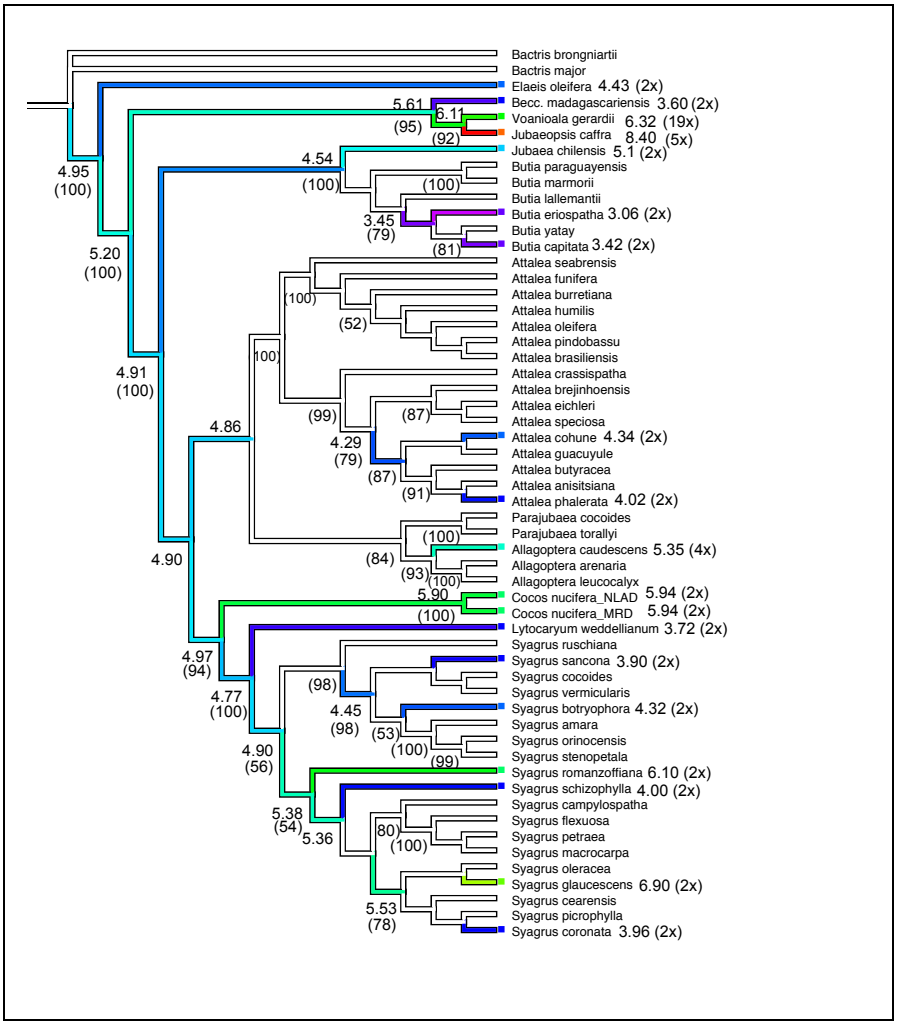


Fig. 3. Ancestral genome size reconstruction: Maximum likelihood phylogenetic tree of Attaleinae based on seven *WRKY* nuclear loci using PhyML (Phylogenetic Analysis of Maximum Likelihood). ML bootstrap supports are in parenthesis below the branches. Sequence alignment will be deposited in Dryad database (http://datadryad.org/). The numbers at the nodes refer to the inferred ancestral genome sizes using maximum likelihood reconstruction approach implemented in Mesquite. Numbers adjacent to the OTUs are the holoploid genome size (2Cx) estimated using flow cytometry with ploidy levels in parenthesis, where 2x denote diploids and >2x denote polyploids. Outgroups included were *Elaeis oleifera*, *Bactris major* and *B. brongniartii*.

5.4.4 ***Ancestral genome size*** **(2Cx)** ***of Attaleinae***—The ancestral genome size of the most recent common ancestor (TMRCA) based on the maximum likelihood topology (Fig. 3) was 4.15 pg and was 4.55 pg for the African/Malagasy and South American clades. The most recent common ancestor of *Beccariophoenix* and *Voaniaola* + *Jubaeopsis* was 4.89 pg and the inferred genome size for TMRCA oc *Voaniaola* + *Jubaeopsis* was 6.53 pg. The inferred ancestral genome size for *Cocos nucifera* was 5.61 pg. The genome size of TMRCA of the *Cocos*/*Syagrus* clades was 4.95 pg and for paraphylectic *Syagrus*, the genome size of the TMRCA of the two major clades was 4.43 pg. As expected, the genome sizes of the speciose *Syagrus* showed some variation between the Rainforest and Eastern Brazil clades. TMRCA of *Attalea /*(*Allagoptera* + *Polyandrococos* + *Parajubaea*) clades was 4.85 pg (Fig. 3). Genome size among *Butia* appears to be smallest (3.06 pg) with inferred ancestral genome size leading to the TMRCA of (*Jubaea chilensis* + *Butia*) clade as 4.45 pg, showing a reduction in *Butia* but an increase in the closely related *J. chilensis* (5.1 pg).

5.5 DISCUSSION

Plant domestication is an evolutionary process that involves artificial selection and may lead to population bottlenecks that can reduce the genetic diversity relative to the wild progenitors through selection of preferred phenotypes (Doebley, Gaut, and Smith, 2006). In the case of coconuts, as shown by our comparison of Tall and Dwarf cultivars, human selection for traits such as dwarfism, precocity and higher water contents may have affected the patterns of their genome architecture (Olsen and Wendel 2013). Meiotic abnormalities occur at a higher percentage in Dwarf than in Talls and may be associated with the shift from out-crossing in the Tall ecotypes to predominantly self-pollination in Dwarfs (Swaminathan et al. 1961). The consequences of this shift in breeding systems are observed in the poor endosperm development and reduced vegetative vigor in Dwarfs (Swaminathan and Nambiar, 1961). Dwarfism in

coconuts may be due to pleiotropic gene effects and preliminary studies have suggested that dwarfism involves at least five independently segregating genes. (Baudouin, unpubl. res.).

Tanksley (2004) has shown that in domesticated plants such as the tomato, the wide range of fruit morphologies and phenotypic variation are controlled by only four QTLs encoding for fruit shape and size. In contrast, Baudouin et al. (2006) found that 34 putative QTLs were accounted for by six pleiotropic genes associated with traits regulating coconut fruit component whilst six QTLs were detected for precocity or early germination trait (Herran et al., 2000).

Our screening has demonstrated that the DNA contents in cultivars and wild-sown coconut genotypes are variable. Our results show that Dwarfs express significantly less variation in genome size compared to Talls. One possible explanation is that all Dwarf cultivars originated from a single Tall associated with the shift from allogamy to autogamy and thus likely to have the genome size similar to the ancestral Tall. This is consistent with the effect of a domestication bottleneck reducing the genetic diversity accompanying the process of artificial selection for traits related to the Dwarf phenotype in combination with retention of a high proportion of genetic variation in the Tall ecotypes (eg., (Miller and Gross, 2011). We did not discover any tetraploids in coconuts, in contrast to other domesticated plants such as maize, wheat, barley, rice and cotton, where DNA polyploidy occurs commonly among cultivars (Laurie and Bennett, 1985; Gaut and Doebley, 1997; Wendel and Cronn, 2003; Zhang et al., 2005).

Leitch et al. (2005) examined genome size data for 4,538 angiosperms and used time of divergence to reconstruct the ancestral genome size. In the Attaleinae, there is some variability in genome size at the generic level but overall there is conservation of genome size at the interspecific level (Fig. 3). The Attaleinae is monophyletic and

includes all members of the Cocoseae except the spiny cocosoids (Bactridinae) and, *Barcella* and *Elaeis* (Elaeidinae), (see (Dransfield et al., 2008).

A study by Shapcott et al. (2007) on the genetic diversity of *Becc. madagascariensis* found highly inbred populations. *Becc. alfredii,* although a distinct taxon, microsatellite data did not show differentiation from the northern *Becc. madagascariensis* population. It is possible that selfing within these northern populations potentially led to polyploidy with subsequent dispersal by frugivores to new habitats for speciation process. Evidence from the current study indicated that two other members of the Attaleinae: *Polyandrococos caudescens* and *Beccariophoenix alfredii* were polyploids suggesting evolution of polyploidy occurred at least four times within the Cocoseae.

The outgroup *Roystonea* (tribe: Areceae) has a genome size of 9.6 pg (Röser, Johnson, and Hanson, 1997). The Cocoseae tribe diverged from its closest relatives *Roystonea* /*Reinhardtia* ca. 55 – 58 mya (Gunn, 2004; Roncal et al., 2013). These data suggest that the ancestral genome size for the Attaleinae may have been small (ca. 4.80 pg) and *Polyandrococos caudescens*, *Becc. alfredii* Rakotoarin *et al.*, *Jubaeopsis caffra* and *Voanioala* have undergone polyploidization events in the past and have retained their duplicated genomes. The African oil palm *Elaeis guineensis* Jacq., subtribe Bactridinae, has a genome size = 3.76 ± 0.09 pg (Rival et al., 1997) which is two-thirds that of the coconut.

We inferred the ancestral genome sizes across the Attaleinae using a maximum likelihood approach (Pagel, 1999). The Attaleinae diversified in South America and for the highly speciose taxa such as *Syagrus, Attalea* and *Butia,* genome size shows variation at the generic level. Their genome sizes are much smaller than the species poor Malagasy/African clade and it is possible that small genome size have played a role providing competitive advantages for these South American taxa to diversify into

different biomes as small genome size has been shown to correlate with shorter minimum generation time (MGT), increased reproductive rate and reduced reproductive costs especially in perennial diploid monocots (Bennett, 1972; Midgley, 1991).

In a study on the nuclear content among 411 Angiosperms species using flow cytometry, Zonneveld et al. (2005) found that 1C – values ranged from 0.6 to 95.0 pg and that the median and mean estimates were 6.6 and 11.7 pg respectively. A surprising finding of the current study is that the genome size of the coconut (1C = 2.98 pg) is lower than the mean for both Angiosperms and the Arecaceae (1C = 3.55 pg; across 56 genera and 90 species), yet the coconut has a slow generation time of 4 – 7 years for Dwarfs and Talls, respectively. We know that the minimum generation time is positively correlated with the C-value of the species in annuals, perennials and obligate perennials (Bennett, 1987) .

In a comparative study on the relationship between cell size and genome size, Beaulieu et al. (2008) found a negative relationship between stomatal density and genome size. Their study demonstrated that trees in comparison to shrubs and herbs had the smallest genome sizes and cell sizes but highest stomatal density. Thus it is possible that genome size fixes the minimum size of guard cells and epidermal cells leading to variation in stomatal density, providing adaptations for certain environments and life history strategies. For example in dry environments, small stomata are more responsive to water stress whilst high density optimizes CO_2 exchanges. Rajagopal *et al*. (1990) recorded stomatal densities for 23 coconut cultivars, with means of 208 mm^{-2} (Talls) and 232 mm^{-2} (Dwarfs), which is about twice what is observed in shade adapted palms such as *Scheelea* (71.9 mm^{-2}) and *Socratea* (120.3 mm^{-2}) studied by Hogan (1988). Given the negative correlation of genomes size and stomatal density reported above and the positive correlation between MGT, we would expect that Tall ecotypes should have larger genome size than Dwarfs but our study found that Dwarf genome

sizes were in the higher quartile of the Tall ecotypes range. A possible explanation for this discordance could be that the genome size had little time to fluctuate since the domestication event.

A novel finding in this study is the evidence for significant intraspecific genome size variation between Tall and Dwarf ecotypes. We also found that genome size variation among Talls was greater than that in Dwarfs. Human-mediated selection for lower MGT in Dwarfs may in the long-term result in lower genome size.

Our research has implications for future of whole genome sequencing and annotation of coconut and understanding of the complexities of the nuclear DNA content and its ploidy levels. Our results indicate that the coconut is diploid and its genome size is 5.966 ± 0.111 pg or 5.757 Gbp which is consistent with the estimate found by Sandoval et al. (2003) based on different cell phases.

Whole genome sequencing involves both nuclear and chloroplast genomes. The chloroplast genome is maternally inherited and consists mostly of coding DNA. Nuclear genomes are inherited bi-parentally and have a higher chance of accumulating mutations, genetic recombination and gene duplication events (Soltis and Soltis, 1999) and if gene duplications were undetected could lead to erroneous phylogenetic inferences and homologies. SNPs discoveries from genome wide association studies (GWAS) may be critically influenced by gene duplications affecting the outcomes of candidate genes for QTLs.

The transcriptome of the Hainan Tall coconut cultivar has recently been sequenced using Next Generation Sequencing techniques (Fan et al., 2013). The whole genome sequence of the coconut will provide us with insights into decoding the traits associated with fruit morphology and selection and importantly to enable the discovery of QTLs associated with disease resistance such as for lethal yellowing.

This study enlightens our understanding of the role of domestication in genome size evolution and revealed that polyploidy is relatively common in the Attaleinae and has evolved multiple times independently. Polyploidy is an important process in the evolution of plants with far reaching effects from molecular to ecological levels and contributes to reproductive isolation, novel gene expressions leading to divergence and potentially to speciation (Adams and Wendel, 2005; Comai, 2005). Detection of ploidy levels using flow cytometric methods provides a practical tool for plant breeders interested in polyploidy because ploidy variation may be exploited for desirable phenotypic traits for horticultural purposes (Parris et al., 2010) or for plant conservation biologists as polyploidy may also be a hindrance to reproduction because of sterility of polyploids.

LITERATURE CITED

ABRAHAM, A., AND P. M. MATHEW. 1963. Cytology of coconut endosperm. *Annals of Botany* 27: 505–513.

ADAMS, K. L., AND J. F. WENDEL. 2005. Polyploidy and genome evolution in plants. *Current Opinions in Plant Biology* 8: 135–141.

AL-DOUS, E. K., G. BINU, M. E. AL-MAHMOUD, M. Y. AL-JABER, H. WANG, Y. M. SALAMEH, E. K. AL-AZWANI, et al. 2011. *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* 29: 521–527.

BATUGAL, P., P. R. BOURDEIX, AND L. BAUDOUIN. 2009. Coconut breeding. *In* S. M. Jain AND P. M. Priyadarshan [eds.], Breeding plantation tree crops: Tropical species. Springer.

BAUDOUIN, L., P. LEBRUN, J. L. KONAN, E. RITTER, A. BERGER, AND N. BILLOTTE. 2006. QTL analysis of fruit components in the progeny of a Rennell Island Tall coconut (*Cocos nucifera* L.) individual. *Theoretical and Applied Genetics* 112: 258–268.

BEAULIEU, J. M., I. J. LEITCH, S. PATEL, A. PENDHARKAR, AND C. A. KNIGHT. 2008. Genome size is a strong predictor of cell size and stomatal density. *New Phytologist* 179: 975–986.

BENNETT, M. D. 1972. Nuclear DNA content and minimum genration time in herbaceous plants. *Proc. R. Soc. Lond.* 181: 109 - 135.

BENNETT, M. D. 1987. Variations in genomic form in plants and its ecological implications. *New Phytologist* 106: 177–200.

------. 1998. Plant genome values: How much do we know? *Proceedings of the National Academy of Sciences* 95: 2011–2016.

BENNETT, M. D., P. BHANDOL, AND I. J. LEITCH. 2000. Nuclear DNA amounts un Angiosperms and their modern uses - 807 new estimates. *Annals of Botany* 86: 859–909.

BOURDEIX, R., L. BAUDOUIN, T. BAMBRIDGE, H. I. JOLY, S. PLANES, AND M. L. GEORGE. 2009. Dynamics and conservation of the coconut palm Cocos nucifera L. in the Pacific region: towards a new conservation approach, 11th Pacific Science inter-congress.

CHAMBERS, J. M., W. S. CLEVELAND, B. KLEINER, AND P. A. TUKEY. 1983. Graphical methods for data analysis, 395 pp. Wadsworth International Group.

CHURCH, G. L. 1940. Cytotaxonomic studies in Gramineae: *Spartina, Andropogon*, and *Panicum*. *American Journal of Botany* 27: 263–271.

COBA DE LA PEÑA, T., AND S. C. BROWN. 2001. Flow cytometry. *In* C. Hawes AND B. Satiat-Jeunemaitre [eds.], Plant Cell Biology: A practical approach. Oxford University Press, Oxford.

COMAI, L. 2005. The advantages and disadvantages of being a polyploid. *Nature Reviews. Genetics* 6: 836–846.

DOEBLEY, J. F., B. S. GAUT, AND B. D. SMITH. 2006. The molecular genetics of crop domesticatiom. *Cell* 127: 1309–1321.

DOLEZEL, J., AND J. BARTOS. 2005. Plant DNA Flow Cytometry and estimation of nuclear genome size. *Annals of Botany* 95: 99–110.

DOLEZEL, J., P. BINAROVA, AND S. LUCRETTI. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum (PRAHA)* 31: 113–120

DOLEZEL, J., J. GREILHUBER, AND J. SUDA. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2: 2233–2244.

DRANSFIELD, J., N. W. UHL, C. B. ASMUSSEN-LANGE, W. J. BAKER, M. M. HARLEY, AND C. E. LEWIS. 2008. Genera Palmarum: The evolution and classification of palms., 397–454. Kew Publishing, Royal Botanic Gardens, Kew.

FAN, H., Y. XIAO, Y. YANG, W. XIA, A. MASON, Z. XIA, F. QIAO, et al. 2013. RNA-Seq analysis of *Cocos nucifera:* Trancsriptome sequencing and *De Novo* assembly for subsequent functional genomics approaches. *PloS One* 8: (3): e59997.

FOALE, M. 2005. An introduction to the coconut palm. *Coconut Genetic Resources*: 1.

GAUT, B. S., AND J. F. DOEBLEY. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences* 94: 6809–6814.

GREILHUBER, J. 2005. Intraspecific variation in genome size in Angiosperms: Identifying its existence. *Annals of Botany* 95: 91—98.

