

Relatedness of *Saccharum* species hybrids and wild relatives in eastern South Africa

Ms H Khanyi

 **orcid.org 0000-0001-6706-2116**

Dissertation submitted in fulfilment of the requirements for the degree Master of Science in Environmental Sciences at the North-West University

Supervisor: Prof S Barnard

Co-supervisor: Prof SJ Siebert

Assistant Supervisor: Dr SJ Snyman

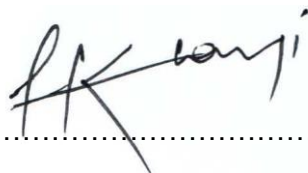
Graduation ceremony: July 2018

26948370

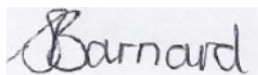
DECLARATION

I declare that the work presented in this *Magister Scientiae* dissertation is my own work, that it has not been submitted for any degree or examination at any other university, and that all the sources I have used or quoted have been acknowledged by complete reference.

Signature of the student

A handwritten signature in black ink, appearing to read 'H. K. van der', is placed on a light blue rectangular background. The signature is written in a cursive style with a long horizontal stroke extending to the right.

Signature of the supervisor

A handwritten signature in black ink, appearing to read 'Barnard', is placed on a light blue rectangular background. The signature is written in a cursive style with a long horizontal stroke extending to the right.

Signature of the co-supervisor

A handwritten signature in black ink, appearing to read 'H. K. van der', is placed on a light blue rectangular background. The signature is written in a cursive style with a long horizontal stroke extending to the right.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to those who have contributed significantly to the success of this dissertation and supported me in one way or the other during this amazing journey. Thus, I would like to extend my special thanks to:

- My parents for emotional support and my son for being patient with my absence throughout the duration of this study
- My supervisors: Prof Sandra Barnard and Prof Stefan Siebert for granting me an opportunity to be part of this project. I appreciate all your guidance, immense knowledge and time invested throughout the stages of this study.
- Dr Sandy Snyman (assistant supervisor) for her fruitful and timely assistance in contacting the relevant people prior to field work and providing helpful supplementary information in this study.
- Dr Deborah Sweby for molecular work advice and conducting sequencing reactions.
- Extension specialists, biosecurity officers and pest and disease teams from SASRI in the Mpumalanga and KwaZulu-Natal regions visited for sampling
- Mr Sifiso Thwala for his interest in the project and willingness to assist with extra information required after the sampling period
- Mr Dennis Komape for field identification of wild species and assisting throughout fieldwork
- Mr Perfection Chauke for assisting with driving during field work
- Biosafety South Africa, NWU, SASRI and NRF for financial support

All Glory be to God; For from Him and through Him and to Him are all things.

ABSTRACT

Gene flow between crops and their cross-compatible wild relatives is undesirable in commercial production systems. The development of genetically modified (GM) sugarcane is set to provide new opportunities to increase yield and grow the global competitiveness of the South African sugar industry. However, the concern when cultivating GM plants is that the transfer of transgenes to related species may enhance their capacity for invasiveness. Therefore, biosafety studies are a legal requirement to evaluate the potential impact of GM crops on the environment before commercial release. The aim of the study was to contribute data to such an initiative by assessing the gene flow potential from sugarcane to its wild relatives in the major sugar production regions of the Mpumalanga and KwaZulu-Natal provinces of South Africa. Three approaches were followed: (1) a systematic literature review was conducted to identify individuals reported to have spontaneously hybridised with sugarcane in the past; (2) two chloroplast (*matK* and *rbcL*) and one nucleic (ITS) DNA barcodes of related species were sequenced and phylogenetic analyses were done to investigate relatedness with *Saccharum* species hybrids; and (3) field assessments were conducted to determine pollen viability of commercial sugarcane varieties using the Iodine Potassium Iodide (IKI) and Triphenyl Tetrazolium Chloride (TTC) staining methods.

