

Andrew H. Paterson *Editor*

Genomics of the Saccharinae

Plant Genetics and Genomics: Crops and Models

Volume 11

Series Editor

Richard A. Jorgensen

For further volumes:

<http://www.springer.com/series/7397>

Andrew H. Paterson
Editor

Genomics of the Saccharinae

 Springer

Editor

Andrew H. Paterson, Ph.D.
Plant Genome Mapping Laboratory
and Departments of Crop and Soil Science
Plant Biology, and Genetics
University of Georgia
Athens, GA, USA

ISBN 978-1-4419-5946-1 ISBN 978-1-4419-5947-8 (eBook)
DOI 10.1007/978-1-4419-5947-8
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012942104

© Springer Science+Business Media New York 2013

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

This volume is dedicated to the memories of James E. Irvine and Keith F. Schertz, Saccharinae scientists who inspired and mentored the editor, many of the chapter authors, and many others.

Preface

Among flowering plants, several “warm-season” grasses are the most efficient at fixing atmospheric carbon, thanks to “C4” photosynthesis, a complex combination of biochemical and morphological specializations discovered in sugarcane that increase net carbon assimilation at high temperature. The Saccharinae clade of grasses is of singular importance, including one cereal that is fifth in importance among the world’s grain crops, as well as three leading biofuel crops, and several of the world’s most noxious weeds.

Sorghum bicolor L. Moench. is a leading cereal, fifth in importance among the world’s grain crops. Introduced into the USA about 200 years ago, sorghum is grown on 8–10 million acres and has a farm-gate value of ~\$1 billion/yr. *Sorghum* is unusually tolerant of drought, a feature essential in the US Southern Plains that often receive too little rain for other grains. In arid countries of northeast Africa, sorghum contributes 26–39% of calories in the human diet. Increased demand for limited fresh water, along with rising global temperature and aridity, suggest that sorghum will be of growing importance.

Expansion of agriculture to provide plant biomass for production of fuels and/or feedstocks will require additions to our present repertoire of crops. The Saccharinae clade of grasses shows singular promise, including three leading biofuel crops, *Saccharum* (sugarcane, the world’s #1 fuel ethanol crop), *Sorghum* (currently the #2 source of seed-based fuel ethanol in the USA, and a promising potential source of cellulosic ethanol), and *Miscanthus*, a promising potential cellulosic ethanol crop with much higher yield than another leading candidate, switchgrass, in the US Midwest. Its adaptability to continental Europe shows the feasibility of producing *Miscanthus* in temperate latitudes.

The *Sorghum* genus also includes one of the world’s worst weeds “Johnsongrass” (*S. halepense*), a naturally occurring polyploid hybrid that reduces yields of many crops by up to 45 %. The first federal appropriation for weed control research targeted Johnsongrass. Functional genomic data may lead to new strategies for environmentally benign plant growth regulation, suppressing weed dispersal. Better understanding of reproductive barriers in sorghum may lead to strategies to reduce risk that transgenic *S. bicolor* outcrosses with *S. halepense*.

An important breeding line of *Sorghum bicolor* recently became only the second monocot to have its genome essentially fully sequenced, providing an important complement to the previously sequenced genome of rice and opening new doors into the study and improvement of members of the clade. As a model for the tropical grasses, sorghum is a logical complement to *Oryza* (rice). Sorghum is representative of warm-season grasses in that it has “C4” photosynthesis, while rice is more representative of temperate grasses, using “C3” photosynthesis. The ~740 megabase sorghum genome, with ~90 % of DNA and ~98 % of genes placed in their chromosomal context, is a logical bridge to the ~2,500 megabase maize genome that is also being sequenced, and the ~4,000 megabase genome of sugarcane, the world’s leading biomass/biofuels crop. Sorghum shared common ancestry with maize (12–15 million years ago, mya) and sugarcane (5–9 mya) much more recently than rice (42–47 mya). The most recent genome duplication in sorghum appears to be ~70 mya versus ~12 mya in maize and <5–9 mya in sugarcane with lower genetic redundancy promising a higher success rate in relating sorghum genes to phenotypes.

For a multitude of reasons—invigorated interest in biofuels, concerns about a looming worldwide water crisis, the need for more precise and more environmentally benign methods of weed control—the Saccharinae clade has seen a resurgence of interest in the past few years. The Saccharinae have an important role to play in a more bio-based economy and a more sustainable agroecosystem. Sequencing of additional members of the clade has begun, building on their rich histories of conventional breeding and genetics research, but constrained by the challenges of their large and complex genomes. In this book, we seek to share with you, the reader, our enthusiasm about the advances in genetics and genomics of the Saccharinae of the past few years and those that loom on the horizon.

In closing, a clarification of nomenclature is important. As described in detail in Chap. 1, the taxonomic nomenclature of the species that are the focus of this book remains unclear. For the purposes of this book, the authors have been encouraged to adopt the view expressed and explained in Chap. 1, that it appears appropriate that subtribe Sorghine (presently including sorghum) should be merged into subtribe Saccharinae (including *Saccharum* and *Miscanthus*). Accordingly, the chapter authors have been encouraged to refer to the Saccharinae as inclusive of all three taxa.

Athens, GA, USA

Andrew H. Paterson, Ph.D.

Contents

Part I Natural History and Genetic Diversity

- 1 Phylogenetic Relationships of Saccharinae and Sorghinae.....** 3
Elizabeth A. Kellogg
- 2 The Gene Pool of *Sorghum bicolor* and Its Improvement** 23
Clarissa T. Kimber, Jeff A. Dahlberg, and Stephen Kresovich
- 3 The Gene Pool of *Saccharum* Species
and Their Improvement** 43
Andrew H. Paterson, Paul H. Moore, and Tom L. Tew
- 4 The Gene Pool of *Miscanthus* Species
and Its Improvement** 73
Erik J. Sacks, John A. Juvik, Qi Lin, J. Ryan Stewart,
and Toshihiko Yamada

Part II Genomic Tools, Resources and Approaches

- 5 The Sorghum Genome Sequence: A Core Resource
for Saccharinae Genomics.....** 105
Andrew H. Paterson
- 6 Transcriptome Analysis in the Saccharinae** 121
Milton Yutaka Nishiyama-Jr, Fabio Vicente,
Paloma Mieke Sato, Savio Siqueira Ferreira,
Frank Alex Feltus, and Glaucia Mendes Souza
- 7 Sorghum and Sugarcane Proteomics** 141
Bongani Kaiser Ndimba and Rudo Ngara