GREILHUBER, J., J. DOLEZEL, M. A. LYSAK, AND M. D. BENNETT. 2005. The origin, evolution and proposed stabilization of the terms "Genome Size" nd "C-value" to describe nuclear DNA contents. *Annals of Botany* 95: 255–260.

GRIME, J. P. 1983. Prediction of weed and crop response to climate based upon measurements of nuclear DNA content, 4. National Vegetable Research Station: The Association of Applied Biologists, Warwick.

GRIME, J. P., AND M. A. MOWFORTH. 1982. Variation in genome size - an ecological interpretation. *Nature* 299: 151—153.

GUINDON, S., AND O. GASCEUL. 2003. PhyML: A simple., fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.

GUNN, B. F. 2004. The phyogeny of the Cocoeae (Arecaceae) with emphasis on Cocos nucifera. *Annals of the Missouri Botanical Garden* 91: 505–522.

GUNN, B. F., L. BAUDOUIN, AND K. M. OLSEN. 2011. Independent origins of cultivated coconut (Cocos nucifera L.) in the Old World Tropics. *PloS One* 6: e21143.

HARRIES, H. 1978. The evolution, dissemination and classification of Cocos nucifera. L. *Botanical Review* 44: 265-319.

HERRAN, A., L. ESTIOKO, D. BECKER, M. J. B. RODRIGUEZ, W. ROHDE, AND E. RITTER. 2000. Linkage mapping and QTL analysis in coconut (Cocos nucifera L.). *Theoretical and Applied Genetics* 101: 292–300.

HESLOP-HARRISON, J. S., AND T. SCHWARZACHER. 2007. Domestication, genomics and the future for banana. *Annals of Botany* 100: 1073–1084.

HOGAN, K. P. 1988. Photosynthesis in two neotropical palm species. *Functional Ecology* 2: 371–377.

KNIGHT, C. A., N. A. MOLINARI, AND D. PETROV. 2005. Large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany* 95: 177–190.

LAURIE, D. A., AND M. D. BENNETT. 1985. Nuclear DNA content in the genera Zea and Sorghum. Intergeneric, interspecific and intraspecific variation. *Heredity* 55: 307—313.

LEITCH, I. J., AND M. D. BENNETT. 1997. Polyploidy in Angiosperms. *Trends in Plant Science* 2: 470—476.

LEITCH, I. J., D. E. SOLTIS, P. S. SOLTIS, AND M. D. BENNETT. 2005. Evolution of DNA amounts across land plants (Embryophyta). *Ann Bot* 95: 207—217.

MADDISON, W. P., AND D. R. MADDISON. 2008. Mesquite, a modular system for evolutionary analysis. Version 2.5. Computer program distributed by authors, http://mesquiteproject.org [accessed 22 June 2014].

MASTERTON, J. 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of Angiosperms. *Science* 264: 421–424.

MEEROW, A., L. NOBLICK, J. W. BORRONE, T. COUVREUR, M. MAURO-HERRERA, W. HAHN, D. N. KUHN, et al. 2009. Phylogenetic analyses of seven WRKY genes across the palm subtribe Attaleinae (Arecaceae) identifies Syagrus as sister group of the coconut. *PloS One* 4: (10):e7353.

MIDGLEY, J. J. 1991. Ecological aspects of the rise of angiosperms: a challenge to the reproductive superiority hypotheses. *Biological Journal of the Linnean Soiety* 44: 81–91.

MILLER, A. J., AND B. L. GROSS. 2011. From forest to field: Perennial fruit crop domestication. *Am J Botany* 98: 1389-1414.

NAMBIAR, M. C., AND M. S. SWAMINATHAN. 1960. Chromosome morphology, microsporogenesis and pollen fertility in some varieties of coconut. *Indian Journal of Genetics and Plant Breeding* 20: 200–211.

PAGEL, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* 48: 612–622.

PARRIS, K., J., T. G. RANNEY, H. T. KNAP, AND V. BAIRD, W. 2010. Ploidy levels, relative genome sizes, and base pair composition in Magnolia. *Journal of American Society of Horticultural Science* 135: 533–547.

PICHERSKY, E. 1990. Nomad DNA - A model for movement and duplication of DNA sequences in plant genomes. *Plant Molecular Biology* 15: 437–448.

PRICE, J. P., AND M. BARANOVA. 1976. Evolution of DNA content in higher plants. *Botanical Review* 42: 27–52.

R DEVELOPMENT CORETEAM. 2011. R: A language and environment for statistical computing. Computer program distributed by R Foundation for Statistical Computing, website: http://www.R-project.org [accessed on June 22, 2014].

RAJAGOPAL, V., B. KASTURI, K. V. KUMAR, AND S. R. VOLETI. 1990. Screening of coconut genotypes for drought tolerance. *Oléagineux* 45: 215–223.

RIESEBERG, L. H. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–1216.

RILEY, R. D., AND K. P. VOGEL. 1982. Chromosome numbers of released cultivars of switchgrass, indiangrass, big bluestem, and sand blue-stem. *Crop Science* 22: 1082–1083.

RIVAL, A., T. BEULE, P. BARRE, S. HAMON, Y. DUVAL, AND M. NOIROT. 1997. Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq) tissue cultures and seed-derived plants. *Plant Cell Reports*: 884–887.

RONCAL, J., F. KAHN, B. MILLAN, T. COUVREUR, AND J.-C. PINTAUD. 2013. Cenozoic colonization and diversification patterns of tropical American palms: evidence from Astrocaryum (Arecaceae). *Botanical Journal of the Linnean Society* 171: 120–139.

RÖSER, M., M. A. T. JOHNSON, AND L. HANSON. 1997. Nuclear DNA amounts in palms (Arecaceae). *Botanica Acta* 110: 79–89.

SANDOVAL, A., V. HOCHER, AND J.-L. VERDEIL. 2003. Flow cytometric analysis of the cell cycle in different coconut palm (*Cocos nucifera* L.) tissues cultured in vitro. *Plant Cell Reports* 22: 25–31.

SAUER, J. D. 1971. A re-evaluation of the coconut as an indicator of human dispersal. *In* C. Riley [ed.], Man across the sea, 309-319. University of Texas Press, Austin.

SHAPCOTT, A., M. RAKOTOARINIVO, R. J. SMITH, G. LYSAOVA, M. F. FAY, AND J. DRANSFIELD. 2007. Can we bring Madagascar's critically endangered palms back from the brink? Genetics, ecology and conservation of the critically endangered palm *Beccariophoenix madagascariensis. Botanical Journal of the Linnean Society* 154: 589–608.

SINGH, R., M. ONG-ABDULLAH, E.-T. LOW, M. A. A. MANAF, AND R. ROSLI. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. *Nature* 500: 335–339.

SOLTIS, D. E., AND P. S. SOLTIS. 1999. Polyploidy: recurrent formation and genome evolution. *Trends in Ecology and Evolution* 14: 348–352.

SPARROW, A. H., AND J. P. MIKSCHE. 1961. Correlations of nuclear volume and DNA content with higer tolerance to chronic radiation. *Science* 134: 282–283.

SPARROW, A. H., S. S. SCHWEMMER, AND P. J. BOTTINO. 1971. The effects of external gamma radiation from radioactive fallout of plants with special reference to crop production. *Radiation Botany* 11: 85–118.

SPARROW, R. C., AND A. H. SPARROW. 1965. Relative radiosensitivities of woody and herbaceous spermatophytes. *Science* 147: 1449–1451.

STEBBINS, G. L. 1959. The role of hybridization in evolution. *Proceedings of the American Philosophical Society* 103: 231–251.

SWAMINATHAN, M. S., AND M. C. NAMBIAR. 1961. Cytology and origin of the Dwarf Coconut Palm. *Nature* 192: 85–86.

SWIFT, H. 1950. The constancy of desoxyribose nucleic acid in plant nuclei. *Proceedings of the National Academy of Sciences* 36: 643–654.

TANKSLEY, S. D. 2004. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *The Plant Cell* 16: S181–S189.

WENDEL, J. F., AND R. C. CRONN. 2003. Polyploidy and the evolutionary history of cotton, Advances in Agronomy, vol. 78, 139–186. Academic Press.

ZHANG, Y., G.-H. XU, X.-Y. GUO, AND L.-J. FAN. 2005. Two ancient rounds of polyploidy in rice genome. *Journal of Zhejiang University SCIENCE* 6B: 87–90.

ZONNEVELD, B. J. M., I. J. LEITCH, AND M. D. BENNETT. 2005. First nuclear DNA amounts in more than 300 Angiosperms. *Annals of Botany* 96: 229–244.

Appendix 1. Absolute genome sizes, 2C-values (pg) estimated for *Cocos nucifera* L. cultivars sampled with *Petunia hybrida* internal standard, from flow cytometry.

Cultivar	*Internat. abbrev.*	*Habit*	*N*	*Abs. genome size/pg (mean±sd)*	*Origin*	*Collection Locality*
Andaman Ordinary Tall	ADOT	Tall	4	6.02 ± 0.09	Andaman Island	Sta. MD_L03A13
Brazil Green Dwarf	BGD	Dwarf	4	5.94 ± 0.03	Brazil	Sta. MD_L13A28
Catigan Green Dwarf	CATD	Dwarf	4	6.04 ± 0.04	Philippines	Sta. MD_L05A15
Cameroon Kribi Tall	CKT	Tall	2	5.87 ± 0.20	Cameroon	Sta. MD_L12A09
Cameroon Red Dwarf	CRD	Tall	3	6.02 ± 0.02	Cameroon	Sta. MD_L06A13
Gazelle Peninsular Tall	GPT	Tall	3	5.89 ± 0.08	Papua New Guinea	Sta. MD_L08A12
Ghana Yellow Dwarf	GYD	Dwarf	3	5.96 ± 0.03	Ghana	Sta. MD_L02A30
Lizard Island Tall	LIZ	Tall	4	5.89 ± 0.05	Australia	ANBG_BG753A
Laccadive Micro Tall	LMT	Tall	3	6.13 ± 0.00	Laccadives	Sta. MD_L08A18
Mission Beach	MISB	Tall	2	5.87 ± 0.00	Australia	RBG SYD_20101370
Malayan Tall	MLT	Tall	4	5.79 ± 0.06	Malaysia	Sta.MD_L03A18
Malayan Yellow Dwarf	MYD	Dwarf	2	5.94 ± 0.02	Malaysia	RBG SYD_903153
Mozambique Tall	MZT	Tall	3	6.19 ± 0.04	Mozambique	Sta. MD_L03A13
Niu Leka Dwarf	NLAD	Compact	4	5.94 ± 0.06	Fiji	Sta. MD_L08A09
Pilipog Green Dwarf	PILD	Dwarf	6	6.01 ± 0.08	Philippines	Sta. MD_L35A28
Panama Tall	PNT	Tall	4	6.01 ± 0.03	Panama	Sta. MD_L03A12
Solomon Island Tall	SIT	Tall	3	5.96 ± 0.03	Solomon Islands	Sta. MD_L21A13
Sri Lanka Tall	SLT	Tall	4	6.07 ± 0.08	Sri Lanka	Sta. MD_L36A24
Tagnanan Tall	TAGT	Tall	3	5.93 ± 0.00	Philippines	Sta. MD_L38A25
Tahiti Tall	TAT	Tall	3	5.75 ± 0.03	Tahiti	Sta. MD_L03A08
Tahiti Red Dwarf	TRD	Dwarf	3	6.04 ± 0.13	Tahiti	Sta. MD_L14A26
Vanuatu Tall	VTT	Tall	3	5.95 ± 0.03	Vanuatu	Sta. MD_L44A24
West Africa Tall	WAT3	Tall	6	5.89 ± 0.06	West Africa	Sta. MD_L09A14

Appendix 2. Absolute genome sizes, 2Cx (pg) estimated for Attaleinae species

Species	*2Cx (pg)*	*x*	*Locality*	*Accession No.*
Allagoptera caudescens (Mart.) Kunze	5.35	4	RBG, Sydney	20091679
Attalea cohune Mart.	4.34	2	RBG, Sydney	20091583
Attalea phalerata Mart. ex Spreng.	4.02	2	RBG, Sydney	20091585
Beccariophoenix madagascariensis Jum. & H.Perrier	3.6	2	RBG, Sydney	20040914
Butia capitata (Mart.) Becc.	3.42	2	RBG, Sydney	932392
Butia eriospatha (Mart. ex Drude) Becc.	3.06	2	RBG, Sydney	780035
Cocos nucifera L. (MYD)	5.94	2	RBG, Sydney	903153
Cocos nucifera L. (NLAD)	5.94	2	Sta.MD	L08A09
Elaeis guineensis Jacq.	3.76	2	Kew C-values website	
Jubaea chilensis (Molina) Baill.	5.10	2	Kew C-values website	20090098
Jubaeopsis caffra Becc.	8.40	5	RBG, Sydney	801080
Lytocaryum weddellianum (H. Wendl.) Toledo	3.72	2	RBG, Sydney	14451
Syagrus botryophora (Mart.) Mart.	4.32	2	RBG, Sydney	20090788
Syagrus coronata (Mart.) Becc.	3.96	2	RBG, Sydney	20091730
Syagrus glaucescens Glaz. ex. Becc.	6.90	2	Kew C-values website	
Syagrus romanzoffiana (Cham.) Glassman	6.10	2	Kew C-values website	
Syagrus sancona (Kunth) H.Karst.	3.90	2	RBG, Sydney	20091729
Syagrus schizophylla (Mart.) Glassman	4.00	2	RBG, Sydney	20091652
Voanioala gerardii J. Dransf.	6.32	19	Kew C-values website	

Notes: Abbrev: Sta. MD = CNRA Marc Delorme Coconut Research Centre in Côte d'Ivoire, Africa; ANBG = Australian National Botanic Gardens Canberra and RBG SYD = Royal Botanic Gardens Sydney, Australia.

APPENDIX 1

Annals of Botany **113**: 1–5, 2014
doi:10.1093/aob/mct244, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

VIEWPOINT

The presence of coconut in southern Panama in pre-Columbian times: clearing up the confusion

Luc Baudouin[1,*], Bee F. Gunn[2] and Kenneth M. Olsen[3]

[1]*CIRAD, UMR-AGAP, 34398 Montpellier, France,* [2]*Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, Australia and* [3]*Biology Department, Washington University, St. Louis, MO, USA*
** For correspondence. E-mail luc.baudouin@cirad.fr*

Received: 12 April 2013 Returned for revision: 26 July 2013 Accepted: 3 September 2013 Published electronically: 13 November 2013

• *Background* The pre-Columbian presence of coconut on the Pacific coast of Panama is attested by a number of independent written accounts. However, recent papers question their accuracy and conclude that coconut was introduced to the region by the Spaniards after their conquests.
• *Scope* In order to examine the value of such claims, an extensive search was conducted of the relevant historical accounts of coconut in America and in the Orient.
• *Key Results* The Spanish chronicler Oviedo (1478–1557) is found to have effectively used fruit and seed size to distinguish coconut from other palms. In addition, it is shown that he has been inaccurately faulted with incorrectly representing a cluster of coconuts. The original drawing, a cluster of a native *Bactris*, was in the marginalia and was only assigned to coconut after Oviedo's death. Finally, the location is identified of a coastal Panamanian site described by Pedro Mártir de Anglería and where tidal dispersal of coconuts was observed.
• *Conclusions* This previously overlooked evidence confirms the pre-historical presence of coconut in Panama. Genetic data indicate that it must have been brought there directly or indirectly from the Philippines. But when, where and by whom remains a subject of research. Further molecular marker studies, computer simulation of natural drift and archaeological research could contribute to this research.

Key words: Coconut, *Cocos nucifera*, New World flora, Panama, oceanic current dissemination, Spanish explorations, Central America, early trans-Pacific voyaging.

Downloaded from http://aob.oxfordjournals.org/ at Australian National University on January 3, 2014

INTRODUCTION

The presence of coconut on the Pacific coast of Panama is attested by a number of historical documents scattered over a 23-year period, from 1516 to 1539, mostly attributable to the chronicles of Pedro Mártir de Anglería and Gonzalo Fernández de Oviedo y Valdés (Oviedo). The complete work of the latter (Amador de los Ríos, 1851) remained unpublished for three centuries. These testimonies were compiled with a number of shorter accounts in Patiño (1964, 2002 pp 241–270), which left no doubt about the presence of coconut in the Americas at the time of European contact (see Zizumbo and Queros, 1998 for an English translation of significant extracts).

Yet, surprisingly, two recent papers (Harries, 2012; Clement *et al.*, 2013) claim that the presence of coconut at the time of contact lacks sufficient evidence and is unlikely. These claims are based on little if any new evidence and rely on a strongly biased selection of texts. In reality, their thesis is based on two extremely strong suppositions: (1) the various witnesses were *systematically* mistaken when they claimed they had seen coconut palms in America; and (2) whenever a document unambiguously describes coconut, it *must* be in reference to the Orient. Neither of these papers actually proves these suppositions; at best they assemble a number of quotations tending to present Oviedo as an incredibly poor observer.

Oviedo's descriptions are not always perfectly accurate by modern standards. The dimensions or volumes he mentioned are rather approximate, partly because he wrote his account in Spain and he may have been betrayed by his memory; however, we did not find any instance where he was obviously misidentifying coconut. For instance, both Clement *et al.* (2013) and Harries (2012) quote the following sentence in Oviedo's account: 'After I wrote the report I have mentioned, I was in the province and headland of Borica, and I ate some of these cocos and carried many with me to Nicaragua, and came to loathe them, and others did as I did and said the same thing as well'. The hypothesis of a misidentification (of some *Bactris* species) was cautiously suggested by Allen (1965) and Clement *et al.* and Harries seem to hold it as established truth. They claim that Oviedo's cocos had little water because he says he *ate* rather than *drank* them. They apparently did not notice that Oviedo indicates the usual way of consuming coconut: coconut milk was incorporated into mazamorra (a porridge-like meal made with bread or corn). They add that 'There are people who find coconut kernel indigestible, but it is not usually a group phenomenon'. However, Oviedo makes it very clear that what his group was complaining about was *massive and continuous consumption* of coconut, not coconut itself. He concludes 'Finally, it is food for men who work and who are very strong, but for the rest a little of this fruit is enough, or if eaten continually, as it was done there, it is not for all stomachs'. Considering the recipe of coconut-based mazamorra, we believe that few nutritionists would disagree.