A total of 36 hybridization incidents were reported in the literature, of which none were spontaneous hybridization with sugarcane. These crosses were mainly made with the genera *Saccharum* (Syn. *Erianthus*) and *Sorghum*. Regardless of breeder's efforts to cross sugarcane or its progenitors with various species, hybrid generation was rare, and if successful, hybrids were reported to be weak with a slow growth rate. Since members of the Saccharinae and Sorghinae, which have successfully been crossed with sugarcane, are prominent in the study area, phylogenetic studies were conducted to determine the degree of relatedness at the genome level. Generally, sugarcane was found to be closely related to its wild relatives, as the average nucleotide identity (ITS, *rbcL* and *matK*) ranged from 78-98%. Based on the three locus barcode clustering, pairwise distances and sequence identity of related species against the sugarcane varieties, *Sorghum arundinaceum*, *Miscanthus ecklonii* and *Imperata cylindrica* are the most closely related species to commercial sugarcane varieties in the cultivation areas. A test for statistical difference revealed that there was no significant difference (p -value= 0.622) between varying percentage pollen viability obtained from the IKI and TTC stains respectively. Pollen viability was found to decrease from the northern regions (85%) to the southern regions (0%) of the study area. The irrigated regions showed the highest pollen viability, which was frequently measured from sugarcane varieties N14 and N36. A significant association between pollen viability with day length, maximum temperature (positive) and soil water content (negative) was observed. However, the study suggested that pollen viability is largely determined by the genetic makeup of the individual and mean maximum temperature. None of the commercial varieties with

inserted transgenes showed any pollen viability in this study. Even though sugarcane has the potential to hybridize with the closely related members of the Sorghinae and Saccharinae which are also present in the sugar production regions, gene flow would not occur without the production of viable pollen. However, some of the GM developments are with unreleased genotypes, thus, future studies are required to evaluate pollen viability from these genotypes especially when cultivated in the irrigated regions.

Keywords: Hybridization, pollen viability, relatedness, sugarcane varieties, gene flow, wild related species

TABLE OF CONTENTS

DECLARATION	I
LIST OF ABBREVIATIONS	X
LIST OF FIGURES.....	XII
LIST OF TABLES	XVIII
CHAPTER 1: INTRODUCTION	1
1.1 Background and Rationale.....	1
1.2 Aim and objectives	3
1.3 Profile of the sugar industry	3
CHAPTER 2: LITERATURE REVIEW	6
2.1 Origin and classification.....	6
2.2 Morphology and developmental stages	6
2.2.1 Plant morphology.....	6
2.2.2 Reproductive morphology	7
2.2.3 Environmental gradients affecting flowering and pollen viability.....	9
2.2.4 Physiological and morphological changes throughout plant development.....	11
2.3 Uses of sugarcane.....	13
2.3.1 Sugar production	13
2.3.2 By-products: Fuel, energy and fertiliser	13
2.4 Overview of sugarcane breeding in South Africa	13

2.5	Challenges with sugarcane breeding	15
2.6	Genetically Modified sugarcane	15
2.6.1	Approaching biotechnology tools	15
2.6.2	Progress with transgenic sugarcane	16
2.7	International and national status of GM crops	16
2.8	Benefits of GM crops	16
2.9	Gene flow	17
2.10	Cases of transgene escape	18
2.11	Biosafety.....	19
2.12	Environmental risk analysis and regulation.....	19
CHAPTER 3: SPECIES REPORTED TO HAVE HYBRIDIZED WITH SACCHARUM		21
3.1	Introduction.....	21
3.2	Methodology	21
3.3	Results	22
3.3.1	Type of hybridization involving <i>Saccharum</i>	22
3.3.2	Parental lines in artificial crosses with <i>Saccharum</i>	23
3.3.3	Motives for crosses.....	23
3.3.4	Artificial crosses achieved and seedling survival rate	23
3.3.5	Distribution of relative parent lines	24
3.3.6	Using molecular markers to verify hybrids	24
3.4	Discussion	34
3.4.1	Spontaneous hybridization between <i>Saccharum</i> species and hybrids with wild relatives	34

3.4.2	Artificial crosses achieved and seedling survival rate	34
3.4.3	Using molecular markers to verify hybrids	35
3.5	Conclusion.....	36
CHAPTER 4: DNA BARCODING.....		37
4.1	Introduction.....	37
4.2	Material and methods	38
4.2.1	Selection of <i>Saccharum</i> wild relatives.....	38
4.2.2	Parental lines and outgroups	38
4.2.3	Collection of plant material	38
4.2.4	DNA sequences sourced from genetic databases	43
4.2.5	DNA Extraction.....	43
4.2.6	Quantity and quality measurements.....	43
4.2.7	Polymerase Chain Reaction (PCR).....	45
4.2.8	Agarose gel electrophoresis	46
4.2.9	Post PCR Purification	49
4.2.10	Sequencing	49
4.2.11	Editing and alignment of sequences	49
4.2.12	Results analysis.....	50
4.3	Results	52
4.3.1	Sequence statistics of each gene fragment	52
4.3.2	Nucleotide identity and pairwise differences between wild relatives and <i>Saccharum</i> species	52
4.3.3	Maximum Likelihood method	54