8 Gene Mutagenesis Systems and Resources for the Saccharinae	169
Zhanguo Xin, Ming-Li Wang, Surinder Chopra, and Pohao Wang	
9 Association Genetics Strategies and Resources	187
Jianming Yu, Martha T. Hamblin, and Mitchell R. Tuinstra	
10 Sorghum Transformation: Overview and Utility	205
Tejinder Kumar, Arlene Howe, Shirley Sato, Ismail Dweikat, and Tom Clemente	
11 Genetic Engineering of <i>Saccharum</i>	223
Getu Beyene, Ian S. Curtis, Mona B. Damaj, Marco T. Buenrostro-Nava, and T. Erik Mirkov	
12 Genetic Engineering of <i>Miscanthus</i>	255
Dean Engler and Katrin Jakob	
13 <i>Saccharinae</i> Bioinformatics Resources	303
Alan R. Gingle and F. Alex Feltus	
 Part III Bridging Classical and Genomic Investigations of Sorghum Biology	
14 Bridging Classical and Molecular Genetics of Sorghum Plant Stature and Maturity	333
Patrick J. Brown and Andrew H. Paterson	
15 Bridging Classical and Molecular Genetics of Sorghum Disease Resistance	347
Clint W. Magill	
16 Bridging Conventional and Molecular Genetics of Sorghum Insect Resistance	367
Yinghua Huang, Hari C. Sharma, and Mukesh K. Dhillon	
17 Genetic Enhancement of Sorghum for Biomass Utilization	391
Wilfred Vermerris and Ana Saballos	
 Part IV Early Messages pertinent to Other Crops, from Saccharinae Research	
18 Comparative Genomics of Grasses: A Saccharinae-Centric View	429
Andrew H. Paterson, Xiyin Wang, Haibao Tang, and Changsoo Kim	

19 Comparative Genomic Analysis of C4 Photosynthesis Pathway Evolution in Grasses 447
 Xiyin Wang and Andrew H. Paterson

20 Differentiation of Seed, Sugar, and Biomass-Producing Genotypes in *Saccharinae* Species 479
 Seth C. Murray

21 Perennialism and Weediness in the *Saccharinae*..... 503
 Russell W. Jessup

Part V Synthetic and Futuristic Perspectives

22 Bringing the Benefits of Sorghum Genomics to Africa 519
 Segenet Kelemu, Brhane Gebrekidan, and Jagger Harvey

23 Synthesis: Fundamental Insights and Practical Applications from the *Saccharinae* Clade 541
 Andrew H. Paterson

Index..... 551

Contributors

Getu Beyene Texas AgriLife Research, Texas A&M System, College Station, Weslaco, TX, USA

International Laboratory for Tropical Agricultural Biotechnology (ILTAB), Donald Danforth Plant Science Center, 975 North Warson Road, St. Louis, MO, USA

Patrick J. Brown Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Marco T. Buenrostro-Nava Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Surinder Chopra Crop and Soil Science Department, Pennsylvania State University, University Park, PA, USA

Tom Clemente Department of Agronomy & Horticulture, Center for Plant Science Innovation, Center for Biotechnology, University of Nebraska–Lincoln, Lincoln, NE, USA

Ian S. Curtis Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Jeff A. Dahlberg U.S. Sorghum Checkoff Program, Lubbock, TX, USA

Mona B. Damaj Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Mukesh K. Dhillon Division of Entomology, Indian Agricultural Research Institute (IARI), New Delhi, India

Ismail Dweikat Department of Agronomy & Horticulture, University of Nebraska–Lincoln, Lincoln, NE, USA

Dean Engler Mendel Biotechnology, Inc, Hayward, CA, USA

F. Alex Feltus Department of Genetics and Biochemistry, Clemson University, Clemson, SC, USA

Savio Siqueira Ferreira Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Brhane Gebrekidan Agriculture Working Group, Ethiopian Academy of Sciences, Addis Ababa, Ethiopia

Alan R. Gingle Plant Genome Mapping Laboratory, University of Georgia, Athens, GA, USA

Martha T. Hamblin Institute for Genomic Diversity, Cornell University, Ithaca, NY, USA

Jagger Harvey Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya

Arlene Howe Center for Biotechnology, University of Nebraska–Lincoln, Lincoln, NE, USA

Monsanto Company, Chesterfield, MO, USA

Yinghua Huang Plant Science Research Laboratory, USDA–ARS, Stillwater, OK, USA

Katrin Jakob Mendel Biotechnology, Inc, Hayward, CA, USA

Russell W. Jessup Department of Soil & Crop Sciences, Texas A&M University, College Station, TX, USA

John A. Juvik Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Segenet Kelemu Biosciences eastern and central Africa-International Livestock Research Institute (BecA-ILRI) Hub, Nairobi, Kenya

Elizabeth A. Kellogg Department of Biology, University of Missouri-St. Louis, St. Louis, MO, USA

Changsoo Kim Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Clarissa T. Kimber Department of Geography, Texas A&M University, College Station, TX, USA

Stephen Kresovich Department of Biology, University of South Carolina, Columbia, SC, USA

Tejinder Kumar Department of Agronomy & Horticulture, University of Nebraska–Lincoln, Lincoln, NE, USA

Qi Lin Institute of Botany, Chinese Academy of Sciences, Beijing, China

Clint W. Magill Department of Plant Pathology and Microbiology, Texas A&M University, College Station, TX, USA

T. Erik Mirkov Texas AgriLife Research, Texas A&M System, College Station, TX, USA

Paul H. Moore Hawaii Agriculture Research Center, Kunia, HI, USA

Seth C. Murray Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA

Bongani Kaiser Ndimba Proteomics Research Group, Department of Biotechnology, University of the Western Cape, Bellville, South Africa
Proteomics Unit, Biotechnology Platform, Agricultural Research Council, Stellenbosch, South Africa

Rudo Ngara Proteomics Research Group, Department of Biotechnology, University of the Western Cape, Bellville, South Africa

Milton Yutaka Nishiyama-Jr Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Andrew H. Paterson Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Ana Saballos Genetics Institute and Agronomy Department, University of Florida, Gainesville, FL, USA

Erik J. Sacks Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Paloma Mieko Sato Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Shirley Sato Center for Biotechnology, University of Nebraska-Lincoln, Lincoln, NE, USA