Starting with Patiño's work, we conducted an extensive search throughout relevant historical accounts of coconut in America and in the Orient. We have found evidence that specifically

© The Author 2013. Published by Oxford University Press on behalf of the Annals of Botany Company. All rights reserved.
For Permissions, please email: journals.permissions@oup.com

refutes the above suppositions: (1) a review of Oviedo's writings, including early editions of his manuscripts, demonstrates that he clearly distinguished between coconut and other smaller-fruited local palms; and (2) we have identified the location of a Panamanian coastal site that was described by Pedro Mártir de Anglería as containing coconut palms, with tidal dispersal of the fruits (a key indicator that the fruits in question were indeed coconut).

THE EVIDENCE

Fruit size as an effective classification criterion

Like everyone who had heard about coconut in his time, Oviedo knew that a coconut was the size of a human head and that it grew on a tall tree that looked like a date palm. Confusion with any other local palm thus seems highly unlikely because of the huge differences in fruit size. This is confirmed by what Oviedo says about the dozen palm species he describes in Book 9, Chapter 4 (Amador de los Ríos, 1851, pp 332–337 of Tome 1). He extends the name 'coco' to various palm seeds, which, like the coconut, exhibit three apertures (e.g. *Elaeis oleifera* or *Bactris*), but he always makes clear that these 'cocos' are small (like a walnut or an olive) and thus different from the 'big coco' he saw in the province of Cacique Chimán (Oviedo y Valdés, 1526), which is bigger than a human head. Oviedo probably had personal experience with the vessels made out of coconut shells that he mentions because such goblets were relatively common in European courts of his time (Tripps, 2005). Most palm fruits of Central America are much too small for this kind of use. *Attalea cuatrecasana* has large fruits (14 cm. long) but, unlike coconut, it grows inland in the rainforests of Colombia and has only a short subterranean stem. Its fruits do not contain any liquid and seeds with two or more kernels are not infrequent.

A drawing erroneously assigned to coconut

One of the most serious reasons for doubting Oviedo's botanical ability was a drawing represented as Figure 15 of Plate 3 in Amador de los Ríos (1851). It is referred to in the coconut section, but is not convincing because it mixes traits of coconut and of *Bactris*. Actually, the original drawing does not represent coconuts at all. It is found in folio 53v of manuscript HM177 conserved at the Huntington Library (Myers, 2007) and represents *Bactris* fruits—recognized by their fused, shallowly lobed calyces—borne at the end of spiny branches (Fig. 1). Contrary to most of Oviedo's illustrations, it is not located within the text but at the bottom of the right margin, partly embedded in a long marginal addition devoted to 'pixabay' and 'cañaspalmas', two species of the genus *Bactris*. Thus, it represents one of these species. The error is due to Amador de los Ríos (he was not a botanist and the drawing was in front of the coconut section) and to his engraver, who apparently felt he should make the fruits look more coconut-like and modified the calyces accordingly (Fig. 1B, C).

Coconut growing spontaneously in Aguadulce (Panama)

Our next line of evidence comes from Pedro Mártir de Anglería's *De orbe novo*. His accounts of coconut have been discounted because he never left Europe and thus they were

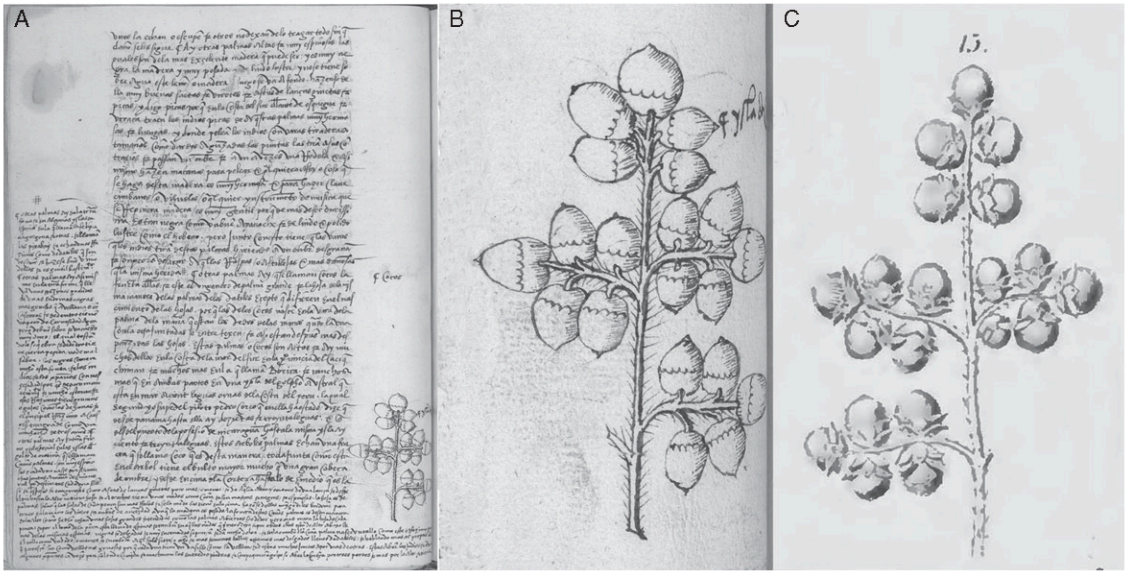


Fig. 1. (A) Reproduction of folio 53v of Oviedo's original manuscript. The rectangular layout of the initial version (written during 1540–1542) is clearly visible. The illustration is placed in continuity with a marginal addition made in the period 1546–1549. (B) Enlarged version of the original drawing. (C) The 1851 interpretation, redrawn and inverted as a result of the lithography technique. Note the difference in the calyces. The original version represents *Bactris* fruits while the modern version was 'improved' to make it look more like the coconut fruits it was supposed to represent. Sources: (A, B) ms. HM 177 (Vol. 2), Huntington Library; (C) Amador de los Ríos (1851).

second-hand testimony. However, his lack of expertise is precisely what would have made him unable to make a convincing description of the natural dissemination of coconuts by the oceanic currents if he was not repeating faithfully what was told to him. Close to Natá (a historic city in the Coclé province of Panama), he says, 'a great abundance of the cocos I mentioned earlier exist there, mainly in the austral region, where the tide penetrates widely in the neighbouring plains. *In one of them, they say that there is a two league space which is washed by the high tide and left dry at low tide.* Such places are those where they say that these trees sprout and grow spontaneously. In the other places, there are none unless they are transplanted when still young. Some think that the high tide leaves there the seeds of these trees from unknown regions' (Torres Asensio, 1892).

Here we undoubtedly have the coconut palm flourishing in its natural environment, precisely close to the town of Aguadulce (Panama). The habit of sprouting where the high tide leaves them is a unique trait of coconuts and the sentence in italics refers to the mouth of the Santa María River, near Aguadulce, which was converted into a salt works centuries ago. The uncommon geographical feature described here matches perfectly with the place represented in Fig. 2 in terms of topography, size and location. Moreover, Mártir de Anglería's anonymous informer would never have discovered a connection between the distribution of the coconut palm and the variations of the slope of the beach if he had not observed it on the spot. Likewise, he would never have added that 'in the other places, there are none unless they are transplanted when still young' if he was referring to India, because, according to Varthema, coconut in India was exclusively cultivated (Teyssier, 2004). There is thus no way in which these three elements—coconut, natural dissemination and Aguadulce—can be dissociated and the above text shows that coconut grew spontaneously in America.

DISCUSSION

Systematically tracking the sources of the documents has proved effective in confirming the pre-Columbian presence of coconut in America, which had been firmly established by Patiño. Coconut has a few unique features that even the poorest observer would not miss, and the texts tell us that Oviedo noticed its uncommonly large fruit size and used it as a classification criterion. In addition, it is now clear that Oviedo was not responsible for

FIG. 2. The main salt works in Aguadulce, located 8°9′ N, 8°31′ W, 17 km south of Natá. It is protected from the sea by a 5–10 m high dam (CD). Points A and B are located 7·5 km (~2 leagues) from the sea and are only 1 m above sea level. Before the construction of the dam, the space between them and the sea was inundated by the high tide and left dry at low tide, as stated in the text. This phenomenon, not unlike what is observed in the bay of Mount St Michel (France), is rare enough to warrant that this perfect matching is not merely coincidental. In total, the salt works stretch for 32 km around Aguadulce. Source: Google Earth.

incorrectly assigning Figure 15 of Plate 3 in Amador de los Ríos (1851) to coconut. Finally, we identified the site described by Mártir de Anglería as Aguadulce, Panama, a place where, five centuries later, J. L. Renard would collect one of the representative samples of the Panama Tall. It is significant that the populations from the Pacific coast of Panama, including this one, along with others from Costa Rica and from the north of Peru, can all be traced back to the same origin, a very small number of palms (effective population size was estimated to have been between 2 and 5) originating from the Philippines.

Did coconut originate in America?

The hypothesis of an American origin of coconut was defended by Cook (1910). It is indisputable that its closest ancestors were American (Gunn, 2004; Meerow *et al.*, 2009) and a putative *Cocos* fossil dating back 60 m.y.a. (Gomez-Navarro, 2009) was found in Colombia. However, a permanent presence of coconut in America during the Holocene is extremely unlikely given the absence of linguistic, archaeological and ethnobotanical evidence (Patiño, 2002; Clement *et al.*, 2013). In addition, genetic studies do not reveal an American centre of diversification (Gunn *et al.*, 2011). On the contrary, they demonstrate that, while all of the alleles of the Panama Tall exist in the Philippines, the reciprocal is not true, which indicates a close relationship between the coconuts from both regions and the direction of the migration (Baudouin and Lebrun, 2009). Finally, diagnostic features of *Cocos* could not be observed in the Colombian fossil due to incomplete preservation and its assignment to the genus *Cocos* genus is uncertain. Systematicians tend to place it at the root of the subtribe Attaleinae (Eisenhardt *et al.*, 2011; Meerow *et al.*, 2009). *Cocos nucifera* has American ancestors but its lineage probably became extinct on the continent until it was introduced during the late Holocene, but before Columbus.

Coconut grew spontaneously in America

The historical documents make it clear that coconut in America was not cultivated, with the possible exception—mentioned by Mártir de Anglería—of the Pearl Islands Archipelago, to the East of the Gulf of Panama. This is confirmed by Clement *et al.* (2013) and may be surprising because the size of the populations was such that the natives must have co-existed with coconut for at least four generations (Patiño, 2002), without developing a tradition of growing or even using coconut. But Patiño (2002) cites a similar case two centuries later in the bay of Bocas de Toro (Atlantic coast of Panama). Anglería's description of natural dissemination gives us clues about the pre-historic distribution of coconut in Central America. It was abundant in a limited number of places, where the topography was favourable and absent elsewhere. Another factor inevitably played a major role: the direction of oceanic currents. Computer simulation studies of the same kind as those made by Ward and Brookfield (1972), but at a regional scale, could help our understanding of this distribution and (possibly by reversing time) give indications about the place where coconut first reached America.

From the Philippines to America

Yet the Panama Tall is no doubt descended from cultivated populations. It must have been brought to the Americas, because the distance from the Philippines to Panama prevents unaided drifting. At the same time, it is clear that the tradition of coconut cultivation was not passed to the natives of Central America, maybe because those who brought it had little contact with them, because they did not stay long enough, or because they reached America in another region, possibly more to the south. It could be the Bay of Caráquez, as proposed by Baudouin and Lebrun (2009) or the Gulf of Guayaquil, one of the three regions highlighted by Jones *et al.* (2011) for pre-Columbian contact. The journey from the Philippines to America was not necessarily direct. An intermediate stage in the Polynesian triangle is unlikely because the genetic structure of the populations is different (Gunn *et al.*, 2011) but a more northern route, via the Polynesian outliers (whose coconut populations are yet to be characterized molecularly) can be envisaged. Further research in this area is needed.

CONCLUSIONS

We show in this paper that at least part of the accounts of coconut in America resulted from genuinely local observation and that the hypothesis of systematic confusion between coconut and some undetermined palm species is contradicted by the evidence. The pre-historic presence of coconut is thus demonstrated beyond reasonable doubt. How precisely and when it was brought to the Americas and came to form spontaneous populations in Panama remains an open field of inquiry, although hypotheses can be proposed (Baudouin and Lebrun, 2009).

A growing amount of evidence attests to the existence of ancient trans-Pacific travels from Polynesia to America (Jones *et al.*, 2011) and in the reverse direction (Roullier *et al.*, 2013). A more detailed understanding of the conditions of these travels, the dates and the people who undertook them, as well as of the consequences in the regions of arrival, will require combining results of the application of various disciplines to different animal and plant species (in addition to artefacts and human features). Coconut fully deserves its place in the set of commensal models proposed in Storey *et al.* (2013).

ACKNOWLEDGEMENTS

We are grateful to the Huntington Library, San Marino, CA, USA, for kindly providing the illustration of Fig. 1 and to Geo Coppens d'Eckenbrugge for manuscript revision and useful suggestions.

LITERATURE CITED

Allen PH. 1965. Oviedo on "Cocos". *Principes* **9**: 60–66.

Amador de los Ríos J. ed 1851–55. Historia general y natural de las Indias, islas y tierra-firme del mar océano – por el capitán Gonzalo Fernández de Oviedo y Valdés, primer cronista del Nuevo mundo. In five volumes. Real Academia de la Historia.

Baudouin L, Lebrun P. 2009. Coconut (*Cocos nucifera* L.) DNA studies support the hypothesis of an ancient Austronesian migration from Southeast Asia to America. *Genetic Resources and Crop Evolution* **56**: 257–262.

Clement CR, Zizumbo-Villarreal D, Brown CH, Ward RG, Alves-Pereira A, Harries HC. 2013. Coconuts in the Americas. *Botanical Review* **79**: 342–370.

Cook OF. 1910. History of the coconut palm in America. *Contributions from the US National Herbarium* **14**: 271–342.

Eiserhardt WL, Pintaud JC, Asmussen-Lange C, Hahn W, Bernal R, Balslev H, Borchsenius F. 2011. Phylogeny and divergence times of Bactridinae (Arecaceae, Palmae) based on plastid and nuclear DNA sequences. *Genetic Taxon* **60**: 485–498.

Gomez-Navarro C, Jaramillo C, Herrera F, Wing SL, Callejas R. 2009. Palms (Arecaceae) from a Paleocene rainforest of northern Colombia. *American Journal of Botany* **96**: 1300–1312.

Gunn BF. 2004. The phylogeny of the Cocoeae (Arecaceae) with emphasis on *Cocos nucifera*. *Annals of the Missouri Botanical Garden* **91**: 505–522.

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

Harries HC. 2012. Key to coconut cultivation on the American Pacific coast: the Manila-Acapulco galleon route (1565–1815). *Palms* **56**: 72–77.

Jones TL, Storey AA, Matisoo-Smith EA, Ramirez-Aliaga JM. 2011. Polynesians in America: pre-Columbian contacts with the New World. Lanham, MD: AltaMira Press.

Meerow A, Noblick L, Borrone JW, *et al*. 2009. Phylogenetic analysis of seven WRKY genes across the palm subtribe Attaleinae (Arecaceae) identifies *Syagrus* as sister group of the coconut. *PloS ONE* **4**:e7353. doi:10.1371/journal.pone.0007353.

Myers KA. 2007. Fernández de Oviedo's chronicle of America: a new history for a New World. Austin: University of Texas Press.

de Oviedo y Valdés GF. 1526. *De la natural historia de las Indias.* Toledo: Maestre Remon de Petras.

Patiño VM. 1964. *Plantas cultivadas y animales domésticos en América equinoccial. Tomo I: frutales.* Cali: Imprenta Departamental.

Patiño VM. 2002. *Historia y dispersión de los frutales nativos del Neotrópico. CIAT Publication no. 368.* Palmira: CIAT, 241–270.

Roullier C, Benoit L, McKey DB, Velot V. 2013. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. *Proceedings of the National Academy of Sciences USA* **110**: 2205–2210.

Storey AA, Clarke AC, Ladefoged T, Robins J, Matisoo-Smith E. 2013. DNA and Pacific commensal models: applications, construction, limitations, and future prospects. *Journal of Island and Coastal Archaeology* **8**: 37–65.

Teyssier P. 2004. *Voyage de Ludovico di Varthema en Arabie et aux Indes Orientales.* Paris: Editions Chandeigne.

Torres Asensio J. 1892. *Fuentes históricas sobre Colón y América – Pedro Mártir de Anglería.* In four volumes. Imp. de la SE de San Francisco de Sales.

Tripps J. 2005. Pilgerfahrten als kreative Impulse für die Goldschmiedekunst der Spätgotik: die Stiftungen von Pilger- und Reiseandenken durch Adel und Stadtpatriziat in Kirchenschätze. In Babel R, Paravicini W Ed. *Grand Tour. Adeliges Reisen und Europäische Kultur vom 14. bis zum 18. Jahrhundert (Beihefte der Francia no. 60).* Paris: Deutsches Historisches Institut.

Ward RG, Brookfield M. 1972. The dispersal of the coconut: did it float or was it carried to Panama? *Journal of Biogeography* **19**: 467–480.

Zizumbo-Villareal D, Quero HJ. 1998. Re-evaluation of early observations on coconut in the New World. *Economic Botany* **52**: 68–77.

APPENDIX 2

RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

Ploidy and domestication are associated with genome size variation in Palms[1*]

Bee F. Gunn[2,7], Luc Baudouin[3], Thierry Beulé[4], Pascal Ilbert[4], Christophe Duperray[5], Michael Crisp[2], Auguste Issali[6], Jean-Louis Konan[6], and Alain Rival[4]

PREMISE OF THE STUDY: The genome size of a species (C-value) is associated with growth, development and adaptation to environmental changes. Angiosperm C-values range 1200-fold and frequently vary within species, although little is known about the impacts of domestication on genome size. Genome size variation among related species of palms is of evolutionary significance because changes characterize clades and may be associated with polyploidy, transposon amplifications, deletions, or rearrangements. Further knowledge of genome size will provide crucial information needed for planning of whole genome sequencing and accurate annotations. We studied the genome size of *Cocos nucifera* and its variation among cultivars, and compared it to values for related palms from the Attaleinae subtribe.