4.3.4	Neighbor Joining method.....	59
4.3.5	Congruency.....	63
4.4	Discussion	69
4.4.1	Efficiency of barcodes in phylogeny construction.....	69
4.4.2	Relatedness of the <i>Saccharum</i> complex with sugarcane varieties.....	69
4.4.3	Relatedness with the <i>Sorghum</i> genus	70
4.4.4	Relatedness with the <i>Miscanthus</i> genus.....	71
4.4.5	Relatedness with the <i>Saccharum</i> genus (previously classified as <i>Erianthus</i>).....	71
4.4.6	Relatedness with the <i>Imperata</i> genus.....	71
4.5	Conclusion.....	72
CHAPTER 5: POLLEN VIABILITY		73
5.1	Introduction.....	73
5.2	Materials and methods	73
5.2.1	Plant material	73
5.2.2	Sampling sites	74
5.2.3	Stains used to test for pollen viability	76
5.2.4	Testing pollen viability using staining methods	78
5.2.5	Weather data.....	78
5.2.6	Data analysis.....	78
5.3	Results	80
5.3.1	Pollen viability of sugarcane varieties	80
5.3.2	Variance between the staining methods	81

5.3.3	Flowering sugarcane varieties and geographic range of pollen viability ...	82
5.3.4	Environmental variability across sites	84
5.3.5	Optimal environmental variables for pollen viability.....	90
5.3.6	Correlation between pollen viability and environmental variables	91
5.3.7	Multivariate analysis of principal components	94
5.4	Discussion	96
5.4.1	Correlation of staining methods	96
5.4.2	Pollen viability and the real determinant factors	96
5.4.3	Current sugarcane varieties in relation to the unreleased GM varieties ...	99
5.5	Conclusion.....	99
CHAPTER 6: CONCLUSION AND RECOMMENDATIONS		101
REFERENCE LIST		103
APPENDICES		113
ADDENDUM		129

LIST OF ABBREVIATIONS

<u>Acronym</u>	<u>Definition</u>
2n	Somatic chromosome number
AC	Advisory Committee
AE	Acetic Acid; EDTA
AT	adenine-thymine
BOLD	Barcode of Life Database
Bt	<i>Bacillus thuringiensis</i>
CBD	Convention of Biological Diversity
CBOL	Consortium for the Barcode of Life
CO1	Cytochrome c oxidase subunit I
DAFF	The Department of Agriculture, Forestry and Fisheries
DL	Day length
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleotide Triphosphate
EC	Executive Council
EDTA	Ethylenediaminetetraacetic acid
eNGOs	environmental nongovernmental organizations
GC	Guanine-Cytosine
GIS	Geographic information systems
GM	Genetic Modification
GMOs	Genetic Modified organisms
IDT	Integrated DNA Technologies
ITS	Internal transcribed spacer
IKI	Iodine Potassium Iodide
JC	Jukes-Cantor
K2	Kimura 2-parameter
Kb	Kilobyte
LPD	Local Pest Disease
Mya	million years ago
MAST	maximum agreement subtrees
matK	maturase K/ Megakaryocyte-Associated Tyrosine Kinase
MCL	Maximum Composite Likelihood
MCMC	Markov chain Monte Carlo
mega 7	Molecular Evolutionary Genetics Analysis Version 7.0
ML	Maximum Likelihood
MP	Maximum Parsimony
Muscle	Multiple Sequence Comparison by Log-Expectation
n	Gametic chromosome number
NCBI	National Centre for Biotechnology Information
NJ	Neighbor-Joining
NNI	Nearest-Neighbor-Interchange
OECD	Organisation for Economic Cooperation and Development
PCA	Principal Component Analysis
PCR	Polymerase chain reaction

rbcl	Ribulose biphosphate carboxylase large chain
RH	Relative humidity
RSA	Republic of South Africa
Sa.	<i>Saccharum</i>
SD	Standard deviation
SE	Standard error
SANBI	South African National Biodiversity Institute
SASRI	South African Sugarcane Research Institute
SSR	Simple sequence repeat
So.	<i>Sorghum</i>
SWC	Soil water content
TAE	Tris; Acetic Acid; EDTA
TE	Tris EDTA
TMN	Maximum temperature
TMX	Minimum temperature
tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
TSB	Transvaal Suiker Beperk
TTC	Triphenyl Tetrazolium Chloride
UCL	Union cooperative limited
UV	Ultra Violet
VCC	Variety Control Committee
VRC	Variety Release Committee