Hari C. Sharma International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, Andhra Pradesh, India

Glauca Mendes Souza Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

J. Ryan Stewart Department of Crop Sciences, University of Illinois, Urbana, IL, USA

Haibao Tang Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Tom L. Tew USDA, ARS, Houma, LA, USA

Mitchell R. Tuinstra Department of Agronomy, Purdue University, Richmond, IN, USA

Wilfred Vermerris Genetics Institute and Agronomy Department, University of Florida, Gainesville, FL, USA

Department of Agricultural & Biological Engineering and Laboratory of Renewable Resources Engineering, Purdue University, West Lafayette, IN, USA

Fabio Vicente Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

Ming-Li Wang PGRCU, USDA-ARS, Griffin, GA, USA

Pohao Wang Crop and Soil Science Department, Pennsylvania State University, University Park, PA, USA

Xiyin Wang Plant Genome Mapping Laboratory and Departments of Crop and Soil Science, Plant Biology, and Genetics, University of Georgia, Athens, GA, USA

Center for Genomics and Computational Biology, School of Sciences, Hebei Polytechnic University, Tangshan, Hebei, China

Zhanguo Xin Plant Stress and Germplasm Development Unit, USDA-ARS, Lubbock, TX, USA

Toshihiko Yamada Field Science Center for Northern Biosphere, Hokkaido University, Hokkaidō, Japan

Jianming Yu Department of Agronomy, Kansas State University, Manhattan, KS, USA

Part I
Natural History and Genetic Diversity

Chapter 1

Phylogenetic Relationships of Saccharinae and Sorghinae

Elizabeth A. Kellogg

Abstract Multiple taxonomic and phylogenetic studies have been conducted on sugarcane, *Miscanthus*, and sorghum, but to date the results have been contradictory and somewhat confusing. A few generalities have emerged. The Andropogoneae is clearly monophyletic. *Saccharum* and *Miscanthus* are closely related to each other. Their relationship with *Sorghum* is less clear, although they are probably more closely related to *Sorghum* than any of them is to maize or to *Andropogon* and its immediate relatives. The phylogeny of Andropogoneae is largely unresolved, which leads to a number of problems of taxonomic nomenclature. The solution will require considerably more phylogenetic data on a much broader set of species than has been sampled to date.

Keywords Evolution • Phylogeny • Andropogoneae • *Miscanthus* • Polyploid • Classification • Sorghum • Sarga

1 Introduction

Sugarcane, *Miscanthus*, and sorghum are all members of the grass family, Poaceae. The first two are currently placed in the subtribe Saccharinae, whereas sorghum is often given its own subtribe, Sorghinae; both subtribes belong to the tribe Andropogoneae. Understanding the relationships of the three taxa is complicated by considerable phylogenetic uncertainty and taxonomic confusion. As described below, the phylogeny represents a difficult phylogenetic problem, in which rapid speciation early in the history of the group led to a phylogenetic tree with many very short internal branches. Disentangling this history has been complicated by a

E.A. Kellogg (✉)

Department of Biology, University of Missouri, One University Blvd, St. Louis, MO 63121, USA
e-mail: tkellogg@umsl.edu

limited sample of species in all studies to date. In addition, the group includes many polyploids, sugarcane and *Miscanthus* among them, whose evolution is likely to be highly reticulate; while molecular phylogenetic approaches can disentangle net-like histories, this has not yet been attempted in the Andropogoneae.

In this chapter, I take a hierarchical approach, beginning with the placement of Saccharinae and Sorghinae in the angiosperms, and noting the characteristics that have originated at various stages in evolution. I then consider the placement of the groups within the phylogeny of Andropogoneae, and discuss some of the implications and limitations of the phylogeny for understanding morphological and chromosomal evolution. In the final section, I consider the implications of the phylogenetic data for classification. Although I may not eliminate any taxonomic confusion, I hope to explain at least why it has occurred and what information might address the problems in the future. One likely conclusion from the available data is that the subtribe Saccharinae will ultimately be merged with the subtribe Sorghinae, and that the former name will take precedence.

2 Placement of Sugarcane, *Miscanthus*, and Sorghum within the Angiosperms

The grasses are angiosperms, and are members of the large monocot clade, which includes about 20 % of known flowering plants. Within the monocots, the grasses belong to the commelinid clade (Fig. 1.1). Thus sugarcane, *Miscanthus*, and sorghum all inherit the molecular and morphological characteristics of each of the larger groups to which they belong (summarized in Stevens 2008). As angiosperms, they have ovules enclosed in ovaries, double fertilization, and the many familiar angiosperm characteristics. As monocots they have a single cotyledon, and vascular bundles scattered in the stem. As commelinid monocots, they have cell walls that contain ferulic acid such that they fluoresce under ultraviolet light. Most commelinid monocots, including sugarcane, *Miscanthus*, sorghum, and the grasses, also produce silica bodies (SiO_2) in their leaves and have bracteate inflorescences. The stomata of the commelinids are characteristically paracytic, meaning that they have subsidiary cells that are parallel to the long axis of the stomate. (Some species have tetracytic stomata, which have two additional subsidiaries, one at either end with the long axis of the cell perpendicular to the opening of the stomate.) The commelinids, except for the palms, accumulate starch in the endosperm; the endosperm that is the source of much human nutrition is thus an ancient characteristic that has been retained in the grasses.

Within the commelinids, sugarcane, *Miscanthus* and sorghum are members of the order Poales (Fig. 1.1). This is a monophyletic group of 17 families, all of which have silica bodies in their epidermis. All members of this group also are characterized by having nuclear endosperm, in which multiple nuclear divisions occur before cell walls are formed. A clade within the Poales includes the grass family (Poaceae), plus the families Anarthriaceae, Centrolepidaceae, Restionaceae, Flagellariaceae, Ecdeiocoleaceae, and Joinvilleaceae; together these are known as the graminoid Poales (Campbell and Kellogg 1987; Kellogg and Linder 1995) (Fig. 1.2). As members of

