

RESEARCH

Fernanda Antunes Carvalho

Molecular Phylogeny,  
Biogeography and an  
e-Monograph of the Papaya  
Family (Caricaceae) as an  
Example of Taxonomy in  
the Electronic Age



Springer Spektrum

---

# Molecular Phylogeny, Biogeography and an e-Monograph of the Papaya Family (Caricaceae) as an Example of Taxonomy in the Electronic Age

---

Fernanda Antunes Carvalho

Molecular Phylogeny,  
Biogeography and  
an e-Monograph of  
the Papaya Family  
(Caricaceae) as an  
Example of Taxonomy  
in the Electronic Age

With a foreword by Prof. Dr. Susanne S. Renner

 Springer Spektrum

Fernanda Antunes Carvalho  
Munich, Germany

Dissertation Ludwig-Maximilians-Universität München, Germany, 2014

ISBN 978-3-658-10266-1      ISBN 978-3-658-10267-8 (eBook)  
DOI 10.1007/978-3-658-10267-8

Library of Congress Control Number: 2015941425

Springer Spektrum

© Springer Fachmedien Wiesbaden 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer Spektrum is a brand of Springer Fachmedien Wiesbaden  
Springer Fachmedien Wiesbaden is part of Springer Science+Business Media  
([www.springer.com](http://www.springer.com))

*Para Bia, Quincas e Tânia*

## Foreword by the Supervisor

In the flowering plants, there are currently 3–4 synonyms for every accepted name (<http://www.theplantlist.org>). This fact is not a harmless nuisance. Synonymous names cause at least two kinds of errors: they result in wrong assumptions about species' geographic ranges, and they make it difficult to find information about species because users cannot know which name refers to what. The judgment about what is a good biological species and which names are synonyms is made during monographic research. Such research consists in evaluating the information pertaining to all names published for the species, subspecies, or other forms in a genus or family, along with studying all specimens to which these names have been applied (rightly or wrongly). The very best monographs also include geographic distribution maps and DNA sequences from specimens representing the group's species. Based on these data, a monographer reaches a conclusion about which names refer to true biological species. He or she then constructs a key to identify the accepted species and prepares authoritative species descriptions and maps.

**Fernanda Antunes Carvalho** carried out monographic research on the papaya family, Caricaceae, between April 2010 and December 2013. Her dissertation consisted of several chapters, three of which are included in the present book. A novel aspect of Dr. Carvalho's monograph of the Caricaceae was that she availed herself of the cybertaxonomy platform 'Botanical Research and Herbarium Management System' (BRAHMS) to also develop an online version of her monograph that will be updated at certain intervals. In her work, she allocated the about 200 names relevant to the family Caricaceae to 34 biological species and one hybrid. To arrive at this conclusion Dr. Carvalho studied of about 2000 herbarium collections as well as relevant type specimens (specimens on which particular names are based).

Fernanda Antunes Carvalho also used molecular-phylogenetic methods to infer the phylogenetic relationships in the papaya family. Her

phylogenetic work in 2011/2012 led to the discovery that the closest relatives of papaya (*Carica papaya*) all live in Mexico and Guatemala, not in the Andes as often suggested. This discovery fits with other data that likewise indicate that the papaya was domesticated in Mexico/Guatemala.

The book that Fernanda Antunes Carvalho is presenting here brings together all information that is available today about the evolution, distribution, ecology, morphology, and conservation status of the species in the papaya family. Together with the online data and image database that she maintains about this family, it forms the foundation of all current and future research on this family.

Munich, March 2015

Prof. Dr. Susanne S. Renner

## Declaration of Contribution

In this thesis, I present the results of my doctoral research, carried out in Munich from April 2010 to December 2013 under the guidance of Prof. Susanne S. Renner. My thesis resulted in four manuscripts of which two are published (Chaps. III and IV), one is in review (Chap. II) and the fourth (Chap. V) has yet to be submitted. In addition, I worked on an electronic monograph of Caricaceae available at <http://herbaria.plants.ox.ac.uk/bol/caricaceae>, two book chapters and a tutorial, and gave the talks listed below. I generated all data and conducted all analyses myself, except for the karyological analyses (Chap. V), which were done in collaboration with Alexander Rockinger, Martina Silber, and Aretuza Sousa. Writing and discussion of all manuscripts involved collaboration with Prof. Dr. Susanne Renner.