LIST OF FIGURES

Figure 1.1: The South African sugarcane industry and mills located in the Mpumalanga and KwaZulu-Natal provinces. (Map produced by Mzimase Jalisa- SASRI GIS intern).....	4
Figure 2.1: Morphology of sugarcane illustrated by a) the phylotaxy and numbering of the leaves (taken from Cheavegatti-Gianotto <i>et al.</i> 2011) and b) the root system indicating sett and shoot roots (taken from Smith <i>et al.</i> , 2005)	7
Figure 2.2: Reproductive morphology of a sugarcane crop. a) Rachis formed from the apex of the meristem (Photo: H. Khanyi), b) the panicle carrying secondary and tertiary branches of the inflorescence (Photo: H. Khanyi), c) sessile spikelets, d) purple and fluffy stigmas on the gynoecium and e) anthers in the androecium (Photos: Amaral <i>et al.</i> , 2013)	8
Figure 2.3: Reproductive morphology of a sugarcane crop. a) Mature anthers with pollen grains (Photo: Amaral <i>et al.</i> , 2013), b) flowering of the sugarcane inflorescence on the first two thirds with no flowering at the base (Photo: H. Khanyi), c) post flowering of the inflorescence with branches closed (Photo: H. Khanyi), and d) the formation of fuzz and subsequent seed production (Photo: D.M. Komape).	9
Figure 2.4: Flowering of sugarcane varieties a) N36 and b) N23 at Malelane, Mpumalanga (05 July 2017) (Photos: H. Khanyi).....	10
Figure 2.5: Developmental stages of sugarcane from planting to post harvest. (Taken from Moore & Botha, 2013).	12
Figure 4.1: Localities where species were collected. N14 & N36: sugarcane commercial varieties; SV: <i>Sorghum versicolor</i> ; TS: <i>Trachypogon spicatus</i> MJ; <i>Miscanthus junceus</i> ; SD: <i>Sorghum x drummondii</i> ; ME: <i>Miscanthus ecklonii</i> ; SASRI: Site where NCo376 (sugarcane commercial variety), <i>Saccharum arundinaceum</i> , <i>Saccharum ravennae</i> , <i>Saccharum officinarum</i> , <i>Saccharum sinense</i> , <i>Saccharum robustum</i> , <i>Saccharum spontaneum</i> ; <i>Sorghum arundinaceum</i> and <i>Imperata cylindrica</i> were collected. Names of these localities can be found in Appendix A-1.	39

Figure 4.2: Nanodrop measurement window: a) Bad DNA quality with low concentration; b) Acceptable DNA quality and concentration to be used in downstream applications..... 44

Figure 4.3: Successful DNA amplification of the ITS DNA barcode with length of ± 650 bp for samples 1,3 -12. Sample 2 produced multiple non-specific bands while sample 13 had a weak band, these were used for the second round of PCR. The first well marked 'M' is the DNA molecular weight marker, ranging from 250-10000 bp. 1: *Saccharum robustum*; 2: *Saccharum spontaneum*; 3: *Saccharum ravennae*; 4: *Saccharum arundinaceum*; 5: *Saccharum officinarum*; 6: *Saccharum sinense*; 7: *Saccharum* sp. 'Rowan Green'; 8: *Saccharum* sp. Co745; 9: *Saccharum* sp. N14; 10: *Saccharum* sp. N36; 11: *Saccharum* sp. NCo376; 12: *Imperata cylindrica* and 13: *Miscanthus ecklonii*..... 47