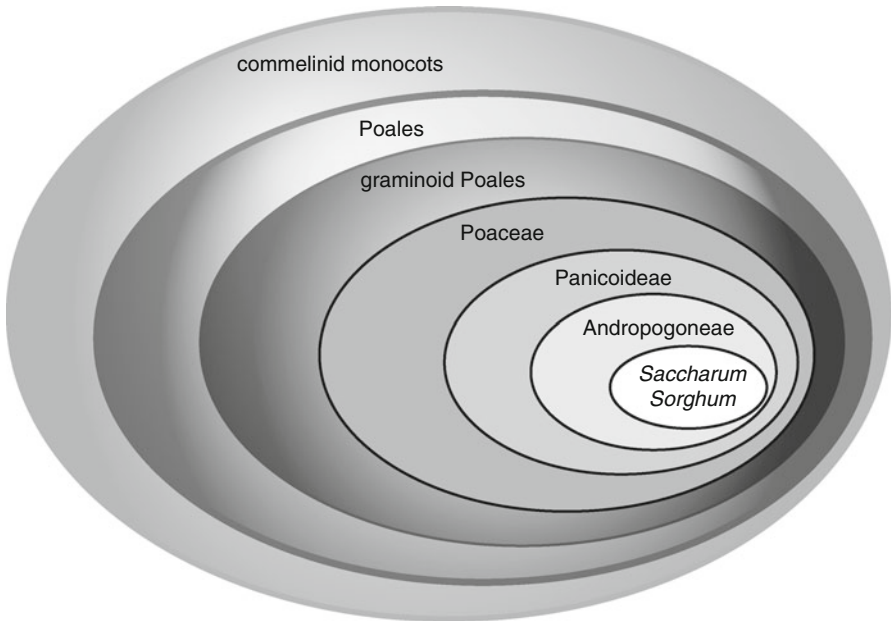


Fig. 1.1 Venn diagram showing the nested relationships of the commelinid clade, Poales, and included taxa

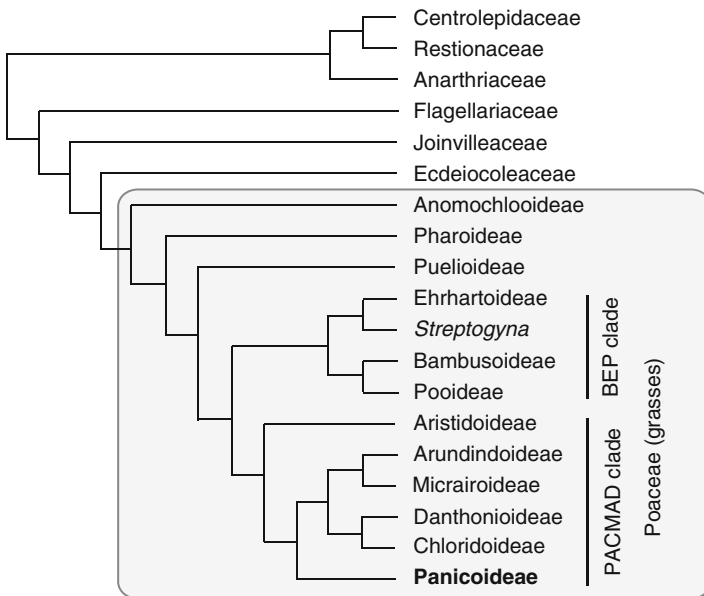


Fig. 1.2 Phylogeny of the graminoid Poales, including relationships among the subfamilies of grasses, following Christin et al. (2008)

the graminoid Poales, the Saccharinae and Sorghinae have distichous leaves, primary cell walls with (1-3,1-4)- β -D-glucans, and sieve tube plastids with cuneate crystals; the pollen has a single pore (monoporate) surrounded by a raised ring (annulus). All graminoid Poales have one anatropous ovule per carpel; the number of locules is often reduced to one. Most members of the clade are wind pollinated, and thus have tiny flowers with few or no tepals, and feathery stigmas.

As members of the family Poaceae, sugarcane, *Miscanthus*, and sorghum have a grass embryo, which is far more differentiated than embryos in other monocots, and is the result of a heterochronic shift in the timing of embryo maturation relative to the maturation of the seed (Kellogg 2000b). The embryo has a clear apical meristem and several seedling leaves, as well as a unique haustorial organ, the scutellum, which is thought to be a highly modified cotyledon. Like almost all grasses, sugarcane, *Miscanthus*, and sorghum have flowers in spikelets, clusters of one or more flowers with the whole cluster subtended by two bracts, the glumes.

Characteristics of morphology plus extensive DNA data place sugarcane, *Miscanthus*, and sorghum in the subfamily Panicoideae, within the grass family (Fig. 1.2). This subfamily is one of the most distinctive groups within the grasses, having been recognized initially by Robert Brown (1810, 1814) based on its two-flowered spikelets, with the upper flower being bisexual and the lower flower staminate or sterile.

3 The Tribe Andropogoneae

3.1 Molecular Phylogenetics

Sugarcane, *Miscanthus*, and sorghum belong to the tribe Andropogoneae, a group that includes 85–90 genera (Clayton and Renvoize 1986). All molecular phylogenetic studies of Andropogoneae (Bomblies and Doebley 2005; Chen et al. 2009; Hodkinson et al. 2002a; Kellogg 2000a; Lukens and Doebley 2001; Mathews et al. 2002; Skendzic et al. 2007; Spangler et al. 1999) show that it is monophyletic and that the genus *Arundinella* (formerly the type genus of the tribe Arundinelleae) is sister to all Andropogoneae. The common ancestor of *Arundinella* plus all other Andropogoneae is estimated to have lived about 19 million years (My) ago (19.1 ± 4.5 ; Vicentini et al. 2008).

All studies also find good support for linking *Zea* plus *Tripsacum*, and for a clade composed of *Bothriochloa*, *Dichanthium*, and *Capillipedium*; both these groups are expected based on considerable previous work (e.g., deWet and Harlan 1974; Harlan and deWet 1963; Hitchcock 1950; Mangelsdorf and Reeves 1931). Other large groups that were identified in one or more studies include (1) the awned Andropogoneae (Bomblies and Doebley 2005; Kellogg 2000a; Lukens and Doebley 2001; Mathews et al. 2002), dated by Vicentini et al. (2008) at 11 My (11.4 ± 3.1); (2) “core Andropogoneae,” including *Andropogon*, *Schizachyrium*, *Hyparrhenia*,

Cymbopogon, and *Heteropogon*, plus the clade of *Bothriochloa*, *Dichanthium*, and *Capillipedium* (Bomblies and Doebley 2005; Kellogg 2000a; Lukens and Doebley 2001; Mathews et al. 2002; Skendzic et al. 2007; Vicentini et al. 2008)(9.1 ± 2.7 My); (3) *Saccharum* plus *Miscanthus* (Hodkinson et al. 2002a; Mathews et al. 2002; Skendzic et al. 2007); (4) African sorghum (*S. bicolor*, *S. arundinaceum*, *S. halepense*, *S. propinquum*) (Chen et al. 2009; Hodkinson et al. 2002a; Skendzic et al. 2007).