METHODS: Flow cytometric analysis of isolated nuclei from young palm leaves was used to estimate genome sizes of 23 coconut cultivars (Talls, Dwarfs, and hybrids) worldwide and 17 Cocoseae species. Ancestral genome size was reconstructed on a maximum likelihood phylogeny of Attaleinae from seven *WRKY* loci.

KEY RESULTS: The coconut genome is large—averaging 5.966 pg—and shows intraspecific variation associated with domestication. Variation among Tall coconuts was significantly greater than among Dwarfs. Attaleinae genomes showed moderate size variation across genera, except polyploids *Jubaeopsis caffra*, *Voanioala gerardii*, *Beccariophoenix alfredii*, and *Allagoptera caudescens*, which had larger genomes.

CONCLUSIONS: Our results contribute to the understanding of the relationship between domestication and genome size in long-lived tree crops and provide a basis for whole-genome sequencing of the coconut and other domesticated plants. Polyploidy evolved independently in two clades within Attaleinae.

KEY WORDS Attaleinae; C-value; *Cocos nucifera*; domestication; flow cytometry; evolution; holoploidy; minimum generation time; nuclear DNA content; polyploidy

Polyploidy is an important product of plant evolution with far-reaching effects from molecular to ecological levels. It also contributes to reproductive isolation, as novel gene expressions lead to divergence and potentially to speciation (Adams and Wendel, 2005; Comai, 2005). Polyploidy is known to occur among 80% of angiosperms (Masterson, 1994) and it is also common in domesticated plants. Indeed it is detectable in major crops such as cereals (wheat and rye), maize, cotton, potato, banana, sugar cane, and coffee (Gaut and Doebley, 1997; Wendel and Cronn, 2003; Heslop-Harrison and Schwarzacher, 2007). More, polyploidy adds complexity when identifying the wild ancestors of a domesticated plant (Olsen and Wendel, 2013). Understanding the impacts of ploidy levels on the genome size provides an assessment of gene duplications and transposable elements which may play an important role in epigenetic gene silencing or expression and which also provide protection against harmful viruses and transposons (Pichersky, 1990).

Detection of ploidy levels using flow cytometric methods is a practical tool for plant breeders because polyploidy may be exploited for desirable phenotypic traits. Indeed, studies on hybrids of *Magnolia* [Magnoliaceae] have shown that greater differences in ploidy levels between the parents lead to greater sterility in the progenies (Parris et al., 2010). For plant conservation biologists, polyploidy may also be a hindrance to reproduction because of the sterility of polyploids.

[1] Manuscript received 12 April 2015; revision accepted 24 August 2015.

[2] Research School of Biology, Division of Evolution, Ecology and Genetics, The Australian National University, Acton, ACT 2601, Australia;
[3] CIRAD, UMR AGAP, F-34398 Montpellier, France;
[4] CIRAD, UMR DIADE, F-34398 Montpellier, France;
[5] Montpellier Rio Imaging, IRB, INSERM U1040, CHU de Montpellier 34295 Montpellier, France; and
[6] Station de Recherche Marc Delorme, Port Bouët. 07 BP 13 Abidjan, Côte d'Ivoire
[7] Author for correspondence (email: bee.gunn@anu.edu.au)

* The present research work is dedicated to the memory of our respected colleague Jean-Christophe Pintaud, who suddenly passed away in August 2015.

doi:10.3732/ajb.1500164

The C-value is equivalent to genome size in diploid species but it is always greater than the genome size(s) in polyploids (Bennett et al., 2000). A diploid plant has two genomes after gametic fertilization whereas a polyploid has more than two genomes, as a result of either autopolyploidization or allopolyploidization following hybridization (Stebbins, 1959). The C-value (holoploid genome size) of a species corresponds to the DNA amount in its unreplicated haploid or gametic nucleus (pollen or sperm), regardless of its ploidy level (Swift, 1950; Greilhuber et al., 2005) and it is measured in picograms (pg) or base pairs (bp).

The genome size of a species has major effects on the growth, meiotic, and mitotic cycles and on the expansion of cells. Cellular DNA content or nucleotypic changes therefore affect the individual's morphological and physiological development as well as adaptations to its environment (Price and Baranova, 1976; Bennett, 1998; Hardie and Hebert, 2003; Knight et al., 2005). Large variation in C-values may have consequences or costs to the organisms. Indeed, several studies have shown that C-values in plants are often associated with ecological constraints (Bennett, 1987; Knight et al., 2005), temporal shifts in phenology such as the early flowering of *Fritillaria* sp. [Liliaceae] (2C = 96.5 – 254.8) (Grime and Mowforth, 1982), or sensitivity to ionizing radiations and climatic changes in plants and possibly also in animals (Sparrow and Miksche, 1961; Sparrow and Sparrow, 1965; Sparrow et al., 1971).

Chromosome numbers (2n), C-values, and ploidy levels are tightly linked and remain constant for most species; nevertheless, there are exceptions in which variation does occur. Intraspecific variation in C-values is not rare in plant species despite the absence of any change in chromosome number; for example, the domesticated crop *Zea mays* L. [Poaceae] (2n = 20) shows 37% variation among various cultivar lines (Laurie and Bennett, 1985) and *Poa annua* L. [Poaceae] (2n = 28) shows a 100% variation rate (Grime, 1983). The switchgrass, *Panicum virgatum* L. [Poaceae] is a North American native perennial cultivated for pastures, rangelands, and fuel biomass. Cytological studies reveal that the latter species presents a series of karyotypes ranging from diploid (2n = 18) to dodecaploid (2n = 12C = 108) (Church, 1940; Riley and Vogel, 1982).

The palm family (Arecaceae) is among the most diverse in the plant kingdom, with C-values ranging from 0.9 to 30 pg (Angiosperm 1C-values database (http://data.kew.org/cvalues/)). *Cocos nucifera* L. (Arecaceae) has 16 chromosomes (Nambiar and Swaminathan, 1960; Abraham and Mathew, 1963) and is the only species of its genus. The coconut palm is cultivated globally on over 12 million hectares in the humid tropics. It is best regarded as a semidomesticated species, a complex of local populations with all degrees of dependency upon humans, from nil to complete (Sauer, 1971). Harries (1978) distinguishes "domesticated" and "wild" coconuts, but this distinction refers to an ancient domestication event and both types are cultivated nowadays. Wild populations do exist in a few locations (Foale, 2005), although some of them might be feral, i.e., formerly cultivated populations surviving spontaneously (Baudouin et al., 2014).

At the other end of the domestication continuum, the Dwarf coconut type can be regarded as the most completely domesticated type (Gunn et al., 2011). The Dwarf type is usually grown near human habitations and accounts for only 5% of coconuts globally (Bourdeix et al., 2010). Its self-pollinating floral biology enables the true-to-type propagation of desirable genotypes and the screening for rare off-types based on recognizable phenotypic markers such as fruit color and shape. It is precocious, maturing usually after four years. Dwarf coconut is especially appreciated for the liquid in its immature nuts and its slow growth makes harvesting relatively easy for most of its relatively short lifespan (ca. 35 yr) (Bourdeix et al., 2010). Finally, the Dwarf type is dependent on human protection because it is a poor competitor in natural stands or in mixed plantings due to its short lifespan and limited vigor.

The Tall coconut type—which is more frequently cultivated—lacks most of the "domesticated" features found in its Dwarf counterpart. It is predominantly cross-pollinated and thus highly heterozygous. Tall coconuts are fast-growing, i. e., they become reproductively mature after seven years and they live for 70 years or more (Bourdeix et al., 2010). Besides Talls and Dwarfs, some relatively rare types are observed, among them Semi-Talls, which are self-pollinating like Dwarfs but relatively more robust. The "compact Dwarf" represented by the *Niu Leka* Dwarf from the South Pacific is not closely related to other Dwarfs. It is cross-pollinating, as vigorous as a Tall and its small size is due to a marked reduction in internode length and in the distance between leaflets (Lebrun et al., 2005) .

Determination of the genome sizes of cultivated coconuts and ploidy level are essential prerequisites for the sequencing of the coconut genome. This will provide a precise calculation for the optimal depth of reads required for the accurate assembly and annotations of the coconut genome. Genome sequences have been recently generated and made publicly available for two palm species of major economic importance, namely the date palm (Al-Dous et al., 2011) and the oil palm (Singh et al., 2013). For the coconut palm, future genome sequencing will be important in identifying genes responsible for disease resistance and characters of agro-ecological interest such as drought or salt tolerance (Fan et al., 2013). The integration of gene discovery and marker-assisted breeding will pave the way for the generation of new coconut cultivars, which will be better adapted to changing agro-climatic conditions and agricultural practices.

We are keen to know whether the phenotypic differences such as dwarfism and fruit morphology observed between Dwarf and Tall cultivars, and their different generation times (three vs. seven years), are related to their genome size. In this study, we explored genome-size variation through the flow cytometric analysis of 23 coconut genotypes from around the globe, including two Australian wild-sown coconuts. Our objectives were: (1) to determine the actual genome size of coconut, for which contradictory values were published; (2) to identify and study intraspecific variation, and the impact of domestication on genome size; (3) to test whether genome size is less variable in Dwarf than in Tall coconut types; and (4) to reconstruct ancestral genome sizes across the subtribe Attaleinae.

MATERIALS AND METHODS

Plant material—We sampled immature leaves from 23 adult palms originating from 23 coconut populations, which were selected with the aim of covering most of the genetic diversity of the genus (Appendix 1). Two of them were self-sown, putatively wild, populations from Australia (Mission Beach, lat. −17.869121°, long. 146.106338° and Lizard Island, lat. −14.667717°, long. 145.446729°). The other coconut types under study were traditional and advanced cultivars from the germplasm collection preserved at Marc Delorme Research Station (CNRA, Côte d'Ivoire). They include seven self-pollinating Dwarf cultivars, 15 cross-pollinating Tall cultivars, one cross-pollinating "compact Dwarf" cultivar and three population hybrids (one Tall × Tall and two Dwarf × Tall).

Fresh leaf material was collected from the unopened spear leaf of the palm whenever possible. In addition, we sampled leaf material for 17 other species across 9 genera of the tribe Cocoseae: *Allagoptera*, *Astrocaryum*, *Attalea*, *Bactris*, *Beccariophoenix*, *Butia*, *Jubaeopsis*, *Lytocaryum*, and *Syagrus* from the living collections of the Royal Botanic Gardens in Sydney, Australia. We obtained genome size values for four additional species from the Angiosperm 1C-values database. We wrapped approx. 4 cm length of each leaf in moistened tissue paper and placed it into an envelope kept at 4°C to preserve it during transportation to the IRB laboratory in Montpellier, France.

Estimation of 2C-value—To extract nuclei we chopped coconut and *Petunia ×hybrida* hort. ex E. Vilm. leaves using razor blades. The *P. ×hybrida* Px PC6 (Vilmorin), 2C = 2.85 pg was grown in the greenhouse and leaves were used as a calibration standard following Coba de la Peña and Brown (2001). Approximately 1 cm^2 of fresh leaves were chopped in 500 µL of Doležel's lysis buffer (Doležel et al., 1989) with the following modifications: (1) no spermine was added; and (2) we replaced β-mercaptoethanol with 10 mM sodium metabisulphite which was added immediately before use (Rival et al., 1997). The lysate was then filtered through disposable filters using 20 µm nylon mesh (Partec CellTrics, Görlitz, Germany) to isolate nuclei from cell debris and aggregates. Then 500 µL of the filtrate were pipetted into a new disposable tube and 20 µL of DAPI (4',6-diamidino-2-phenylindole, dihydrochloride) fluorochrome solution (0.1 mg mL^{-1}) were added, for a final DAPI concentration of 4 µg mL^{-1}. After homogenizing and stabilizing for 5 min at room temperature, the stained nuclei suspensions were analyzed.

We measured relative fluorescence intensities from stained nuclei using a Beckman-Coulter CyAN™ ADP flow cytometer (Beckman Coulter, Brea, California, USA) with at least 500 nuclei analyzed per run. We repeated measurements of the G1 peaks (nonreplicated phase of the cell cycle) for each coconut cultivar 3-5 times with internal standards and used the means (µ ±SD) in our assessment of the absolute value of the coconut's genome size, yielding graphical outputs such as illustrated in Fig. 1.

Data analysis—The first step of data analysis consisted of a visual examination of the cytometer plots (Fig. 1) to exclude unreliable runs (i.e., with low signal to noise ratio and, as a consequence, fluorescence peaks that could not be adequately discerned which were scattered among genotypes, mainly due to inadequate preservation of analyzed plant material). Twenty-four samples out of 80 were discarded as unreliable, leaving 56 available for further analysis (see Table 1).

Proportionality of G1 peak values with internal standard—The proportionality of the G1 peak values between the coconut genotypes and the internal standard (*Petunia ×hybrida*) was checked through regression analyses. As a result, G1 peak values for various coconuts against the internal standard (*Petunia ×hybrida*) were found to be highly correlated (corrected $R^2 = 0.9997$ when the intercept was fixed to 0), thus confirming their proportionality. The proportionality coefficient was 2.0921 ± 0.0041 (mean ± SE). This enabled the use of the ratio of the coconut G1 values to the internal standard to calculate the absolute genome size of the coconut ecotypes (see Appendix 1).

Genome size for each sample was estimated by $G_C = D_C/D_S * G_S$ where D_C is the G1 peak value of coconut, D_S is the G1 peak value of the standard, and G_S is the genome size of the standard (2.85 pg for *Petunia*). We examined variation in genome size among coconut cultivars using ANOVA and we applied an F-test to determine the significance of the values. We tested for possible effects of domestication on genome size of *Cocos nucifera* using ANOVA with two groups: Tall (n = 16), and Dwarf (n = 7). We followed the same method to analyze variation between Indo-Atlantic and Pacific groups of geographical origin. Finally, we visualized changes in DNA amounts in Dwarf and Tall coconuts using boxplots. Calculations and graphical representation were carried out using R software version 3.1.1 (Chambers et al., 1983; R Development Core Team, 2011).

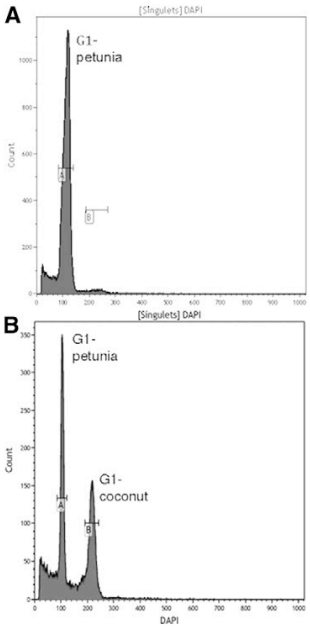


FIGURE 1 Examples of flow cytometry histograms. A: Peak A: *Petunia* standard alone; B is a marker showing relative fluorescence B: Peak A: *Petunia* standard, Peak B: *Cocos nucifera* G1 represents the nonreplicating cell phase.

Ploidy level—Ploidy in flow-cytometric assays equates a constant DNA quantity (C-value) of the complete chromosome complement with respect to a published reference standard of known ploidy. We determined the ploidy level of the coconut and 17 other species of Cocoseae from the positions of the G1 peaks in histograms of relative fluorescence intensities. The presence of polyploidy is reflected in the position of the dominant G1 peak and the appearance of more than one nonreference dominant peak in a single sample apart from the internal standard.

***Evolution of 2C-value in* Attaleinae**—We estimated the absolute genome size of the 17 species using flow cytometry (Appendix 2). To estimate the evolutionary tree of the *Attaleinae*, we used seven *WRKY*

TABLE 1. ANOVA of estimated DNA content in coconut cultivars (pg).

Df		Sum Sq	Mean Sq	F	*P* value
Between types	1	0.04511	0.04511	10.90	0.0017
Within Dwarf type	6	0.03318	0.00553	1.34	0.2568
Within Tall type	15	0.64875	0.04325	10.45	0.0000
Residuals	56	0.23183	0.00414		

nuclear loci from Meerow et al. (2009), concatenated to sequence length of 5.648 kb for 56 taxa across the Attaleinae available from GenBank. We conducted maximum likelihood analyses using PHYML software (Guindon and Gasceul, 2003) implemented through Geneious 6.1.7 with the following criteria: initial BioNJ tree, NNI topology search, GTR substitution model, discrete Gamma model, 4 categories, random seed, and 100 bootstrap replicates.

We applied the maximum likelihood approach as described in Pagel (1999) for ancestral character reconstruction as implemented in the Mesquite software. The maximum likelihood trees (100) were imported into Mesquite version 2.5 (Maddison and Maddison, 2008) and a character matrix of 2Cx-values for 19 taxa were appended to the DNA sequences. We used the 2Cx-values of the taxa to allow for comparability by correcting for ploidy and for optimization on the phylogeny. We traced the 2Cx-values sizes as continuous characters on the ML tree to infer ancestral state likelihoods. *Bactris* and *Elaeis* were used as outgroups for the nonspiny Attaleinae.

RESULTS

Absolute genome size of the coconut—The overall mean of genome size of *Cocos nucifera* was found to be 5.966 pg, after exclusion of the hybrid genotypes. The residual standard deviation was 0.0641 pg. This represents the uncertainty due to the breadth of the peaks and to random fluctuations in the experimental conditions and variation among all cultivars.

Ploidy level in* Cocos nucifera *cultivars—The fluorescence histograms obtained for all coconut cultivars under study clearly showed a single G1 peak, suggesting that all sampled cultivars are diploids (Fig. 1). G1 peaks occurred in the same position relative to the internal standard in all cases. Since the *Petunia ×hybrida* standard used has nearly half the DNA quantity of the coconuts, it is possible that if haploid cells were present in the coconut samples, their peaks might have overlapped with the standard, but leaf cells are somatic and do not undergo meiosis. Nevertheless, the possible presence of spontaneous haploids was checked in several samples without internal standards and it proved constantly negative.