Figure 4.4: Successful DNA amplification of the *rbcL* DNA barcode with length of ± 550 bp for 12 samples. Primer dimers also formed from this reaction. 1: *Saccharum robustum*; 2: *Saccharum spontaneum*; 3: *Saccharum ravennae*; 4: *Saccharum arundinaceum*; 5: *Saccharum officinarum*; 6: *Saccharum sinense*; 7: *Saccharum* sp. Rowan green; 8: *Saccharum* sp. Co745; 9: *Saccharum* sp. N14; 10: *Saccharum* sp. N36; 11: *Saccharum* sp. NCo376 and 12: *Imperata cylindrica*..... 48

Figure 4.5: Successful DNA amplification of the *matK* DNA barcode with length of ± 750 bp for samples 1-11. Samples 12 and 13 produced weak bands thus were used for a second round of PCR. Primer dimers also formed from this reaction. 1: *Saccharum robustum*; 2: *Saccharum spontaneum*; 3: *Saccharum ravennae*; 4: *Saccharum arundinaceum*; 5: *Saccharum officinarum*; 6: *Saccharum sinense*; 7: *Saccharum* sp. Rowan green; 8: *Saccharum* sp. Co745; 9: *Saccharum* sp. N14; 10: *Saccharum* sp. N36; 11: *Saccharum* sp. NCo376; 12: *Imperata cylindrica* and 13: *Miscanthus ecklonii*..... 48

Figure 4.6: Maximum Likelihood phylogeny (K2 model, 1000 bootstrap replicates) based on a DNA fragment of ITS The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. 56

Figure 4.7: Maximum Likelihood phylogeny (K2 model, 1000 bootstrap replicates) based on a DNA fragment of <i>rbcL</i> . The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.	57
Figure 4.8: Maximum Likelihood phylogeny (K2 model, 1000 bootstrap replicates) based on a DNA fragment of <i>matK</i> . The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.	58
Figure 4.9: The evolutionary history was inferred using the Neighbor-Joining method based on a DNA fragment of ITS. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site.....	60
Figure 4.10: The evolutionary history was inferred using the Neighbor-Joining method based on a DNA fragment of <i>rbcL</i> . The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site	61
Figure 4.11: The evolutionary history was inferred using the Neighbor-Joining method based on a DNA fragment of <i>matK</i> . The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site.	62
Figure 4.12: Maximum Likelihood phylogeny for <i>rbcL</i> + <i>matK</i> . The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches.	64

Figure 4.13: Maximum Likelihood phylogeny for <i>rbcL</i> + <i>matK</i> +ITS. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches	65
Figure 4.14: Neighbor-joining tree (Jukes-Contor model) based on core barcode of the <i>matK</i> and <i>rbcL</i> region using Geneious R11. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.	67
Figure 4.15: Neighbor-joining tree (Jukes-Contor model) based on the core barcode and ITS region using Geneious R11. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. There were a total of 1,105 positions in the final dataset.	68
Figure 5.1: Sugarcane production regions and locations of sugar mills in the Mpumalanga and KwaZulu-Natal Provinces. See text for explanation of sites 1-9. (Map produced from data provided by SASRI)	75
Figure 5.2: Viable and non-viable pollen grains of sugarcane observed using the TTC staining method. Nikon AZ100 M Stereoscopic microscope (Nikon Instruments Inc.) 100X magnification.	76
Figure 5.3: Starch test using the IKI staining method, where (a) viable and (b) non-viable pollen grains of sugarcane were observed. Nikon AZ100 M Stereoscopic microscope (Nikon Instruments Inc.) 100X magnification.	77
Figure 5.4: Mean pollen viability of 11 sugarcane varieties measured across six study sites in July 2016. Site 1: Malelane; Site 2: Komatipoort (Mpumalanga); Site 3: Pongola; Site 4: Jozini; Site 6: Empangeni; Site 8: Mount Edgecombe (KwaZulu-Natal).	80
Figure 5.5: Mean pollen viability of eleven different sugarcane varieties measured across nine study sites during the month of July 2017. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5:	

Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe
and site 9: Port Shepstone (KwaZulu-Natal)..... 81

Figure 5.6: Box and whiskers plot of the mean pollen viability of sugarcane tested for the
IKI (Iodine potassium iodide) and TTC (2,3,5- triphenyl tetrazolium
chloride) staining methods. $\pm$ Standard error (SE) and standard deviation
(SD), n=43. 82

Figure 5.7: Colour codes assigned to each of the classes for pollen viability. 82