Comparing the many studies is difficult because each includes a somewhat different set of taxa and a different set of DNA markers. Spangler et al. (1999) used the chloroplast gene *ndhF*, Kellogg (2000a) used *ndhF*, granule bound starch synthase 1 (GBSS1 or *waxy*) and morphology, Lukens and Doebley (2001) used *teosinte branched 1 (tb1)*, Mathews et al. (2002) combined *ndhF*, *waxy*, and phytochrome B (*phyB*), Bomblies and Doebley (2005) used *Leafy (lfy)*, Hodkinson et al. (2002a) and Skendzic et al. (2007) used the Internal Transcribed Spacer of the nuclear ribosomal RNA genes (ITS) plus sequences of the intron and spacer of the chloroplast *trnL-F*, Chen et al. (2009) used ITS, and Vicentini et al. (2008) used *ndhF* and *phyB*. Spangler et al. (1999) included a particularly large sample of the Australian species of *Sorghum*, Bomblies and Doebley (2005) focused on *Zea* and *Tripsacum*, Hodkinson et al. (2002a) included many species of *Saccharum* and *Miscanthus*, and Chen et al. (2009) focused on *Microstegium*. In these studies, other members of Andropogoneae were included simply as placeholders. Plant material and DNA was shared among the authors of several studies, so that studies from different labs used some of the same plant accessions; for example, all sequences of *Capillipedium parviflorum* were produced from a single plant. This is useful, in that any conflict between gene trees cannot be due to misidentification or confusion of specimens. On the other hand, it means that in some cases our view of the history of an entire species is determined by the DNA sequences of a single plant.

More problematical for comparison among studies is the lack of resolution of the trees. No study—even Mathews et al. (2002), which used more base pairs of DNA than any of the others—was able to resolve the early radiation of the Andropogoneae. Few mutations were found to link any of the clades, and many genera remain unplaced relative to each other. This sort of phylogeny, with short internal branches (few mutations) and longer terminal branches, is notoriously difficult to resolve and generally requires large amounts of DNA sequence for large numbers of taxa (e.g., Baurain et al. 2007; Jian et al. 2008; Rokas and Carroll 2005; Wurdack and Davis 2009), an approach that has yet to be tried for the Andropogoneae. An attempt to synthesize the available trees for this chapter by using a supertree approach produced a tree that was almost entirely unresolved (not shown).

The poor phylogenetic resolution of Andropogoneae, and general lack of appropriate taxon sampling, means that we do not know precisely where *Saccharum*, *Miscanthus* and *Sorghum* fall within the tribe and what their closest relatives are. Based on their morphology and molecular data, they clearly belong in the awned Andropogoneae, a clade that includes about two thirds of the genera of the tribe. Molecular data also show that they fall outside the core Andropogoneae, and thus the genera of that group can be ruled out as near relatives. Some studies hint at a

clade that includes both *Saccharum* and *Sorghum* (Bombliès and Doebley 2005; Hodkinson et al. 2002a; Skendzic et al. 2007), but the group is not strongly supported. Early studies linking the two genera included too few taxa to evaluate relationships rigorously (Al-Janabi et al. 1994; Hamby and Zimmer 1988; Sobral et al. 1994). Other genera that have been linked to the *Saccharum/Sorghum* group by one or more studies include *Cleistachne*, *Microstegium*, *Miscanthus*, and *Sorghastrum*. Skendzic et al. (2007) provide data on several species of *Sorghastrum*, but as with all other studies, their relationship to *Sorghum* and to other Andropogoneae is ambiguous.

3.2 Morphological Evolution

Members of Andropogoneae have paired spikelets, one of which is sessile and one of which is pedicellate, although this characteristic is shared with many other Panicoideae (Kellogg 2000a; Zanotti et al. 2010). The ancestral condition for the tribe is for the lower flower of the sessile spikelet to be staminate, and the lemmas acute, lacking awns. Anatomically, epidermal papillae appear to have been ancestrally absent, and costal short cells in long rows (Watson and Dallwitz 1992). In addition, in most species the rachis (inflorescence stalk) breaks up at maturity. This character appears in the common ancestor of all members of the tribe except for *Arundinella*. The disarticulating rachis is lost independently in several genera, including *Miscanthus* and sorghum, although in sugarcane the lateral branches disarticulate as well. It also appears that hardened glumes, as exhibited by sorghum and many other genera, originated at this same point in the phylogeny, but this characteristic was lost later in evolutionary time.

All Andropogoneae use the C_4 photosynthetic pathway, and use NADP-ME as a decarboxylating enzyme. Associated with this C_4 subtype, the vascular bundles have a single sheath (Hattersley and Watson 1975). This photosynthetic pathway might constitute a synapomorphy for the Andropogoneae (Christin et al. 2008), or might have been derived earlier, possibly at the origin of the Panicoideae (Vicentini et al. 2008). The optimization of this character on the phylogeny depends heavily on the taxa included in the tree and also on the particular model of evolution used.

The earliest lineages of Andropogoneae, including *Coix* and the *Zea-Tripsacum* clade, lacked awns on the lemmas. Most phylogenies suggest that these awnless lineages form a paraphyletic grade, rather than a clade. In contrast, the awned Andropogoneae, a group that includes about two thirds of the genera (ca. 55) appear to be monophyletic, based on all molecular analyses to date (see above). (Some analyses also include *Coix* here even though it is clearly awnless.) The awn is usually twisted and hygroscopic, and is borne on the lemma. Awns are thought to be adaptations for seed dispersal (Elbaum et al. 2007; Garnier and Dajoz 2001; Peart 1979, 1981, 1984; Peart and Clifford 1987), although they also affect seed provisioning (e.g., Li et al. 2006; Motzo and Giunta 2002), and possibly also drought stress (Abebe et al. 2010). Awns are lost in some taxa, including a few relatives of sugarcane.