Variation of genome size in coconut—We performed an analysis of variance (ANOVA) based on 16 Tall and 7 Dwarf coconut types (Table 1). On average, Tall and Dwarf coconuts differed in genome size. There were also significant differences among Talls but the studied Dwarfs were not significantly different (Table 1). The estimated mean and confidence interval ($\alpha = 0.05$) of genome size were 6.00 [5.97 – 6.03] and 5.95 [5.74 – 6.16] in Dwarfs and Talls, respectively. This takes into account both experimental errors and the estimated variance of genome size across cultivars. Although the genome size in Dwarf is slightly larger than the *average* genome size of Talls, it remains within the range of Tall coconuts (Fig. 2). This is also the case for the three additional individuals we sampled in hybrid populations (one Tall × Tall, 2C = 6.13 pg, and two Dwarf × Tall, 2C = 5.90 pg and 5.92 pg, respectively).

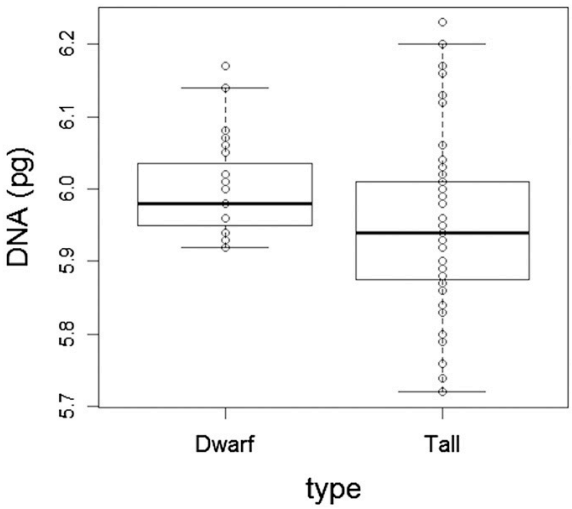


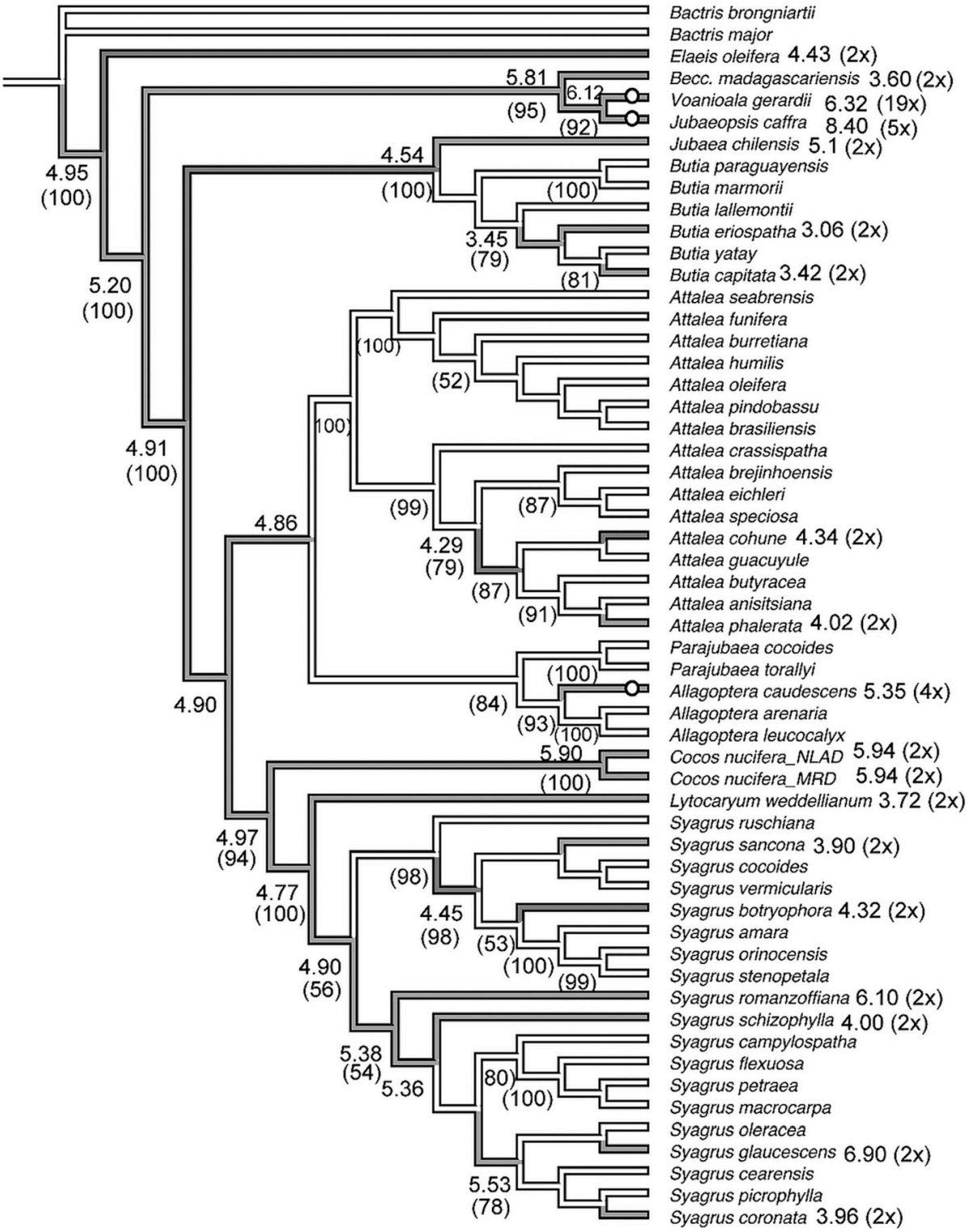
FIGURE 2 Boxplot of estimated nucleus DNA content of Dwarf and Tall types. The thick horizontal line corresponds to the median and the limits of the boxes are the first and the third quartiles. Open circles represent individual observations.

Our results reveal limited (CV = 1.7%) but significant variation in genome size in cultivars of *Cocos nucifera*. Such variations occur both in the Indo-Atlantic and in the Pacific genetic groups (respective means and confidence intervals 6.01 [5.79 – 6.25] and 5.90 [5.76 – 6.09]), but they could not be detected among Dwarfs.

Holoploid genome size in Attaleinae—Within the Attaleinae subtribe, the holoploid genome sizes were as follows: *Voanioala gerardii* J.Drans. = 60 pg (Johnson, 1989); *Allagoptera caudescens* (Mart.) Kunze = 10.70 pg; *Attalea* sp. = 4.02 – 4.34 pg; *Butia* sp. = 3.06 – 3.42 pg; *Beccariophoenix* sp. = 3.6 – 7.47 pg; *Cocos nucifera* = 5.966 ± 0.111 pg; *Jubaeopsis caffra* Becc. = 20.98 pg; *Lytocaryum weddellianum* (H.Wendl.) Toledo = 3.72 pg; and *Syagrus* sp. = 3.9 – 6.9 pg. The holoploid genome size of *Beccariophoenix madagascariensis* Jum. & H.Perrier was 3.6 pg while that of its sister taxon *Beccariophoenix alfredii* Rakotoarin et al. was almost twice as large (7.47 pg) suggesting that the latter is a tetraploid.

Reconstruction of monoploid genome size (2Cx) evolution in Attaleinae—The most recent common ancestor (TMRCA) is defined as the most recent lineage from which two diverging lineages descended. The inferred ancestral genome size of TMRCA of the Attaleinae based on the maximum likelihood topology (second internal node, Fig. 3) was 4.95 pg and it was 5.20 pg for the African/Malagasy and South American clades. The genome size of TMRCA of *Beccariophoenix* and

FIGURE 3 Ancestral genome size reconstruction: Maximum likelihood phylogenetic tree of Attaleinae based on seven *WRKY* nuclear loci using PhyML (Phylogenetic Analysis of Maximum Likelihood). ML bootstrap supports are in parenthesis below the branches. Sequence alignment is deposited in Dryad database (http://dx.doi.org/10.5061/dryad.561hm). The numbers at the nodes refer to the inferred ancestral genome sizes using maximum likelihood reconstruction approach implemented in Mesquite version 2.5 (Maddison and Maddison, 2008). Numbers adjacent to the OTUs are the genome size (2Cx) estimated using flow cytometry with ploidy levels in parenthesis, where 2× denote diploids and >2× denote polyploids. The open circles indicate the polyploidy events. Outgroups included were *Elaeis oleifera*, *Bactris major* Jacq., and *B. brongniartii* Jacq. ex Scop.


Bactris brongniartii
Bactris major
Elaeis oleifera 4.43 (2x)
Becc. madagascariensis 3.60 (2x)
Voanioala gerardii 6.32 (19x)
Jubaeopsis caffra 8.40 (5x)
Jubaea chilensis 5.1 (2x)
Butia paraguayensis
Butia marmorii
Butia lallemontii
Butia eriospatha 3.06 (2x)
Butia yatay
Butia capitata 3.42 (2x)
Attalea seabrensis
Attalea funifera
Attalea burretiana
Attalea humilis
Attalea oleifera
Attalea pindobassu
Attalea brasiliensis
Attalea crassispatha
Attalea brejinhoensis
Attalea eichleri
Attalea speciosa
Attalea cohune 4.34 (2x)
Attalea guacuyule
Attalea butyracea
Attalea anisitsiana
Attalea phalerata 4.02 (2x)
Parajubaea cocoides
Parajubaea torallyi
Allagoptera caudescens 5.35 (4x)
Allagoptera arenaria
Allagoptera leucocalyx
Cocos nucifera_NLAD 5.94 (2x)
Cocos nucifera_MRD 5.94 (2x)
Lytocaryum weddellianum 3.72 (2x)
Syagrus ruschiana
Syagrus sancona 3.90 (2x)
Syagrus cocoides
Syagrus vermicularis
Syagrus botryophora 4.32 (2x)
Syagrus amara
Syagrus orinocensis
Syagrus stenopetala
Syagrus romanzoffiana 6.10 (2x)
Syagrus schizophylla 4.00 (2x)
Syagrus campylospatha
Syagrus flexuosa
Syagrus petraea
Syagrus macrocarpa
Syagrus oleracea
Syagrus glaucescens 6.90 (2x)
Syagrus cearensis
Syagrus picrophylla
Syagrus coronata 3.96 (2x)
4.95 (100)
5.20 (100)
4.91 (100)
4.90
4.97 (94)
4.77 (100)
4.90 (56)
5.38 (54)
5.36
5.53 (78)
4.86
4.29 (79)
4.45 (98)
5.81
6.12
(95)
(92)
4.54
(100)
(100)
3.45 (79)
(81)
(100)
(52)
100)
(99)
(87)
(87)
(91)
(100)
(84)
(93)
(100)
5.90
(100)
(98)
(53)
(100)
(99)
80)
(100)

Voanioala + *Jubaeopsis* was 5.81 pg and the inferred genome size for TMRCA of *Voanioala* + *Jubaeopsis* was 6.12 pg. The inferred ancestral genome size for *Cocos nucifera* was 5.90 pg. The genome size of TMRCA of the *Cocos*/*Syagrus* + *Lytocaryum* clades was 4.97 pg and for paraphylectic *Syagrus*, the genome size of TMRCA of the two major clades was 4.90 pg. The TMRCA of *Attalea* /(*Allagoptera* + *Allagoptera* + *Parajubaea*) clades was 4.86 pg (Fig. 3). Genome size among *Butia* appears to be the smallest (3.06 pg) with inferred ancestral genome size leading to TMRCA of *Jubaea chilensis* (Molina) Baillon + *Butia* clade being 4.54 pg, showing a reduction in *Butia* but an increase in the closely related *J. chilensis* (5.1 pg).

Polyploidy in the Attaleinae—Genome size estimates in the Attaleinae subtribe suggest that polyploidy has occurred in *Beccariophoenix alfredii* (2Cx = 7.47 pg), which has twice the C-value of *B. madagascariensis* (2Cx = 3.6 pg), as well as in *Voanioala gerardii* (2Cx = 6.32 pg), *Jubaeopsis caffra* (2Cx = 8.40 pg) and *Allagoptera caudescens* (2Cx = 5.35 pg).

DISCUSSION

Genome size in coconut and its variations—Our results indicate that the genome size of the coconut is 5.966 ± 0.111 pg or 5.757 Gbp. This value differs from the results previously obtained through Feulgen- microdensitometry by Röser et al. (1997). In addition, the 4C value of *Cocos nucifera* was reported inconsistently by these authors. Indeed in Table 3 in Röser et al. (1997) the value was 14.19 pg while in the Results and Discussion section it was 10.2 pg. Our estimated value is somewhat larger than in Zonneveld et al. (2005) although it is consistent with data published by Sandoval et al. (2003) based on different cell phases.

Flow cytometry has become the predominant method for ploidy studies and determination of absolute DNA contents of cells, due to its high sample throughput and relative ease of sample preparation (Doležel and Bartos, 2005; Doležel et al., 2007). Intraspecific genome size has been shown to vary between cultivars and wild progenitors in angiosperms (Greilhuber, 2005), and such subtle changes may be detected only when using flow cytometry. Karyotyping analyses does not allow for the detection of infraspecific genome size differences because the number of chromosomes is unlikely to vary and when Feulgen- microdensitometry method is used, the presence of tannins in root tissue may interfere with the Feulgen dye causing errors in the measurement of nuclear DNA amounts (Greilhuber, 1986).

It has been proposed that genome size has a nucleotypic impact on a number of life history traits including the minimal generation time (MGT) (Bennett, 1987), which is long in the case of coconut relative to other palms. However, other factors need to be considered such as the adaptation to environmental variations. In particular, families with small genomes are more speciose (Knight et al., 2005). This is the case of Arecaceae, which is a large family with relatively small genomes among perennial plants (Zonneveld et al., 2005). The influence of nucleotype could however still hold at a more restricted evolutionary scale, i. e., the coconut genome is about 1.5 times larger than that of the African oil palm *Elaeis guineensis* Jacq. (3.76 ± 0.09 pg (Rival et al., 1997), which has a shorter MGT and a higher leaf emission rate.

We found that genome size varies significantly among coconuts. Such variation is limited (CV = 1.7%) and affects both Indo-Atlantic and Pacific groups. The genome size of the self-pollinating Dwarfs is within the range of the Talls but above average and uniform. This difference was not expected if we consider the positive correlation of genome size with MGT and the negative correlation with stomatal density. In fact, time to flowering in Talls is 4 to 5 yr, and only 2 to 3 in Dwarfs (Pillai et al., 1973).

Plant domestication is an evolutionary process that involves artificial selection and leads to population bottlenecks that can reduce the genetic diversity relative to the wild progenitors through the selection of preferred phenotypes (Doebley et al., 2006). Human selection may affect the patterns of the genome architecture of domesticated plants (Olsen and Wendel, 2013). In the case of coconuts, phenotypic traits were further influenced by consanguinity resulting from the shift from allogamy to autogamy (see Miller and Gross, 2011). This resulted in the expression of genetic load as shown by an increase in the rate of meiotic abnormalities in Dwarfs compared to Talls, by the poor endosperm development and (at least partly) by reduced vegetative vigor in Dwarfs (Swaminathan and Nambiar, 1961). However, the most likely explanation of the uniform and comparatively large genome of Dwarfs is that they were derived from a single Tall ancestor which happened to have a large genome and that this trait has not evolved since then. Coconuts (including Dwarfs) have a long generation time and the number of generations since the appearance of autogamy is probably less than 100.

***Evolution of genome size in* Attaleinae**—The Attaleinae is monophyletic and includes all members of the Cocoseae except the spiny cocosoids (Bactridinae and Elaeidinae) (see Dransfield et al., 2008). The Cocoseae tribe diverged from its closest relatives *Roystonea*/*Reinhardtia* ca. 55–58 million years ago (mya). Its spiny and non-spiny members diverged about 46 mya (Gunn, 2004; Roncal et al., 2013). Most Attaleinae are diploid while *Allagoptera caudescens*, *Beccariophoenix alfredii*, *Jubaeopsis caffra*, and *Voanioala* have undergone polyploidization events in the past and have retained a duplicated genome. A study by Shapcott et al. (2007) on the genetic diversity of the diploid *Beccariophoenix madagascariensis* found highly inbred populations. Microsatellite data did not show differentiation between *B. alfredii*, and the northern *B. madagascariensis* population. It is possible that selfing within these northern populations led to polyploidy with subsequent dispersal by frugivores to new habitats thus resulting in speciation. Including *Beccariophoenix alfredii*, a tetraploid shown in this current study, we found that polyploidy occurred at least four times within the Cocoseae.

Our phylogenetic analysis suggests that the ancestral genome size for the Attaleinae may have been small (ca. 4.95 pg). We observed some variability in genome size at the generic level but genome size within a given genus was broadly conserved except for *Syagrus glaucescens* Glaz. ex Becc. and *S. romanzoffiana* (Cham.) Glassman (Fig. 3). The Attaleinae diversified in South America and formed highly speciose taxa such as *Syagrus*, *Attalea*, and *Butia*. In general, their genome sizes are much smaller than the species-poor Malagasy/African clade (*Beccariophoenix*, *Voanioala*, and *Jubaeopsis*) and it is possible that small genome size provided competitive advantage for the South American taxa allowing them to diversify into different biomes. Small genome size has been shown to correlate with shorter minimum generation time (MGT), increased reproductive rate and reduced reproductive costs especially in perennial diploid monocots (Bennett, 1972; Midgley, 1991). Our study suggests a role for domestication in genome size evolution and revealed that polyploidy is common within the Attaleinae (3 out of 17 species sampled), relative to the rest of Cocoseae and has evolved multiple times independently.