Figure 5.8: Box and whiskers plot showing the mean minimum temperature (TMN) among
different sites, $\pm$ standard error (SE) and standard deviation (SD) n=43. A
 $p < 0.05$ indicated that there was a significant difference in daily minimum
temperatures between the sites highlighted in Appendix D-1. Box plots
with different letters indicate significant differences. Site 1: Malelane and
site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5:
Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe
and site 9: Port Shepstone (KwaZulu-Natal)..... 85

Figure 5.9: Box and whiskers plot showing the mean maximum temperature (TMX) across
the study sites, $\pm$ standard error (SE) and standard deviation (SD) n=43. A
 $p\text{-value} < 0.05$ indicated that there was a significant difference in maximum
temperatures between the sites highlighted in Appendix D-2. Box plots
with different letters indicate significant differences. Site 1: Malelane and
site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5:
Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe
and site 9: Port Shepstone (KwaZulu-Natal)..... 86

Figure 5.10: Box and whiskers plot showing the mean day length (DL) across the study
sites, $\pm$ standard error (SE) and standard deviation (SD) n=43. A $p < 0.05$
indicated that there was a significant difference in day lengths between
the sites highlighted in Appendix D-3. Box plots with different letters
indicate significant differences. Site 1: Malelane and site 2: Komatipoort
(Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6:
Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port
Shepstone (KwaZulu-Natal). 87

Figure 5.11: Box and whiskers plot showing the mean relative humidity (RH) across the
sites, $\pm$ standard error (SE) and standard deviation (SD) n=43. A $p < 0.05$
indicated that there was a significant difference in relative humidity

between the sites highlighted in Appendix D-4. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal)..... 88

Figure 5.12: Box and whiskers plot showing the mean soil water content (SWC) at 100mm across the study sites, $\pm$ standard error (SE) and standard deviation (SD) $n=43$. A $p<0.05$ indicated a significant statistical difference between the observed SWC at 100mm between the sites highlighted in Appendix D-5. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal)..... 89

Figure 5.13: Box and whiskers plot showing the difference in pollen viability for all varieties between irrigated and non-irrigated regions, $\pm$ standard error (SE) and standard deviation (SD) $n=43$ 90

Figure 5.14: Correlation between mean pollen viability and day length for viability assessed using TTC stain ($p=0.0004$)..... 92

Figure 5.15: Correlation between mean pollen viability and soil water content for viability assessed using TTC stain ($p=0.0084$)..... 92

Figure 5.16: Correlation between mean pollen viability and maximum temperature for viability assessed using TTC stain ($p=0.00009$) 93

Figure 5.17: Correlation between mean pollen viability and minimum temperature for viability assessed using TTC stain ($p=0.5636$) 93

Figure 5.18: Correlation between mean pollen viability and relative humidity for viability assessed using TTC stain ($p=0.5734$)..... 94

Figure 5.19: PCA of variables contributing to the variance within the pollen viability data, their relation to the first and second axes, as well as the individual sugarcane varieties at different sites. Variables represent soil water content (SWC), minimum temperature (TMN), maximum temperature (TMX), relative humidity (RH) and day length (DL) as environmental

variables and 2,3,5 Triphenyl tetrazolium chloride (TTC) and Iodine potassium iodide (IKI) as staining methods	95
---	----

LIST OF TABLES

Table 2.1: Selection stages of a new sugarcane variety by breeding programmes at SASRI (Taken from Parfitt, 2005)	14
Table 3.1: Hybridization incidents reported in the literature to involve either <i>Saccharum</i> species or commercial sugarcane cultivars, with wild related species. (¹ spontaneous and ² artificial hybridization)	25
Table 4.1: Targeted wild relatives of sugarcane from the Andropogoneae, their life form and habitat type.	40
Table 4.2: Species and hybrids included for phylogenetic tree construction within the Andropogoneae tribe.....	42
Table 4.3: Taxa selected as sugarcane related wild and outgroup species not collected in the field during this study. DNA nucleotide sequences (ITS, <i>matK</i> , <i>rbcL</i>) were obtained from GenBank.....	43
Table 4.4: The primers used for the amplification and sequencing of the ITS, <i>matK</i> and <i>rbcL</i> gene fragments.	45
Table 4.5: Amount (µl) of PCR reactants for each DNA fragment for the three rounds that were run during the optimising stages. Annealing temperatures remained at 50°C for ITS and <i>rbcL</i> , and 48°C for <i>matK</i> throughout these rounds.	46
Table 4.6: Statistics for each gene fragment from 24 DNA sequences after alignment and trimming	52
Table 4.7: The percentages of identical nucleotides from pairwise comparison of DNA sequences between wild relative species and commercial sugarcane species (Com.), as well as the parental lines (Parent).....	53
Table 4.8: The congruency analysis of the <i>matK</i> and <i>rbcL</i> gene fragments. Analysis showed trees having 24 leaves and MAST having 12 leaves.	63
Table 4.9: The congruency analysis of the core barcode and ITS gene fragments. Analysis showed trees having 24 leaves and MAST having 12 leaves.	63