3.3 Chromosomal Evolution

The ancestral base chromosome number (n) for the Andropogoneae is most likely 10 (Spangler et al. 1999; Wilson et al. 1999). The subfamily Panicoideae has three clades, one with $x=9$, one with $x=10$, and the Andropogoneae; the latter two are sisters (Christin et al. 2008; Vicentini et al. 2008), further supporting 10 as the base number. The earliest diverging genus in Andropogoneae is *Arundinella*, for which chromosome numbers of $2n=16$ (Basappa and Muniyamma 1981), 18 (Christopher and Samraj 1985; Mehra 1982), 20 (Mehra 1982; Norrmann et al. 1994; Pohl and Davidse 1971; Sahni and Bir 1985), 24 (Mehra 1982; Rudyka 1990), 34 (Mehra 1982; Sinha et al. 1990), 40 (Christopher and Samraj 1985), and 60 (Mehra 1982) have been reported, although not all counts have been confirmed. Most other genera of the tribe have chromosome numbers that are multiples of 10, although a few include multiples of nine.

One popular idea is that the ancestor of the tribe had a $2n$ number of 10, and thus a haploid number of 5 (Celarier 1956; Garber 1950). This creates some appealing arithmetic to explain the origin of the two genomes of *Zea*. Proponents of this idea point to the handful of species in Andropogoneae that exhibit $n=5$ (*Coix aquatica*, and members of *Sorghum* subg. *Parasorghum* (= *Sarga*)). However, the $n=5$ species of *Sorghum* are clearly derived in the phylogeny (Spangler et al. 1999), pointing to a secondary reduction in chromosome number. This evidence is further supported by the finding (Paterson et al 2004, 2009) that the *Sorghum bicolor* ($n=10$) genome has not experienced genome duplication or paleopolyploidy in 70 My or more, ruling out the possibility of formation of its $n=10$ karyotype from $n=5$ sorghums. Price et al. (2005) noted that the direction of evolution between $n=10$ and $n=5$ sorghums is ambiguous, but their analysis did not include any other Andropogoneae. Garber (1950) and Price et al. (2005) found that chromosomes of the $x=5$ species were considerably larger than those of the $x=10$ taxa; the $x=5$ sorghums also have a higher 2C DNA content than the $x=10$ species (Price et al. 2005). These observations raise the possibility that $x=5$ represents an intriguing chromosomal fusion event, rather than the ancestral base chromosome number. One can imagine that a burst in retrotransposon activity could have led simultaneously to the abrupt increase in genome size seen in this group, as well as the genome rearrangements that led to five large rather than ten small chromosomes. Price et al. (2005) correct the published data on *S. leiocladum*, noting that this species is actually $n=5$ rather than $n=10$ as reported by Garber (1950) and cited by Spangler et al. (1999). They also note that the report of $n=5$ for *S. nitidum* is almost certainly an error, and that multiple accessions of this species are $n=10$.

Wilson et al. (1999) suggested an ancestral number of $x=8$ for maize, but this number is quite rare in Andropogoneae. One species of *Arundinella* is reported with this base number (Basappa and Muniyamma 1981), and *Chasmopodium*, with two species living in West Africa, is reported to have $n=8$; *Chasmopodium* has not been included in any phylogeny to date. Thus, if maize were convincingly shown to have arisen from an $x=8$ ancestor, it would reflect a highly unusual set of chromosomal rearrangements.

To summarize the available chromosomal data, by far the majority of Andropogoneae and their sister clade in Paniceae have a chromosome base number of 10, not 5. A handful of taxa have chromosome numbers less than $x=10$. For convincing phylogenetic evidence that these represent the ancestral state, these taxa would have to be sisters to all other members of the tribe. *Coix* species sometimes appear near the base of the tribe, but never with strong support, and *Sorghum* subg. *Parasorghum* is always found to be derived. In addition, the genera *Sorghum* and *Arundinella* (if not others) would have to be nonmonophyletic. Because neither of these conditions holds, the phylogeny provides no evidence for $n=5$ as being ancestral. The evidence from chromosome size and DNA content is also equivocal. The fact that the *Sorghum* species with $x=5$ have larger chromosomes could also indicate that they are derived. The issue will only be resolved by genomic studies on the taxa with chromosome numbers other than $x=10$.

4 Phylogeny of *Saccharum* and *Miscanthus*

Within the awned Andropogoneae is a group of species, including *Saccharum* and *Miscanthus*, characterized by having sessile and pedicellate spikelets alike in form and sex expression. Because of the lack of differentiation between the sessile and pedicellate spikelets, Clayton and Renvoize (1986) postulated that these were the most primitive of the Andropogoneae, but this hypothesis has not been supported by molecular phylogenetic data.

The “*Saccharum* complex” was defined originally by Mukherjee (1957) to include *Narenga*, *Sclerostachya*, *Erianthus* sect. *Ripidium*, and *Saccharum*. In most species in this complex, the main axis of the inflorescence is tough and does not break up at maturity; the lateral branches, however, disarticulate between the spikelet pairs. The inflorescence axis and branches are covered with long hairs, as is the base of the spikelet (callus). Although awned lemmas are common in this group, awns fail to develop in some species.

Mukherjee (1957) did not include *Miscanthus* in his original delimitation of the *Saccharum* group, but it was later added (Daniels and Williams 1975). *Miscanthus* is morphologically similar to *Saccharum*, but the sessile spikelet is actually on a short pedicel, and the lateral inflorescence branches do not break up at maturity.

The most comprehensive molecular phylogenetic study to investigate the “*Saccharum* complex” (including *Miscanthus*) is that of Hodkinson et al. (2002a). This study included multiple species of *Saccharum* and *Miscanthus*, as well as representatives of *Erianthus*, *Eulalia*, *Pogonatherum*, *Imperata*, *Narenga*, and *Spodiopogon*. As with all other phylogenetic studies in the group, the relationships are mostly weakly supported, and the results are somewhat inconclusive.