Toward coconut genome sequencing—Our research has implications for the future sequencing and annotation of coconut nuclear genome. To date, the genome of two economically important palms have been sequenced and published namely for the date palm *Phoenix dactylifera* L. (estimated 1C ~671Mb by Al-Mssallem et al. (2013)) and the African oil palm *Elaeis guineensis* (1C ~1.8 Gb according to Singh et al. (2013)). There is also a draft genome sequence available for *E. oleifera* (Kunth) Cortes (Filho et al., 2015). Long generation time and bulkiness make coconut breeding a lengthy process. Thus, marker-assisted selection and genomic breeding are likely to accelerate progress in genetic manipulation. Transcriptomes produced through Next Generation Sequencing have already been published by Fan et al. (2013) and Huang et al. (2014). A preliminary draft coconut genome sequence was presented by Alsaihati et al. (2014) without prior estimation of genome size and variation among cultivars. The coconut genome is 4 and 1.6 times larger than the date palm and oil palm, respectively, and as such requires a much deeper sequencing effort. In addition, a larger genome means that more repeated sequences are present thus causing increased difficulty for assembly. This difficulty can however be overcome by combining the extension of scaffold using paired-end generation of large sequences with the production of a high density linkage map. Whole-genome sequencing will pave the way to a variety of approaches such as Single Nucleotide Polymorphism discoveries from genome wide association studies (GWAS). The whole genome sequence of the coconut will provide us with insights into decoding the traits associated with fruit morphology and more importantly to enable the discovery of Quantitative Trait Loci associated with disease resistance such as lethal yellowing through association studies and mapping. Comparative genomics involving oil palm and date palm genome sequence will help elucidate key cellular and physiological mechanisms among Arecaceae.

ACKNOWLEDGEMENTS

The authors thank the Centre National de la Recherche Agronomique de Côte d'Ivoire for providing fresh coconut samples, Spencer Brown (CNRS, France) for providing the selected *Petunia* seeds, and the IRB (Institut de Recherche en Biothérapie) in Montpellier for kindly providing access to the flow cytometry facilities at Montpellier Rio Imaging. The Australian National Botanic Gardens (Canberra) are gratefully thanked for providing the coconut leaf material from Lizard Island, and the Royal Botanic Gardens (Sydney) for material from Mission Beach, Malayan Yellow Dwarf type, and Cocoseae. The authors acknowledge editorial support from Prof. Patricia Blackwell, Research Sch. of Biol., Div. of Evolution, Ecology and Genetics, The Australian National University. They are also thankful to anonymous reviewers for their useful comments.

LITERATURE CITED

Abraham, A., and P. M. Mathew. 1963. Cytology of coconut endosperm. *Annals of Botany* 27: 505–513.

Adams, K. L., and J. F. Wendel. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* 8: 135–141.

Al-Dous, E. K., G. Binu, M. E. Al-Mahmoud, M. Y. Al-Jaber, H. Wang, Y. M. Salameh, E. K. Al-Azwani, et al. 2011. *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnology* 29: 521–527.

Al-Mssallem, I. S., S. Hu, X. Zhang, Q. Lin, W. Liu, J. Tan, X. Yu, et al. 2013. Genome sequence of the date palm *Phoenix dactylifera* L. *Nature Communications* 4: 1–9.

Alsaihati, B., W. Liu, Q. Lin, and I. S. AlMssallem. 2014. Coconut genome *de novo* sequencing. Poster PO46, Genome sequencing and ESTs session, January 13, 2014. Plant and Animal Genome XXII (International Plant and Animal Genome Conference), San Diego, California, USA.

Baudouin, L., B. F. Gunn, and K. M. Olsen. 2014. The presence of coconut in southern Panama in pre-Columbian times: Clearing up the confusion. *Annals of Botany* 113: 1–5.

Bennett, M. D. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. *Proceedings of the Royal Society of London, B, Biological Sciences* 181: 109–135.

Bennett, M. D. 1987. Variations in genomic form in plants and its ecological implications. *New Phytologist* 106: 177–200.

Bennett, M. D. 1998. Plant genome values: How much do we know? *Proceedings of the National Academy of Sciences, USA* 95: 2011–2016.

Bennett, M. D., P. Bhandol, and I. J. Leitch. 2000. Nuclear DNA amounts un Angiosperms and their modern uses—807 new estimates. *Annals of Botany* 86: 859–909.

Bourdeix, R., P. Batugal, J. Oliver, and M. George. 2010. Catalogue of conserved coconut germplasm. International Plant Genetic Resources International (IPGRI) [now Biodiversity International], Maccarese, Italy.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey. 1983. Graphical methods for data analysis.Wadsworth, Belmont, California, USA.

Church, G. L. 1940. Cytotaxonomic studies in Gramineae: *Spartina, Andropogon,* and *Panicum*. *American Journal of Botany* 27: 263–271.

Coba de la Peña, T., and S. C. Brown. 2001. Flow cytometry. *In* C. Hawes and B. Satiat-Jeunemaitre [eds.], Plant cell biology: A practical approach. Oxford University Press, Oxford, UK.

Comai, L. 2005. The advantages and disadvantages of being a polyploid. *Nature Reviews. Genetics* 6: 836–846.

Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. The molecular genetics of crop domesticatiom. *Cell* 127: 1309–1321.

Doležel, J., and J. Bartos. 2005. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* 95: 99–110.

Doležel, J., P. Binarova, and S. Lucretti. 1989. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* 31: 113–120.

Doležel, J., J. Greilhuber, and J. Suda. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nature Protocols* 2: 2233–2244.

Dransfield, J., N. W. Uhl, C. B. Asmussen-Lange, W. J. Baker, M. M. Harley, and C. E. Lewis. 2008. Genera palmarum: The evolution and classification of palms. Royal Botanic Gardens, Kew, Richmond, UK.

Fan, H., Y. Xiao, Y. Yang, W. Xia, A. Mason, Z. Xia, F. Qiao, et al. 2013. RNA-Seq analysis of *Cocos nucifera*: Transcriptome sequencing and de *novo* assembly for subsequent functional genomics approaches. *PLoS One* 8: e59997.

Filho, J. A. F., F. P. Lobo, M. T. Souza Junior, A. A. Alves, and E. F. Formighieri. 2015. Genome assembly of Amazonian *Elaeis oleifera* genotype and comparative analysis with other oil palm genomes. Poster P0008, Bioenergy-even session, January 12, 2015, Plant and Animal Genome XXIII (International Plant and Animal Genome Conference), San Diego, California, USA.

Foale, M. 2005. An introduction to the coconut palm. *In* P. Batugal, V. R. Rao, and J. Oliver [eds.], Coconut genetic resources, 1-10. International Plant Genetic Resources Institute-Regional Office for Asia, the Pacific and Oceania (IPGRI-APO), Serdang, Selangor, Malaysia.

Gaut, B. S., and J. F. Doebley. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences, USA* 94: 6809–6814.

Greilhuber, J. 1986. Severely distorted Feulgen-DNA amounts in *Pinus* (Coniferophytina) after nonadditive fixations as a result of meristematic self-tanning with vacuole contents. *Canadian Journal of Genetics and Cytology* 28: 409–415.

Greilhuber, J. 2005. Intraspecific variation in genome size in angiosperms: Identifying its existence. *Annals of Botany* 95: 91–98.

Greilhuber, J., J. Doležel, M. A. Lysák, and M. D. Bennett. 2005. The origin, evolution and proposed stabilization of the terms "Genome size" and "C-value" to describe nuclear DNA contents. *Annals of Botany* 95: 255–260.

Grime, J. P. 1983. Prediction of weed and crop response to climate based upon measurements of nuclear DNA content. *In* Aspects of Applied Biology, vol. 4, 87-98. National Vegetable Research Station, Wellsbourne. Association of Applied Biologists, Warwick, UK.

Grime, J. P., and M. A. Mowforth. 1982. Variation in genome size—an ecological interpretation. *Nature* 299: 151–153.

Guindon, S., and O. Gasceul. 2003. PhyML: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696–704.

Gunn, B. F. 2004. The phyogeny of the Cocoeae (Arecaceae) with emphasis on *Cocos nucifera*. *Annals of the Missouri Botanical Garden* 91: 505–522.

Gunn, B. F., L. Baudouin, and K. M. Olsen. 2011. Independent origins of cultivated coconut (*Cocos nucifera* L.) in the Old World Tropics. *PLoS One* 6: e21143.

Hardie, D. C., and P. D. Hebert. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* 46: 683–706.

Harries, H. 1978. The evolution, dissemination and classification of *Cocos nucifera* L. *Botanical Review* 44: 265–319.

Heslop-Harrison, J. S., and T. Schwarzacher. 2007. Domestication, genomics and the future for banana. *Annals of Botany* 100: 1073–1084.

Huang, Y.-Y., C.-P. Lee, J. L. Fu, B. C.-H. Chang, A. J. M. Matzke, and M. Matzke. 2014. *De novo* transcriptome sequence assembly from coconut leaves and seeds with a focus on factors involved in RNA-directed DNA methylation. *G3: Genes|Genomes|Genetics* 4: 2147–2157.

Johnson, M. A. 1989. An unusually high chromosome number in *Voanioala gerardii* (Palmae: Arecoideae: Cocoeae: Butiinae) from Madagascar. *Kew Bulletin* 44: 207–210.

Knight, C. A., N. A. Molinari, and D. Petrov. 2005. Large genome constraint hypothesis: Evolution, ecology and phenotype. *Annals of Botany* 95: 177–190.

Laurie, D. A., and M. D. Bennett. 1985. Nuclear DNA content in the genera *Zea* and *Sorghum*. Intergeneric, interspecific and intraspecific variation. *Heredity* 55: 307–313.

Lebrun, P., A. Berger, T. Hodgkin, and L. Baudouin. 2005. Biochemical and molecular methods for characterizing coconut diversity. *In* P. Batugal, V. R. Rao, and J. Oliver [eds.], Coconut genetic resources, 225-247. International Plant Genetic Resources Institute-Regional Office for Asia, the Pacific and Oceania (IPGRI-APO), Serdang, Selangor, Malaysia.

Maddison, W. P., and D. R. Maddison. 2008. Mesquite, a modular system for evolutionary analysis. Version 2.5. Computer program distributed by authors, http://mesquiteproject.org [accessed 22 June 2014].

Masterson, J. 1994. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* 264: 421–424.

Meerow, A., L. Noblick, J. W. Borrone, T. Couvreur, M. Mauro-Herrera, W. Hahn, D. N. Kuhn, et al. 2009. Phylogenetic analyses of seven *WRKY* genes across the palm subtribe Attaleinae (Arecaceae) identifies *Syagrus* as sister group of the coconut. *PLoS One* 4: e7353.

Midgley, J. J. 1991. Ecological aspects of the rise of angiosperms: A challenge to the reproductive superiority hypotheses. *Biological Journal of the Linnean Society. Linnean Society of London* 44: 81–92.

Miller, A. J., and B. L. Gross. 2011. From forest to field: Perennial fruit crop domestication. *American Journal of Botany* 98: 1389–1414.

Nambiar, M. C., and M. S. Swaminathan. 1960. Chromosome morphology, microsporogenesis and pollen fertility in some varieties of coconut. *Indian Journal of Genetics and Plant Breeding* 20: 200–211.

Olsen, K. M., and J. F. Wendel. 2013. A bountiful harvest: Genomic insights into crop domestication phenotypes. *Annual Review of Plant Biology* 64: 47–70.

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* 48: 612–622.

Parris, J. K., T. G. Ranney, H. T. Knap, and W. V. Baird. 2010. Ploidy levels, relative genome sizes, and base pair composition in Magnolia. *Journal of the American Society for Horticultural Science* 135: 533–547.

Pichersky, E. 1990. Nomad DNA—A model for movement and duplication of DNA sequences in plant genomes. *Plant Molecular Biology* 15: 437–448.

Pillai, R., R. Nair, C. Mathew, K. Bavappa, and A. Ramadasan. 1973. Studies on photoperiodic response reaction in coconut. *Journal of Plantation Crops* 1: 89–92.

Price, J. P., and M. Baranova. 1976. Evolution of DNA content in higher plants. *Botanical Review* 42: 27–52.

R Development Core Team. 2011. R: A language and environment for statistical computing. Computer program version 3.1.1 distributed by R Foundation for Statistical Computing, website: http://www.R-project.org [accessed on 22 June 2014].

Riley, R. D., and K. P. Vogel. 1982. Chromosome numbers of released cultivars of switchgrass, indiangrass, big bluestem, and sand blue-stem. *Crop Science* 22: 1082–1083.

Rival, A., T. Beule, P. Barre, S. Hamon, Y. Duval, and M. Noirot. 1997. Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq.) tissue cultures and seed-derived plants. *Plant Cell Reports* 16: 884–887.

Roncal, J., F. Kahn, B. Millan, T. Couvreur, and J.-C. Pintaud. 2013. Cenozoic colonization and diversification patterns of tropical American palms: Evidence from *Astrocaryum* (Arecaceae). *Botanical Journal of the Linnean Society. Linnean Society of London* 171: 120–139.

Röser, M., M. A. T. Johnson, and L. Hanson. 1997. Nuclear DNA amounts in palms (Arecaceae). *Botanica Acta* 110: 79–89.

Sandoval, A., V. Hocher, and J.-L. Verdeil. 2003. Flow cytometric analysis of the cell cycle in different coconut palm (*Cocos nucifera* L.) tissues cultured in vitro. *Plant Cell Reports* 22: 25–31.

Sauer, J. D. 1971. A re-evaluation of the coconut as an indicator of human dispersal. *In* C. Riley [ed.], Man across the sea, 309-319. University of Texas Press, Austin, Texas, USA.

Shapcott, A., M. Rakotoarinivo, R. J. Smith, G. Lysaova, M. F. Fay, and J. Dransfield. 2007. Can we bring Madagascar's critically endangered palms back from the brink? Genetics, ecology and conservation of the critically endangered palm *Beccariophoenix madagascariensis*. *Botanical Journal of the Linnean Society. Linnean Society of London* 154: 589–608.

Singh, R., M. Ong-Abdullah, E.-T. Low, M. A. A. Manaf, and R. Rosli. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. *Nature* 500: 335–339.

Sparrow, A. H., and J. P. Miksche. 1961. Correlations of nuclear volume and DNA content with higher tolerance to chronic radiation. *Science* 134: 282–283.

Sparrow, A. H., S. S. Schwemmer, and P. J. Bottino. 1971. The effects of external gamma radiation from radioactive fallout of plants with special reference to crop production. *Radiation Botany* 11: 85–118.

Sparrow, R. C., and A. H. Sparrow. 1965. Relative radiosensitivities of woody and herbaceous spermatophytes. *Science* 147: 1449–1451.

Stebbins, G. L. 1959. The role of hybridization in evolution. *Proceedings of the American Philosophical Society* 103: 231–251.

Swaminathan, M. S., and M. C. Nambiar. 1961. Cytology and origin of the dwarf coconut palm. *Nature* 192: 85–86.

Swift, H. 1950. The constancy of desoxyribose nucleic acid in plant nuclei. *Proceedings of the National Academy of Sciences, USA* 36: 643–654.

Wendel, J. F., and R. C. Cronn. 2003. Polyploidy and the evolutionary history of cotton. *Advances in Agronomy* 78: 139–186.

Zonneveld, B. J. M., I. J. Leitch, and M. D. Bennett. 2005. First nuclear DNA amounts in more than 300 Angiosperms. *Annals of Botany* 96: 229–244.

APPENDIX 1 Absolute genome sizes (pg) estimated for *Cocos nucifera* cultivars sampled with *Petunia xhybrida* internal standard, from flow cytometry.

Cultivar	Internat. abbrev.	Habit	N	Abs. genome size (2C)/pg (mean ± SD)	Origin	Collection Locality
Andaman Ordinary Tall	ADOT	Tall	4	6.02 ± 0.09	Andaman Island	Sta. MD_L03A13
Cameroon Kribi Tall	CKT	Tall	2	5.87 ± 0.20	Cameroon	Sta. MD_L12A09
Gazelle Peninsula Tall	GPT	Tall	3	5.89 ± 0.08	Papua New Guinea	Sta. MD_L08A12
Mission Beach	MISB	Tall (ws)	2	5.87 ± 0.00	Australia	RBG SYD_20101370
Lizard Island Tall	LIZ	Tall (ws)	4	5.89 ± 0.05	Australia	ANBG_BG753A
Laccadive Micro Tall	LMT	Tall	3	6.13 ± 0.00	LaccadiveArchipelago	Sta. MD_L08A18
Malayan Tall	MLT	Tall	4	5.79 ± 0.06	Malaysia	Sta.MD_L03A18
Mozambique Tall	MZT	Tall	3	6.19 ± 0.04	Mozambique	Sta. MD_L03A13
Panama Tall	PNT	Tall	4	6.01 ± 0.03	Panama	Sta. MD_L03A12
Solomon Island Tall	SIT	Tall	3	5.96 ± 0.03	Solomon Islands	Sta. MD_L21A13
Sri Lanka Tall	SLT	Tall	4	6.07 ± 0.08	Sri Lanka	Sta. MD_L36A24
Tagnanan Tall	TAGT	Tall	3	5.93 ± 0.00	Philippines	Sta. MD_L38A25
Tahiti Tall	TAT	Tall	3	5.75 ± 0.03	Tahiti	Sta. MD_L03A08
Vanuatu Tall	VTT	Tall	3	5.95 ± 0.03	Vanuatu	Sta. MD_L44A24
West African Tall	WAT3	Tall	6	5.89 ± 0.06	West Africa	Sta. MD_L09A14
Brazil Green Dwarf	BGD	Dwarf	4	5.94 ± 0.03	Brazil	Sta. MD_L13A28
Cameroon Red Dwarf	CRD	Dwarf	3	6.02 ± 0.02	Cameroon	Sta. MD_L06A13
Catigan Green Dwarf	CATD	Dwarf	4	6.04 ± 0.04	Philippines	Sta. MD_L05A15
Ghana Yellow Dwarf	GYD	Dwarf	3	5.96 ± 0.03	Ghana	Sta. MD_L02A30
Malayan Yellow Dwarf	MYD	Dwarf	2	5.94 ± 0.02	Malaysia	RBG SYD_903153
Pilipog Green Dwarf	PILD	Dwarf	6	6.01 ± 0.08	Philippines	Sta. MD_L35A28
Tahiti Red Dwarf	TRD	Dwarf	3	6.04 ± 0.13	Tahiti	Sta. MD_L14A26
Niu Leka Dwarf	NLAD	Compact Dwarf	4	5.94 ± 0.06	Fiji	Sta. MD_L08A09

Appendix 1 Abbreviations: Sta. MD = CNRA Marc Delorme Coconut Research Centre in Côte d'Ivoire, Africa; ANBG = Australian National Botanic Gardens Canberra; RBG SYD = Royal Botanic Gardens Sydney, Australia; ws = wild-sown, one individual per cultivar was assayed; and N = number of runs per individual.