Table 4.10 Congruency analysis of the <i>matK</i> and <i>rbcL</i> gene fragments. Analysis showed trees having 24 leaves and MAST having 16 leaves.	66
Table 4.11 Congruency analysis of the core barcode and ITS gene fragments. Analysis showed trees having 24 leaves and MAST having 13 leaves.	66
Table 5.1: Pollen viability percentage classes (low, intermediate or high) for 2016 and 2017 for sugarcane varieties and sites. Sites 1: Malelane and 2: Komatipoort (Mpumalanga), sites 3: Pongola, 4: Jozini, 5: Mtubatuba, 6: Empangeni, 7: Umhlali, 8: Mount Edgecombe and 9: Port Shepstone (KwaZulu-Natal). The sum of tests conducted for each variety within a site is shown in each block.	84
Table 5.2: The optimal range of environmental conditions required for pollen viability was compared to the actual mean pollen viability observed in each site in order to assess their relationship. Rows highlighted with the same colour indicate that the sites had the same sets of optimal environmental parameters. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).	91
Table 5.3: Eigenvalues and cumulative percentage variance contributed by the four axes on the PCA ordination.	94

CHAPTER 1: INTRODUCTION

1.1 Background and Rationale

Interest in sugarcane has grown considerably in recent years owing to its economic impact on sustainable energy production (Cheavegatti-Gianotto *et al.*, 2011). In addition to this significant contribution, sugarcane is also the world's leading sugar producing crop ahead of sugar beet, contributing approximately 80% of the world's sugar (Gao *et al.*, 2016; Gao *et al.*, 2014). It is cultivated in more than 90 countries and is a very important commercial crop of tropical and subtropical regions (Bonnett *et al.*, 2008; Mnisi & Dlamini, 2012). The world's largest sugarcane producing country is Brazil, where nearly half of the sugarcane crop is used to produce sugar and the remainder is used for ethanol production (Cheavegatti-Gianotto *et al.*, 2011).

Initial sugarcane production in South Africa was dependent on imported sugarcane varieties. The first sugar produced in 1852 was derived from an import of *Saccharum officinarum* L. The imported varieties were however not adapted to local environmental conditions and over time became susceptible to diseases such as smut and mosaic virus (Zhou, 2013). The South African Sugarcane Research Institute (SASRI) was established in 1925 with the aim of importing, testing and releasing new improved varieties from breeding programs. By 2012, more than 60 commercial hybrids had already been released from these breeding programs (Parfitt, 2005; Zhou, 2013). Improved agronomic traits resulting from the breeding programs, amongst others, included herbicide and insect resistance, vigorous stalk growth and high sucrose accumulation (Snyman & Meyer, 2012).

While numerous improvements have been made to commercial sugarcane cultivars through breeding, more opportunities to grow global competitiveness of the South African sugar industry exist with genetic modification (GM) as complementary to breeding (Snyman & Meyer, 2012). Additionally, due to the increasing industrial demand of sugarcane for energy and sugar production purposes, yield can only be increased in the next 30 years with the supplementary use of biotechnological tools (Cheavegatti-Gianotto *et al.*, 2011). Biotechnical advancements improve crop performance by introducing exotic genetic material and adjusting endogenous gene expression (Snyman & Meyer, 2012). In particular, GM sugarcane that incorporates transgenes that increase resistance to biotic and abiotic stress factors could play a major role in fulfilling this expectation (Cheavegatti-Gianotto *et al.*, 2011).