## List of publications

- Carvalho FA, Renner SS (2012)** A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. *Molecular Phylogenetics and Evolution* 65: 46–53
- Carvalho FA (2013 onwards)** e-Monograph of Caricaceae. Vers. 1, Nov. 2013. [Database continuously updated].  
<http://herbaria.plants.ox.ac.uk/bol/caricaceae>
- Carvalho FA, Filer D, Hopkins M (2013)** Using images to enter data in BRAHMS. Tutorial available for download at:  
<http://herbaria.plants.ox.ac.uk/bol/brahms/GroupResources>
- Carvalho FA, Renner SS (2013)** The phylogeny of Caricaceae. In: Ming R, Moore PH (eds) *Genetics and Genomics of Papaya*. Springer, Heidelberg, New York, pp. 81–82
- Carvalho FA, Renner SS (2013)** Correct names for some of the closest relatives of *Carica papaya*: A review of the Mexican/Guatemalan genera *Jarilla* and *Horovitzia*. *PhytoKeys* 29: 63–74
- Carvalho FA, Filer D, Renner SS (2014)** Taxonomy in the electronic age and an e-monograph of the papaya family (Caricaceae) as an example. *Cladistics* (published online 13 Aug. 2014), DOI: 10.1111/cla.12095
- Carvalho FA (2015)** Caricaceae. In: Davidse G, Sousa M, Knapp S, Chiang Cabrera F & Ulloa Ulloa C (eds) *Flora Mesoamericana Vol. 2, Parte 3: Saururaceae a Zygophyllaceae*, Missouri Botanical Garden Press, Monsanto.



## Oral Presentations

**Carvalho FA (2013)** Biogeography, systematic, and cyber-monograph of the papaya family (Caricaceae) *EES<sup>LMU</sup> Conference 2013*, Ludwig Maximilians University, Munich, Germany, October 8–9, 2013 (c. 20 min)

**Carvalho FA (2013)** BRAHMS (Botanical Research and Herbarium Management System). Introduction and hands-on. *Workshop of the Systematics and Biodiversity group and Alexandre Antonelli's research group at the Department of Biological and Environmental Sciences, University of Gothenburg*, Gothenburg, Sweden, August 19–30, 2013 (one full day workshop)

**Carvalho FA, Renner SS (2013)** Climatic niche divergence in old sister lineage splits of Caricaceae, but not young species pairs. *BioSyst.EU 2013 Global systematics!* Vienna, Austria February 18–22, 2013 (c. 20 min)

**Carvalho FA (2012)** Evolution and biogeography of Caricaceae and the closest relatives of papaya. Invited talk in the *Instituto de Biología, Universidad Autónoma de México*, Mexico City, Mexico Aug. 7, 2012 (c. 40 min)

**Carvalho FA, Renner SS (2011)** The papaya tree (*Carica papaya*) belongs in an herbaceous Meso-american clade. *BioSystematics Berlin 2011*. Berlin, Germany, February 21–27, 2011 (c. 15 min)

**Carvalho FA (2010)** Una introducción sobre BRAHMS (Botanical Research and Herbarium Management System), *Instituto de Botánica Agrícola, Universidad Central de Venezuela*, Maracay, Venezuela March 26, 2010 (c. 60 min)

## Herbaria Visited

- BHC, MBM, UCB, R, RB, HB, VEN, MY, MEXU, GUADA, IBUG, K, BM, OXF, GB, W, WU, B, M. More than 1000 additional specimens borrowed (studied and annotated) from HUEFS, F, and NY

## Field Work

- Paraná, Brazil, January 2012; Mérida, Venezuela, March 2012; Jalisco and Oaxaca, Mexico, August 2012

## Funding

- *Conselho Nacional de Pesquisas* (CNPq 290009/2009-0) provided a 4-year scholarship (April 1, 2010 – March 31, 2014)
- *Deutsche Forschungsgemeinschaft* (DFG RE 603/13) funded the project entitled "A cybermonograph and phylogeny of the papaya family, Caricaceae: providing the context for the fully sequenced genome of a worldwide crop"
- *EES<sup>LMU</sup> Travel Grant* provided financial support to present talks at conferences in Berlin and Vienna. I also visited herbaria in these cities
- Ray Ming supported my fieldwork in Mexico in 2012 with part of his grant from the *U.S. National Science Foundation* (NSF) Plant Genome Research Program (Award No. DBI-0922545)

## Summary

This dissertation addresses an issue of key importance to the field of systematics, namely how to foster taxonomic work and the dissemination of knowledge about species by taking full advantage of electronic data and bioinformatic tools. I tested and applied modern systematic tools to produce an electronic monograph of a family of flowering plants, Caricaceae. In addition to a taxonomic revision, a molecular phylogeny of the family that includes representatives of all biological species clarifies the evolutionary relationships. Based on the plastid and nuclear DNA data, I inferred historical processes that may have shaped the evolution of the Caricaceae and explain their current geographic distribution.