APPENDIX 2 Absolute genome sizes (pg) 2Cx values estimated for Cocoseae species

Species	Abs. genome size /pg (2Cx)	Ploidy level	Collection locality	Accession number
Allagoptera caudescens (Mart.) Kunze	5.35	4	RBG, Sydney	20091679
Attalea cohune Mart.	4.34	2	RBG, Sydney	20091583
Attalea phalerata Mart. ex Spreng.	4.02	2	RBG, Sydney	20091585
Astrocaryum alatum H.F.Loomis	4.36	2	RBG, Sydney	20091582
Bactris bifida Mart.	4.10	2	RBG, Sydney	20091209
Bactris gasipaes Kunth	9.43	4	RBG, Sydney	20100250
Beccariophoenix alfredii Rakotoarin et al.	7.47	4	RBG, Sydney	20100251
Beccariophoenix madagascariensis Jum. & H.Perrier	3.60	2	RBG, Sydney	20040914
Butia capitata (Mart.) Becc.	3.42	2	RBG, Sydney	932392
Butia eriospatha (Mart. ex Drude) Becc.	3.06	2	RBG, Sydney	780035
Elaeis oleifera (Kunth) Cortes	4.43	2		Angiosperm 1C-values db
Jubaea chilensis (Molina) Baill.	5.10	2	RBG, Sydney	20090098
Jubaeopsis caffra Becc.	8.40	5		801080
Lytocaryum weddellianum (H.Wendl.) Toledo	3.72	2	RBG, Sydney	14451
Syagrus botryophora (Mart.) Mart.	4.32	2	RBG, Sydney	20090788
Syagrus coronata (Mart.) Becc.	3.96	2	RBG, Sydney	20091730
Syagrus glaucescens Glaz. ex Becc.	6.90	2		Angiosperm 1C-values db
Syagrus romanzoffiana (Cham.) Glassman	6.10	2		Angiosperm 1C-values db
Syagrus sancona (Kunth) H.Karst.	3.90	2	RBG, Sydney	20091729
Syagrus schizophylla (Mart.) Glassman	4.00	2	RBG, Sydney	20091652
Voanioala gerardii J.Dransf.	6.32	19		Angiosperm 1C-values db

Note: The Cx-value is defined as the (averaged) DNA content of the monoploid genomes in polyploids and nonpolyploids, where x is the symbol for the chromosome number of the monoploid genome and for the chromosome base number in a generatively polyploid series of related organisms (see Greilhuber et al., 2005).

APPENDIX 3 (Chapter 3)

List of coconut samples, region, name, identification number, country of origin and country abbreviation.

Number	*Region*	*Name*	*ID_num.*	*Country*	*Count. Abbrev.*
1	Indo-Atlantic	West African Tall	01_WAT	Africa	AFR
2	Indo-Atlantic	West African Tall	13_WAT	Africa	AFR
3	Australasia	Lizard Island Tall	130_AUS	Australia	AUS
4	Australasia	Lizard Island Tall	131_AUS	Australia	AUS
5	Australasia	Lizard Island Tall	132_AUS	Australia	AUS
6	Australasia	Lizard Island Tall	134_AUS	Australia	AUS
7	Australasia	Lizard Island Tall	135_AUS	Australia	AUS
8	Australasia	Lizard Island Tall	136_AUS	Australia	AUS
9	Australasia	Lizard Island Tall	138_AUS	Australia	AUS
10	Australasia	Lizard Island Tall	139_AUS	Australia	AUS
11	Australasia	Lizard Island Tall	140_AUS	Australia	AUS
12	Australasia	Lizard Island Tall	141_AUS	Australia	AUS
13	Australasia	Lizard Island Tall	142_AUS	Australia	AUS
14	Indo-Atlantic	Brazil Tall	96_BRT	Brazil	BRA
15	Brazil	*Syagrus picrophylla*	40_S.PIC	Brazil	BRA
16	Brazil	*Syagrus coronata*	51_S.COR	Brazil	BRA
17	Indo-Atlantic	Brazil Tall	24_BRT	Brazil	BRA
18	Southeast Asia	Cambodian Tall	05_KAT	Cambodia	KHM
19	Southeast Asia	Cambodian Tall	88_KAT	Cambodia	KHM
20	Indo-Atlantic	Cameroon Kribi Tall	74_CKT	Cameroon	CMR
21	Indo-Atlantic	Comoros Anjouan Tall	28_CMRT	Comoros	COM
22	Indo-Atlantic	Comoros Anjouan Tall	76_CMRT	Comoros	COM
23	Indo-Atlantic	Comoros Anjouan Tall	22_CMRT	Comoros	COM
24	Indo-Atlantic	Comoros Anjouan Tall	58_CMRT	Comoros	COM
25	Indo-Atlantic	Comoros Moheli Tall	85_CMT	Comoros	COM
26	South Pacific	Cook Island Tall	79_COKT	Cook Islands	COK
27	South Pacific	Niu magimagi Tall	18_NNMT	Fiji	FJI
28	South Pacific	Niu magimagi Tall	30_NNMT	Fiji	FJI
29	South Pacific	Fiji Tall	31_TAV	Fiji	FJI
30	South Pacific	Fiji Tall	34_TAV	Fiji	FJI
31	South Pacific	Fiji Tall Korolevu	107_TAV	Fiji	FJI
32	South Pacific	*Niu Leka* Dwarf	37_NLAD	Fiji	FJI
33	South Pacific	*Niu Leka* Dwarf	52_NLAD	Fiji	FJI
34	South Pacific	Rotuma Tall	63_RTMT	Fiji	FJI
35	South Pacific	*Niu Leka* Dwarf	68_NLAD	Fiji	FJI
36	South Pacific	*Niu drau*	144_TAV	Fiji	FJI
37	Indo-Atlantic	Andaman Ordinary Tall	86_ADOT	India	IND
38	Indo-Atlantic	Chowgat Green Dwarf	92_CH	India	IND
39	Indo-Atlantic	Kapadam Tall	03_KPDT	India	IND
40	Indo-Atlantic	Laccadive Micro Tall	14_LMT	India	IND
41	Indo-Atlantic	Laccadive Micro Tall	26_LMT	India	IND
42	Australasia	Tenga Tall	04_TGT	Indonesia	IDN
43	Australasia	*Nyior biasa*	07_IDN	Indonesia	IDN
44	Australasia	*Nyior pendek*	11_IDN	Indonesia	IDN
45	Australasia	Palu Tall	16_PUT	Indonesia	IDN
46	Australasia	*Nyior biasa*	35_IDN	Indonesia	IDN
47	Australasia	*Nyior biasa*	44_IDN	Indonesia	IDN
48	Australasia	Flores Tall	57_IDN	Indonesia	IDN
49	Australasia	Ternate Brown Dwarf	65_TBD	Indonesia	IDN
50	Australasia	*Nyior biasa*	67_IDN	Indonesia	IDN
51	Australasia	*Nyior panda hijau*	70_IDN	Indonesia	IDN
52	Australasia	*Nyior biasa*	81_IDN	Indonesia	IDN
53	Australasia	*Nyior meta hijau*	83_IDN	Indonesia	IDN
54	Australasia	Takome Tall	87_TKT	Indonesia	IDN
55	Australasia	*Nyior sangnu (Kopyior)*	97_IDN	Indonesia	IDN

56	Australasia	Deli	100_IDN	Indonesia	IDN
57	Australasia	Matahari	102_IDN	Indonesia	IDN
58	Indo-Atlantic	Jamaican Tall	106_ALT	Jamaica	JAM
59	South Pacific	Kiribat Tall	20_KIT	Kiribats	KIT
60	South Pacific	Kiribat Green Dwarf	27_KIGD	Kiribats	KIT
61	Indo-Atlantic	Madagascar Tall	98_MDGT	Madagascar	MAD
62	Indo-Atlantic	Madagascar Tall	99-MDGT	Madagascar	MAD
63	Indo-Atlantic	Madagascar Tall	101_MDGT	Madagascar	MAD
64	Indo-Atlantic	Madagascar Tall	103_MDGT	Madagascar	MAD
65	Indo-Atlantic	Madagascar Tall	104_MDGT	Madagascar	MAD
66	Indo-Atlantic	Madagascar Tall	105_MDGT	Madagascar	MAD
67	Indo-Atlantic	Madagascar Tall	137_MDGT	Madagascar	MAD
68	Southeast Asia	Malayan Tall	02_MLT	Malaysia	MYS
69	Southeast Asia	Yellow Malayan Dwarf	45_YMD	Malaysia	MYS
70	Southeast Asia	Malayan Red Dwarf	77_MRD	Malaysia	MYS
71	Southeast Asia	Malayan Green Dwarf	89_MGD	Malaysia	MYS
72	South Pacific	Marshall Island Tall	82_MIT	Marshall Island	MHL
73	South Pacific	Marshall Island Tall	94_MIT	Marshall Islands	MHL
74	South Pacific	New Caledonia Tall	55_NCT	New Caledonia	NCL
75	Panama	Panama Tall	25_PNT	Panama	PAN
76	Panama	Panama Tall Costa Rica	36_PNT	Panama	PAN
77	Panama	Panama Tall Aguadulce	50_PNT	Panama	PAN
78	Panama	Panama Tall	60_PNT	Panama	PAN
79	Panama	Panama Tall Aguadulce	62_PNT	Panama	PAN
80	Australasia	Vailala Tall	09_VLT	Papua New Guinea	PNG
81	Australasia	Markam Valley Tall	10_MVT	Papua New Guinea	PNG
82	Australasia	Kiwai Tall	21_KWT	Papua New Guinea	PNG
83	Australasia	East Sepik Tall	32_ELT	Papua New Guinea	PNG
84	Australasia	West New Britain Tall	33_WLT	Papua New Guinea	PNG
85	Australasia	Baibarra Tall	43_BBRT	Papua New Guinea	PNG
86	Australasia	New Ireland Tall	46_NLT	Papua New Guinea	PNG
87	Australasia	HihishuTall	56_HLT	Papua New Guinea	PNG
88	Australasia	Madang Yellow Tall	69_MADY	Papua New Guinea	PNG
89	Australasia	Karkar Tall	71_KKT	Papua New Guinea	PNG
90	Australasia	Poligolo Tall	80_PLT	Papua New Guinea	PNG
91	Australasia	Karkar Tall	93_KKT	Papua New Guinea	PNG
92	Southeast Asia	Kapatangan Green Dw.	06_KAPD	Philippines	PHL
93	Southeast Asia	Macapuno Tall	12_MACT	Philippines	PHL
94	Southeast Asia	Baybay Tall	15_BAYT	Philippines	PHL
95	Southeast Asia	San Ramon Tall	19_SNRT	Philippines	PHL
96	Southeast Asia	Tacunan Green Dwarf	29_TACD	Philippines	PHL
97	Southeast Asia	Pilipog Green Dwarf	41_PILD	Philippines	PHL
98	Southeast Asia	Catigan Green Dwarf	29_CATD	Philippines	PHL
99	Southeast Asia	Tagnanan Tall	61_TAGT	Philippines	PHL
100	Southeast Asia	Ballesteros Tall	90_BALT	Philippines	PHL
101	Indo-Atlantic	Seychelles Tall	17_SEY	Seychelles	SEY
102	Indo-Atlantic	Seychelles Tall	108_SEY	Seychelles	SEY
103	Indo-Atlantic	Seychelles Tall	112_SEY	Seychelles	SEY
104	Indo-Atlantic	Seychelles Tall	118_SEY	Seychelles	SEY
105	South Pacific	Solomon Island Tall	38_SIT	Solomons	SOL
106	South Pacific	Rennell Island Tall	73_RIT	Solomons	SOL
107	South Pacific	*Niu mweta*	113_SOL	Solomons	SOL
108	South Pacific	*Niu marawa*	114_SOL	Solomons	SOL
109	South Pacific	*Niu marawa*	115_SOL	Solomons	SOL
110	South Pacific	*Niu mera*	116_SOL	Solomons	SOL
111	South Pacific	*Niu mafu*	117_SOL	Solomons	SOL
112	South Pacific	*Niu tangarau*	119_SOL	Solomons	SOL
113	South Pacific	*Niu fara*	120_SOL	Solomons	SOL
114	South Pacific	*Niu ngasi*	121_SOL	Solomons	SOL
115	Indo-Atlantic	Sri Lanka Tall	42_SLT	Sri Lanka	LKA
116	Indo-Atlantic	Pumila Green Dwarf	53_PGD	Sri Lanka	LKA
117	Indo-Atlantic	Sri Lanka Tall	54_SLT	Sri Lanka	LKA

118	Indo-Atlantic	Rath Thembili	66_RTB	Sri Lanka	LKA
119	Indo-Atlantic	East African Tall	47_EAT	Tanzania	TZA
120	Indo-Atlantic	East African Tall	59_EAT	Tanzania	TZA
121	Indo-Atlantic	East African Tall	78_EAT	Tanzania	TZA
122	Southeast Asia	Aromatic Green Dwarf	23_AROD	Thailand	THD
123	Southeast Asia	Thailand Green Dwarf	84_THD	Thailand	THD
124	South Pacific	*Niu mea*	109_TIK	Tikopia	TIK
125	South Pacific	*Niu mea*	110_TIK	Tikopia	TIK
126	South Pacific	*Niu uvea*	111_TIK	Tikopia	TIK
127	South Pacific	Tonga Tall	75_TONT	Tonga	TON
128	South Pacific	Tuvalu Tall	91_TUV	Tuvalu	TUV
129	South Pacific	Tuvalu Tall	08_TUV	Tuvalu	TUV
130	South Pacific	Vanuatu Red Dwarf	39_VRD	Vanuatu	VUT
131	South Pacific	Vanuatu Tall	72_VTT	Vanuatu	VUT
132	South Pacific	*Nean anelec*	122_VTT	Vanuatu	VUT
133	South Pacific	*Nean iwyeugd*	123_VTT	Vanuatu	VUT
134	South Pacific	*Nean rek*	124_VTT	Vanuatu	VUT
135	South Pacific	*Nean imtanor*	125_VTT	Vanuatu	VUT
136	South Pacific	*Nean nohan*	126_VTT	Vanuatu	VUT
137	South Pacific	*Nean apojev*	127_VTT	Vanuatu	VUT
138	South Pacific	*Nean induaa inpuad*	128_VTT	Vanuatu	VUT
139	South Pacific	*Syagrus amara*	S. AMA	West Indies	WI

APPENDIX 4

Sequencing QC Report
Based upon: 46,060,022 sequences in 1 data set
Generated by: kulhcar
Creation date: Mon Nov 12 15:22:28 EST 2012
Software: CLC Genomics Workbench 5.5.1

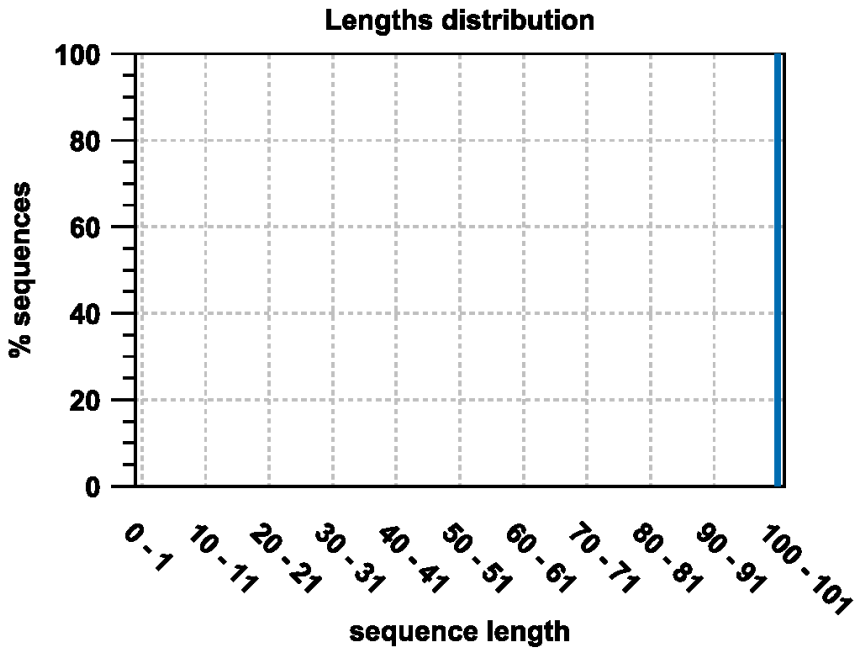
Table of contents

1. Summary

Creation date:	Mon Nov 12 15:22:28 EST 2012
Generated by:	kulhcar
Software:	CLC Genomics Workbench 5.5.1
Based upon:	1 data set
Coco1_Elox3_CAGATC_L007_R1_001 (paired):	46,060,022 sequences in pairs