Figure 1.3 summarizes the ITS phylogeny from Hodkinson et al. (2002a) showing only groups that receive some support from their parsimony bootstrap analysis. A group corresponding to the genus *Saccharum* in the strict sense (*sensu stricto*, or *s.s.*) is well supported by the ITS sequences, and also when the *trnL-F* data are

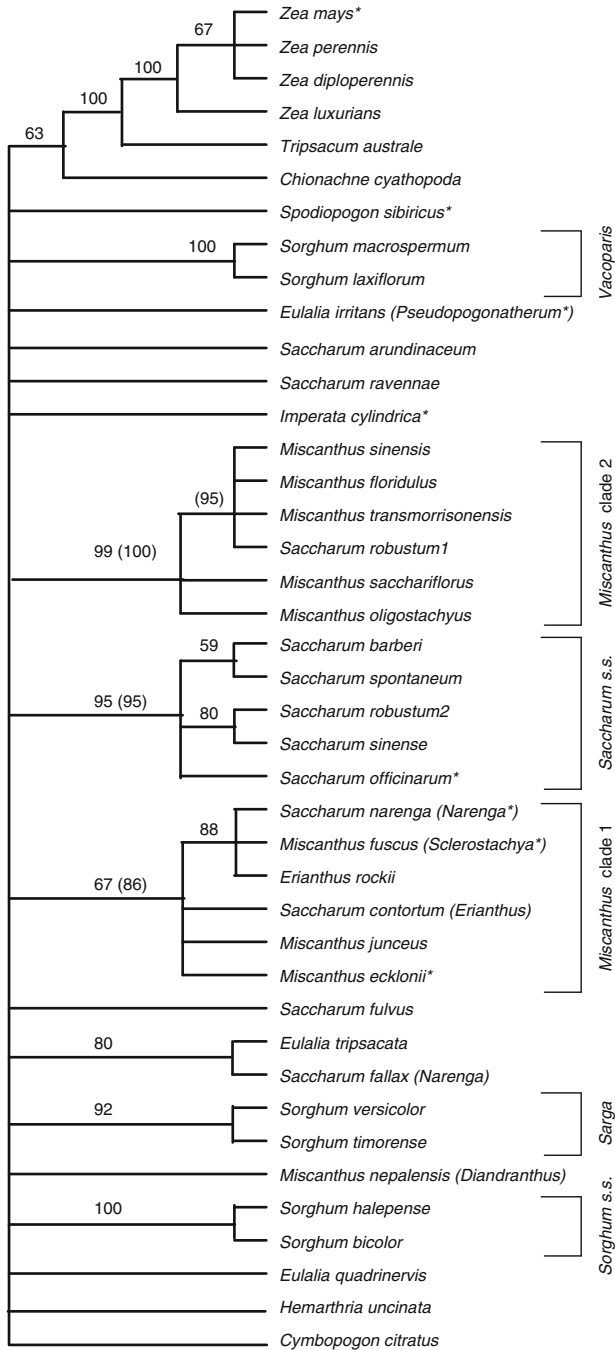


Fig. 1.3 Phylogeny of *Saccharum* and its relatives based on ITS sequences, redrawn from Hodkinson et al. (2002a). Numbers above branches are parsimony bootstrap values. Branches with less than 50 % bootstrap support are collapsed in this figure. Numbers in parentheses refer to support values obtained when the ITS data were combined with data for *trnL-F*. *Sorghum robustum* 1 and 2, and *Miscanthus floridulus* 1 and 2 are distinct paralogues from the same plant. Type species are marked with an asterisk

added (number in parentheses). There are two distinct clades of *Miscanthus* species. *Miscanthus* clade 1 includes the type species of *Miscanthus*, *M. ecklonii*, as well as species previously assigned to *Erianthus*, *Sclerostachya*, and *Narenga*; this group is only moderately supported by ITS data, but receives stronger support when the *trnL-F* data were added. *Miscanthus* clade 2 includes several species of *Miscanthus* plus one ITS paralogue from *Saccharum robustum*. *Miscanthus* × *giganteus* is a triploid derivative of *M. sinensis* and *M. sacchariflorus* (Hodkinson et al. 2002b), so could also be assigned to this clade.

Several earlier studies produced preliminary results that are consistent with those of Hodkinson et al. (2002a). Nair et al. (1999) used RAPD markers and found a group corresponding to *Saccharum* s.s., and another similar to *Miscanthus* clade 1; the analysis was phenetic, however (using Unweighted Pair Group Method with Arithmetic Means, or UPGMA), and cannot be directly compared to phylogenetic studies. Besse et al. (1997) used RFLP data to show that seven species of *Erianthus* were distinct from two species of *Saccharum*, and Selvi et al. (2006) likewise found a clear distinction between *Saccharum* species and *Erianthus* using AFLPs. Bacci et al. (2001) generated an ITS phylogeny of sugarcane and its relatives, and also found a clade corresponding to *Saccharum* s.s.

Within *Saccharum*, six species are commonly recognized: *S. officinarum*, *S. robustum*, *S. spontaneum*, *S. sinense*, *S. barberi*, and *S. edule*, although the grass species index at the Royal Botanic Gardens, Kew lists 37 names (<http://www.kew.org/data/grassbase/index.html>). Relationships among the species are not well resolved by available data. A study of DNA sequences from 18 chloroplast regions suggested that *S. spontaneum* is sister to the remaining species (Takahashi et al. 2005). This contradicts the results of several other studies, including the relationships shown in Fig. 1.3. However, an extensive review of *Saccharum* literature (Chap. 3 of this volume) supports this relationship, also generally viewing *S. sinense*, *S. barberi*, and *S. edule* as forms of *S. officinarum* modified by interspecific hybridization with *S. spontaneum*. Few of the studies have used accepted phylogenetic methods, and none has attempted to dissect the complex reticulate history of the *Saccharum* species using multiple single copy nuclear genes. It is virtually certain that the relationships within the genus *Saccharum* are not strictly divergent, and hence attempts to represent them as a tree are probably misleading.

5 Phylogeny of Sorghum

Sorghum has a highly branched panicle that is superficially quite different from the set of long racemes of *Saccharum*. The glumes are hardened, as is characteristic of many Andropogoneae. The sessile and pedicellate spikelets are quite different morphologically, with the pedicellate spikelet generally much smaller than the sessile one and either staminate or sterile.

Sorghum has been the subject of several molecular phylogenies, of which two have included all or nearly all the species (Dillon et al. 2007; Ng'uni et al. 2010).

The genus has also been the subject of a recent monograph (Spangler 2003). Thus, we know what species are in the genus and what their relationships are. Ng'uni et al. (2010) used noncoding regions of the chloroplast and the ITS; Dillon et al. (2007) added data on a nuclear gene (*Adh1*) to her previous data on ITS1 and *ndhF* (Dillon et al. 2004).