The first part of my thesis focuses on the development of an electronic monograph using existing infrastructures of Information Technology (IT) and bioinformatic tools that together set the stage for a new era of systematics. I address the problem of synonyms and the importance of taxonomic monographs as the portal for the entire information available about species, including all names published since 1753. Using relatively cheap gadgets (a small digital camera and a portable digital microscope), I rather efficiently gathered data from herbarium specimens and organized these data in a dynamically updated electronic monograph of Caricaceae, using the Botanical Research and Herbarium Management System (BRAHMS) developed at the University of Oxford. The e-monograph includes distribution maps (based on 2201 georeferenced collections), photos of 3943 herbarium specimens (and weblinks to high resolution images of type specimens), highly detailed plates illustrating all species, as well as comprehensive data on morphology, chromosome numbers, phenology, uses, and habitat. I revised all extant 233 names, solving nomenclatural and typification problems, and built multi-access identification keys for all species and genera using Xper2, developed at the Université Pierre et Marie Curie in Paris.

The second part of my thesis focuses on the phylogeny and biogeography of Caricaceae. I produced the first complete DNA-based phylogeny of the family including all genera and accepted species and discovered that the closest relatives of papaya are four species endemic to Mexico, Guatemala, and El Salvador. Together with the current distribution of the wild form of papaya (which has smaller and harder fruits than the cultivated form), the phylogeny supports the idea that papaya originated in Central America and was domesticated by a Mesoamerican civilization. The historical biogeography of Caricaceae involved a long-distance dispersal event from Africa to the Neotropics during the Late Eocene. The deepest divergence in the Neotropics dates to the Oligocene-Miocene boundary and involves a split between a Central American and a (mostly) South American clade, suggesting range expansion across the Panamanian Isthmus. In the New World, diversification during the Miocene seems to be related to the main events of mountain building that formed new habitats and barriers, and to the climate cooling responsible by the expansion of dry habitats. The Pleistocene major climate change in Africa parsimoniously relates to the inferred divergence time of ancient West and East African populations.

The last part is dedicated to the evolution of chromosome numbers in the Caricaceae and includes counts for species from three genera (*Cylicomorpha*, *Horovitzia*, *Jarilla*) that have never been investigated before. Before my study, all published counts for Caricaceae were  $2n = 18$ , but preliminary results show that *Horovitzia cnidoscoloides* presents  $2n = 16$ , and two species of *Jarilla* (*J. caudata* and *J. heterophylla*) present  $2n = 14$ , indicating that chromosomal rearrangements resulting in the reduction of the chromosomes number may have occurred in the most recent common ancestor of this small clade.

# Contents

I. General Introduction	1
The Problem of Synonyms and the Importance of Taxonomic Monographs	3
Brassicales and the Caricaceae	5
Distribution and Diversity of Caricaceae	6
Morphology, Pollination, Sexual Systems, and Chromosomes	6
Taxonomic History and Previous Molecular Studies on Caricaceae	9
The Main Geological and Climate Events Related to Caricaceae Biogeography	10
Research Questions and Aims	12
II. Taxonomy in the Electronic Age: An e-Monograph of the Papaya Family (Caricaceae) as an Example	13
III. Correct Names for Some of the Closest Relatives of <i>Carica papaya</i> L.: A Review of the Mexican/Guatemalan Genera <i>Jarilla</i> and <i>Horovitzia</i>	33
IV. A Dated Phylogeny of the Papaya Family (Caricaceae) Reveals the Crop's Closest Relatives and the Family's Biogeographic History	49
V. Chromosome Counts for the Caricaceae Reveal Unexpected Dysploidy	83
VI. General Discussion	93
Taxonomy in the Electronic Age and an e-Monograph of Caricaceae	93
Phylogenetic Relationships Within Caricaceae	96
Historical Biogeography and the Importance of Habitat Diversity in the Evolution of Caricaceae	98
The African Clade ( <i>Cylicomorpha</i> )	98
The Central American Clade	99
The Mostly South American Clade <i>Vasconcellea/Jacaratia</i>	100
Evolution of Chromosome Numbers in the Caricaceae	105

General Conclusions ..... 107  
References ..... 109  
Appendix ..... 117  
Acknowledgments ..... 138  
Curriculum Vitae ..... 141

# I. General Introduction

Biological systematics aims to document and understand the history and diversity of life on Earth. Among other steps, this requires naming biological entities, which involves description, classification, and applying the rules of nomenclature. Before the molecular age, which began in the 1950s, systematics was based mostly on morphology. Today, however, much of systematics focuses on the phylogenetic relationships among species and higher clades, based on DNA sequences. While such “molecular systematics” has revolutionized our understanding of the evolution of organisms, it does not provide all the information required to name and describe the World’s species in a recognizable and universal manner.