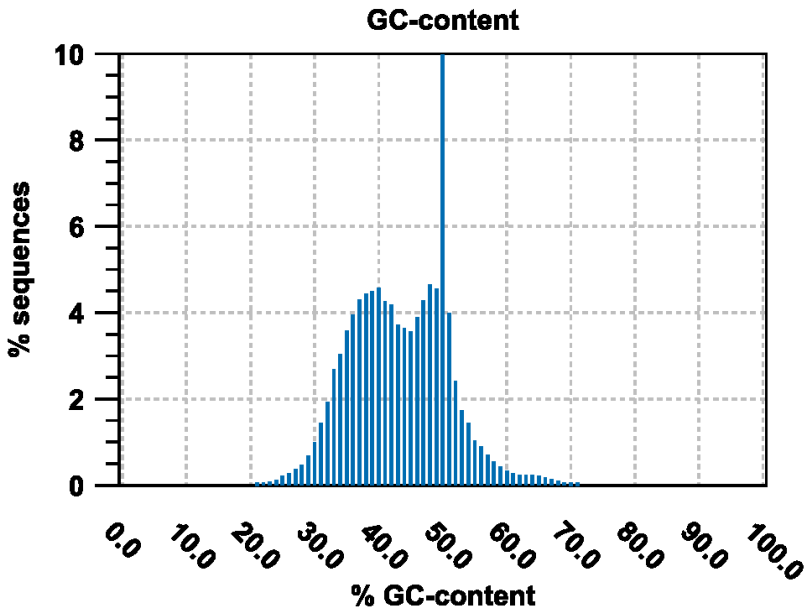
2. Per-sequence analysis

2.1 Lengths distribution



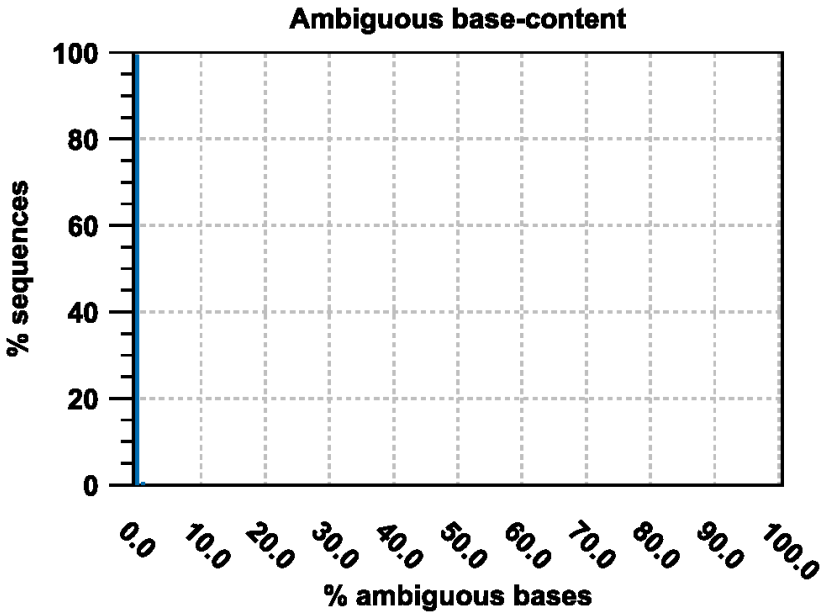
Distribution of sequence lengths. In cases of untrimmed Illumina or SOLiD reads it will ju st contain a single peak.
x: sequence length in base-pairs
y: number of sequences featuring a particular length normalized to the total number of seq uences

2.2 GC-content



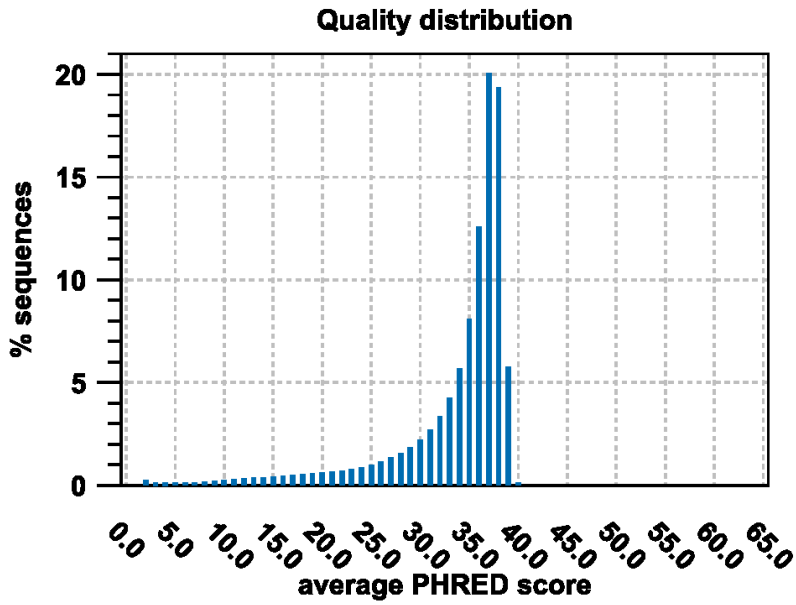
Distribution of GC-contents. The GC-content of a sequence is calculated as the number of G C-bases compared to all bases (including ambiguous bases).
x: relative GC-content of a sequence in percent
y: number of sequences featuring particular GC-percentages normalized to the total number of sequences

2.3 Ambiguous base-content



Distribution of N-contents. The N-content of a sequence is calculated as the number of amb iguous bases compared to all bases.
x: relative N-content of a sequence in percent
y: number of sequences featuring particular N-percentages normalized to the total number o f sequences

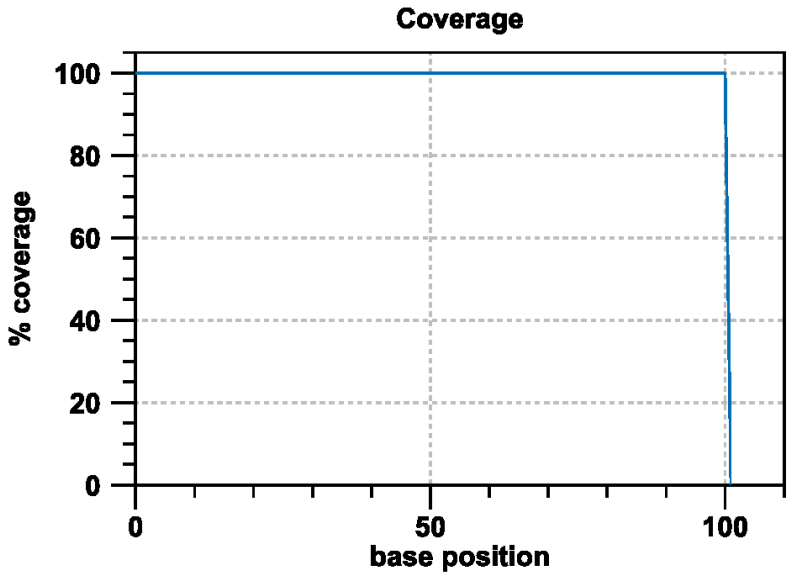
2.4 Quality distribution



Distribution of average sequence qualitie scores. The quality of a sequence is calculated as the arithmetic mean of its base qualities.
x: PHRED-score
y: number of sequences observed at that qual. score normalized to the total number of sequ ences

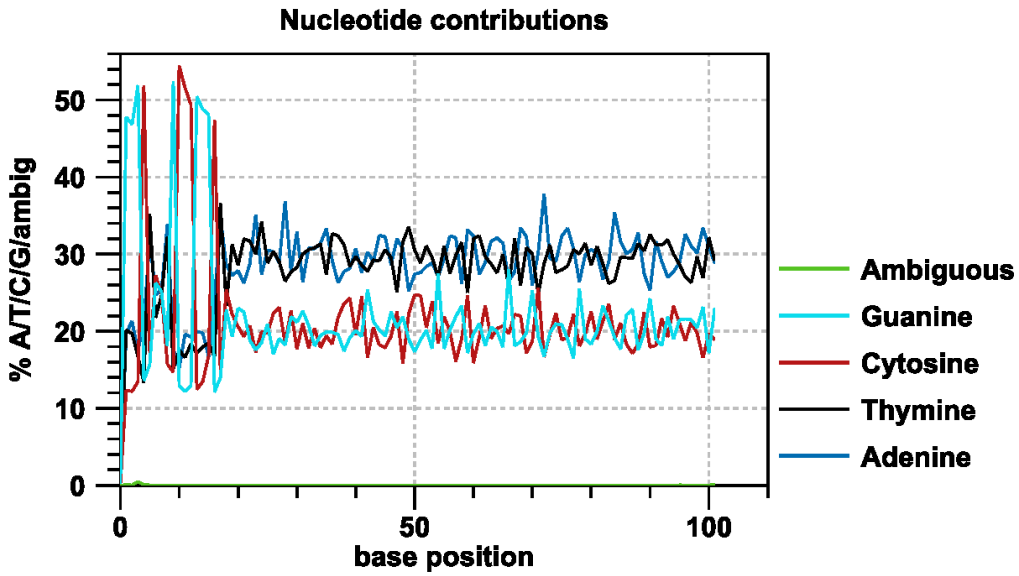
3. Per-base analysis

3.1 Coverage



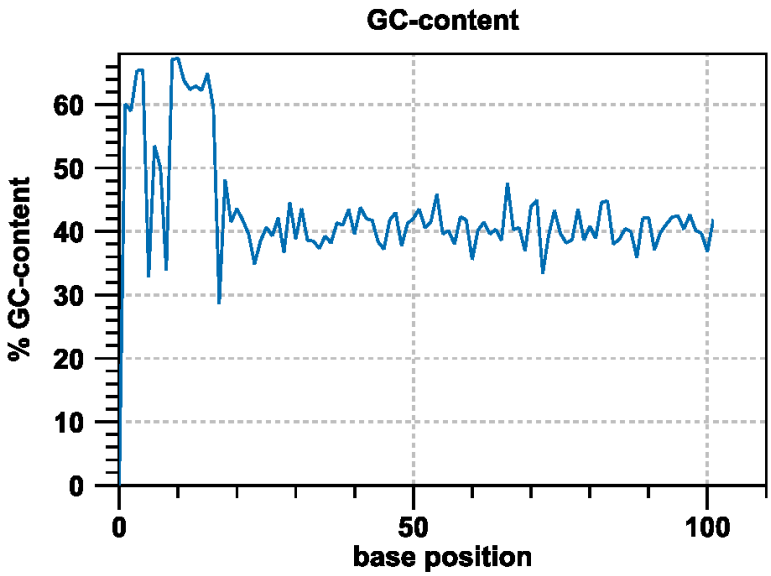
The number of sequences that support (cover) the individual base positions. In cases of un trimmed Illumina or SOLiD reads it will just contain a rectangle.
x: base position
y: number of sequences covering individual base positions normalized to the total number o f sequences

3.2 Nucleotide contributions



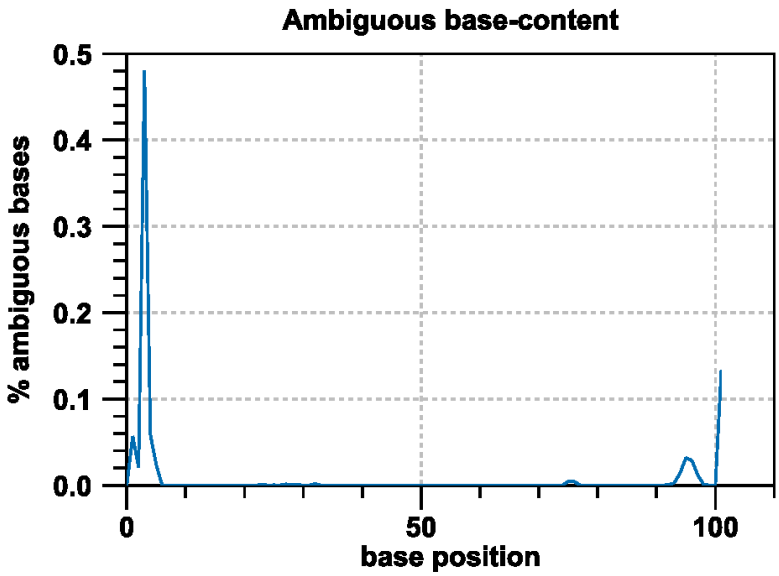
Coverages for the four DNA nucleotides and ambiguous bases.
x: base position
y: number of nucleotides observed per type normalized to the total number of nucleotides o bserved at that position

3.3 GC-content



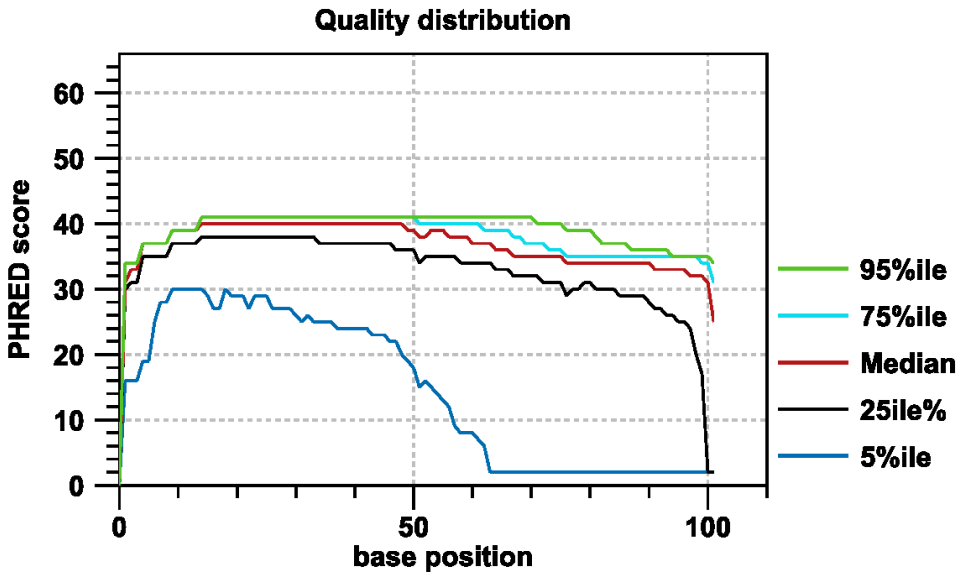
Combined coverage of G- and C-bases.
x: base position
y: number of G- and C-bases observed at current position normalized to the total number of bases observed at that position

3.4 Ambiguous base-content



Combined coverage of ambiguous bases.
x: base position
y: number of ambiguous bases observed at current position normalized to the total number o f bases observed at that position

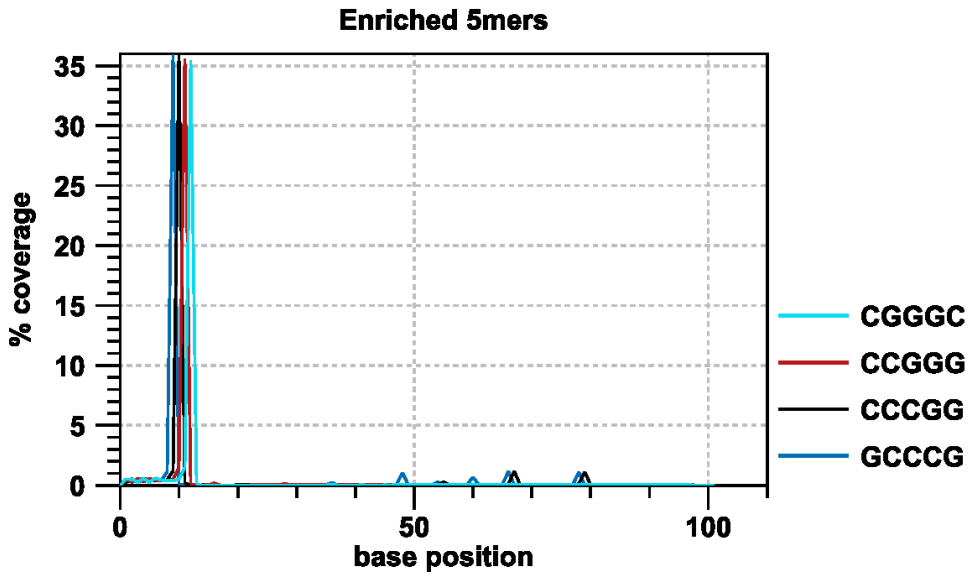
3.5 Quality distribution



Base-quality distribution along the base positions.
x: base position
y: median & percentiles of quality scores observed at that base position

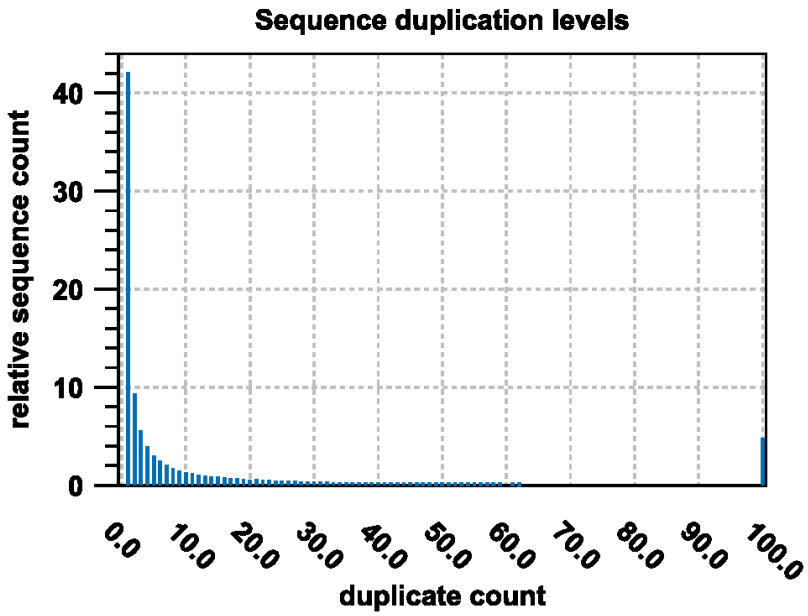
4. Over-representation analyses

4.1 Enriched 5mers



The five most-overrepresented 5mers. The over-representation of a 5mer is calculated as th e ratio of the observed and expected 5mer frequency. The expected frequency is calculated as product of the empirical nucleotide probabilities that make up the 5mer. (5mers that contain ambiguous bases are ignored)
x: base position
y: number of times a 5mer has been observed normalized to all 5mers observed at that posit ion

4.2 Sequence duplication levels



Duplication level distribution. Duplication levels are simply the count of how often a par ticular sequence has been found.
x: duplicate count
y: number of sequences that have been found that many times normalized to the number of un ique sequences

4.3 Duplicated sequences

A table of over-represented sequences is given in the supplementary report