Both Dillon et al. (2007) and Ng'uni et al. (2010) found two well-supported major clades within *Sorghum*. Clade 1 includes *Sorghum bicolor* and its close relatives *S. halepense*, *S. propinquum*, *S. arundinaceum*, *S. alnum*, and *S. drummondii*, supporting earlier work (Sun et al. 1994). (Note that DeWet (1978) places most species of this group except *S. halepense* and *S. propinquum* in the synonymy of *S. bicolor*; however, he does not deal with *S. alnum*.) These species are all African except for *S. propinquum*, which is Asian, and are all part of the secondary gene pool of grain sorghum (*S. bicolor*) (Price et al. 2006). Also in Clade 1 are *Sorghum macrospermum* and *S. laxiflorum*, two Australian species that are clearly sisters and are more closely related to cultivated sorghum than any of the other Australian species (Dillon et al. 2004).

Clade 2 within *Sorghum* includes the 17 species have been assigned to subgenera *Stiposorghum* and *Parasorghum*, all Australian species that can be recognized easily by their bearded nodes (Snowden 1935). The Australian species of *Sorghum* are not interfertile with cultivated sorghum, and pollen from members of one group will not germinate on the stigmas of the other (Garber 1950; Hodnett et al. 2005; Price et al. 2006). All phylogenetic studies support the Australian clade, but show that species of the two subgenera are intermixed (Dillon et al. 2007; Ng'uni et al. 2010; Spangler et al. 1999; Sun et al. 1994). Spangler (2003) evaluated morphological similarities among the 17 species and concluded that there were only seven, although Dillon et al. (2007) argued for reinstating several (Table 1.1).

The trees differ in the placement of *S. nitidum*, a widespread species of Australia and Asia. Ng'uni et al. (2010) place it sister to Clade 2, whereas Dillon et al. (2007) place it within that clade, sister to *S. leiocladum*. Sun et al. (1994) place *S. nitidum* as sister to the African sorghums, and Spangler et al. (1999) place it sister to *S. laxiflorum*. Resolution of its placement will probably require sampling of multiple individuals from different parts of its range.

Cleistachne sorghoides has been included in some studies but not others. The plants look similar to cultivated sorghum (hence the specific epithet) but lack the pedicellate spikelet; a close association of the two genera has been postulated since the 19th century (Hackel 1889). Both Sun et al. (1994) and Dillon et al. (2004) placed *Cleistachne* among the Australian sorghums, a somewhat surprising result because *Cleistachne* is African. However, Dillon et al. (2007) place *Cleistachne* sister to the clade made up of subg. *Sorghum* and *Heterosorghum*. The difference likely reflects the signal from *Adh1*, which was only included by Dillon et al. (2007).

Monophyly of sorghum in its traditional sense (i.e., clades 1 and 2 together) is neither proven nor disproven by available data. Dillon et al. (2007) strongly assert that *Sorghum* is monophyletic. However, both they and Ng'uni et al. (2010) begin with an assumption of monophyly, and do not include *Microstegium*, *Saccharum*, *Miscanthus*, *Erianthus*, *Imperata*, *Sorghastrum*, or any of the other taxa that could

Table 1.1 Alternate classifications of the genus *Sorghum*

Single genus classification (Dillon et al. 2007)	Three-genus classification (Spangler 2003)
<i>Sorghum</i> subg. <i>Parasorghum</i>	<i>Sarga</i>
<i>Sorghum angustum</i>	<i>Sarga angustum</i>
<i>Sorghum ecarinatum</i>	[not named in <i>Sarga</i>]
<i>Sorghum interjectum</i>	[not named in <i>Sarga</i>]
<i>Sorghum intrans</i>	<i>Sarga intrans</i>
<i>Sorghum leiocladum</i>	<i>Sarga leiocladum</i>
<i>Sorghum nitidum</i>	[not named in <i>Sarga</i>]
<i>Sorghum plumosum</i> (= <i>S. grande</i>)	<i>Sarga plumosum</i>
<i>Sorghum purpureo-sericeum</i> (= <i>S. pappii</i> , <i>S. deccanense</i> , <i>S. dimiditum</i>)	<i>Sarga purpureo-sericeum</i>
<i>Sorghum timorense</i> (= <i>S. amplum</i> , <i>S. australiense</i> , <i>S. brachypodum</i> , <i>S. brevicallusum</i> , <i>S. bulbosum</i> , <i>S. matarankense</i> , <i>S. mjoebergii</i> , <i>S. stipoidesum</i>)	<i>Sarga timorense</i>
<i>Sorghum trichocladum</i>	<i>Sarga trichocladum</i>
<i>Sorghum versicolor</i>	<i>Sarga versicolor</i>
<i>Sorghum</i> subg. <i>Sorghum</i>	<i>Sorghum</i>
<i>Sorghum bicolor</i>	<i>Sorghum bicolor</i>
<i>Sorghum halepense</i>	(= <i>S. propinquum</i> , <i>S. × alnum</i> , <i>S. × drummondii</i>)
<i>Sorghum propinquum</i>	<i>Sorghum halepense</i>
<i>Sorghum × alnum</i>	<i>Sorghum nitidum</i>
<i>Sorghum × drummondii</i>	
<i>Sorghum</i> subg. <i>Chaetosorghum</i>	<i>Vacoparis</i>
<i>Sorghum laxiflorum</i>	<i>Vacoparis laxiflorum</i>
<i>Sorghum macrospermum</i>	<i>Vacoparis macrospermum</i>

be related to one or the other of the sorghum clades. Spangler et al (1999) and Hodkinson et al. (2002a) included a broad sample of *Sorghum* species and found that the well-supported clades might not actually particularly closely related (Fig. 1.3). However, because of the lack of resolution of the backbone of the tree, monophyly of the genus cannot be ruled out.

6 Classification of Saccharinae and Sorghinae

In an ideal world, the classification of a plant would be based directly on its position in a phylogenetic tree, such that the name provides unambiguous information on genealogical relationships. Much taxonomic effort in recent years has gone into generating such trees, and adjusting classifications to reflect current knowledge of relationships. In the case of Saccharinae and Sorghinae, unfortunately, the molecular data are inconclusive, so that current classifications rely on a mix of morphological observations (the traditional classification) and fragmentary knowledge of relationships.