Concern about the loss of biodiversity is widespread around the Globe and affects the citizens of all countries, whether on Pacific Islands threatened by raising sea levels or in densely populated industrial regions of Europe. Understanding the extent and causes of biodiversity loss requires, among other fields of science, also systematic research. This is because biodiversity research provides the basis for proper species identification and the permanent preservation and documentation in public collections of examples of as many species and forms as possible. Efficient species identification is hampered by overly technical literature that is not updated and, often, expensive and inaccessible. Traditional taxonomic literature also usually fails to take full advantage of modern tools, such as electronic color images, much less the millions of named and unnamed images already available online and accessible for image recognition software (Kress 2004; Belhumeur et al. 2008; Shamir et al. 2010). Additionally, the high number of synonymous names (see below) contributes to the problem of finding the right name for an organism. In order to identify species properly and efficiently one needs identification keys, detailed species descriptions, and precise distribution maps. Most important is that this information must be accessible, easy to understand and be associated with illustrations. Identifying species correctly is pivotal for exploring chemical and economic properties of wild organisms, prioritizing areas for conservation, and assessing extinction risks. It can also help making the public aware of biodiversity and thereby help conservation.

To produce a meaningful picture of life on Earth it is necessary to bring together the knowledge available on each species (e.g., specimen-based geographic occurrence, morphological descriptions, revised nomenclature, chromosome numbers, reproductive biology, DNA sequences, etc.) and combine it with other types of information, such as climate and geological data. Today's Information Technology (IT)-infrastructure and bioinformatic tools set the stage for a new era of systematics in which the burden of taxonomic work is alleviated by ready access to public repositories with images of specimens, including types, and the literature of the past 260 years (back to Linneaus, 1753, the starting point of botanical nomenclature, and even older literature). Other molecular, geographical, ecological, and physical (geology and climate) data – often freely available – are also increasing exponentially.

For my doctoral research, I decided to use bioinformatic tools to produce an electronic monograph on a family of flowering plants, bringing together all available information and generating new data, including DNA sequences, chromosomes number, a completely revised nomenclature, identification keys, and well-illustrated morphological descriptions. I decided to focus on a suitably-sized family of flowering plants, the Caricaceae, which includes the economically important crop, *Carica papaya*. Caricaceae are well suited for applying and testing modern systematic tools because of (i) their economic importance (almost all species are used at least locally), which has resulted in numerous studies on chemistry, pharma-ceutical properties, and genomics that meant my study would be broadly useful, (ii) their relatively few species (34) but numerous available names (233), which has to do with the economic importance of papaya and the family's distribution in the humid tropics (still under-collected, with the scarcity of specimens [in all reproductive stages] contributing to unclear species boundaries), (iii) their relatively high number of red-listed species (six species listed in the IUCN, 2013), which meant it was important to better document species' geographic ranges, and (iv) the family not having been the focus of a recent monograph. The family's geographic occurrence in Africa, Central America, and South America also made it biogeographically interesting.

In the first part of this General Introduction, I expand on the importance of using bioinformatic tools to make the taxonomic effort more efficient and accessible to different communities of people. In the second part, I summarize basic knowledge on the Caricaceae and clarify the questions about the family's phylogeny, biogeography, and chromosome evolution that motivated my research.

## **The Problem of Synonyms and the Importance of Taxonomic Monographs**

So far, there are c. 1.9 million accepted (named) species on Earth, from which only 66,307 represent microbial diversity (Chapman 2009). For flowering plants, there are 1,040,426 scientific names, 298,900 of them accepted, 263,925 unclear, and 477,601 (45.9%) synonyms (The Plant List 2010). A review of the problem of synonymous names shows that in plant groups that have been monographed, 58 to 78% of the published names turned out to be synonyms (Scotland and Wortley 2003). Such levels of synonymy are a serious problem. Moreover, the rate at which new synonyms are produced seems to be increasing linearly with the rate at which new species are being described (Wortley & Scotland 2004; Fig. 1). The high levels of synonymy and the scarcity of taxonomic treatments of larger groups are major impediments to the recognition of "good" species (because the increasing numbers of synonyms make it ever harder to study and sort the type material). This contributes to the relatively slow rate at which new species are being recognized as such and then described. Thousands of already collected new species await discovery in scientific collections (Bebber et al. 2010).

Besides slowing down systematic and evolutionary research, synonymous names also hamper the prediction of extinction rates, which requires the knowledge of how many species there are and what their range sizes are (Pimm et al. 1995). Synonyms usually result in too small species ranges and thus perhaps exaggerated estimates of endangered species because each name will be associated with its own "species" range. Lastly, synonymous names make it difficult to find published information on a particular biological entity, hampering the