The assessment of the potential environmental impact of GM crops is a fundamental part of the international regulatory process undertaken before any GM crop can be grown under field conditions, either experimentally or commercially (Dale *et al.*, 2002; Mitchell, 2011). Gene flow leading to the escape of transgenes is the main concern associated with the introduction of GM crops (Gepts & Papa, 2003). In many cases, transgenes represent gains of function, potentially relieving wild relatives from constraints that limit their fitness (Gepts & Papa, 2003). Consequently, resistance to abiotic and biotic stress may be enhanced in weedy relatives, leading to the evolution of super weeds (Warwick *et al.*, 2009). This reduces the resilience and stability of an ecosystem, thereby posing a great threat to biodiversity.

According to Goodman and Gepts (2004), gene exchange usually occurs when the interacting individuals belong to the same or closely related biological species and the crop hybrids produce fertile pollen which is able to reach a receptive wild relative. Haygood *et al.* (2003) noted that pollen from cultivated crops frequently reach wild relative plants growing in close proximity, and when closely related (sharing common ancestors), hybridization usually follows. In fact, thirteen of the world's most important crops are known to have hybridized with their wild relatives (Cornille *et al.*, 2013; Ellstrand, 2003) somewhere in their agricultural ranges. Gene flow between seven of these crops and their wild relatives is implicated in the evolution of weediness (e.g. *Phaseolus vulgaris* L.) (Andow & Zwahlen, 2006; Ellstrand *et al.*, 1999). Sugarcane may have the potential to hybridize with its own wild relatives since Asante (2008) noted that every cultivated crop is related to a wild species that spontaneously occurs somewhere in its centre of origin.

Pollen is crucial for seed production and serves as the primary means of gene flow in outcrossing species such as the *Saccharum* hybrids (Ge *et al.*, 2011). However, there is little information available on the pollen viability from commercial sugarcane plantations, because seed is of no economic importance and is not propagated to produce the next crop (Parfitt, 2005). Therefore, to assess the likelihood of GM sugarcane outcrossing with a related plant, there first needs to be an understanding of the level of genetic relatedness between sugarcane and its wild relatives, and whether the current commercial sugarcane varieties do produce viable pollen during their flowering period. This together with additional data (e.g. genetic relatedness) would enable an assessment of the potential gene flow from the *Saccharum* species hybrids to wild related species.

1.2 Aim and objectives

The aim of this study was to:

- Assess the gene flow potential from sugarcane to wild relatives.

To achieve the aim of the study, the objectives were set to:

- Review pertinent literature to determine whether members of the subtribes (Saccharinae and Sorghinae) have spontaneously hybridized with sugarcane in the past;
- Sequence the barcode regions of the internal transcribed spacer (ITS) regions of the 5.8S ribosomal gene as well as the barcode regions of two chloroplast genes, ribulose-bisphosphate carboxylase (*rbcL*) and maturase K (*matK*) to determine relatedness between *Saccharum* and its wild relatives using phylogenetic analysis;
- Assess pollen viability of commercial sugarcane varieties cultivated in Mpumalanga and KwaZulu-Natal.

1.3 Profile of the sugar industry

South Africa is Africa's largest sugarcane producer and features in the top 15 countries with the highest sugarcane production globally (Mohlala *et al.*, 2016). Sugarcane cultivation extends from the latitudes 25° S to 31°S (Figure 1.1), where over 428 000 hectares are under sugarcane production (Snyman *et al.*, 2008). Approximately 2.5 million tons of sugar are produced seasonally from 14 sugar mills (Mohlala *et al.*, 2016; Snyman *et al.*, 2008). These mills are owned by Illovo Sugar Limited (five mills), Tongaat Hulett Sugar Limited (four mills), TSB Sugar RSA Limited (two mills), Umvoti Transport (Pty) Limited (one mill), Ushukela Milling (Pty) Limited (one mill) and UCL Company Limited (one mill) (Mohlala *et al.*, 2016). Commercial sugarcane production is confined to the KwaZulu-Natal and Mpumalanga provinces. Within these provinces, 85% is produced in rain-fed regions while 15% comes from irrigated fields (Snyman *et al.*, 2008). SASRI is responsible for agricultural research and contributes to the cost-effectiveness of the sugar industry while promoting sustainability in farming practices (Mnisi & Dlamini, 2012). With over 85 000 employees in cane production and processing (Mnisi & Dlamini, 2012), the industry generates an estimated R8 billion per annum direct income to the South African national economy (Mashoko *et al.*, 2010).

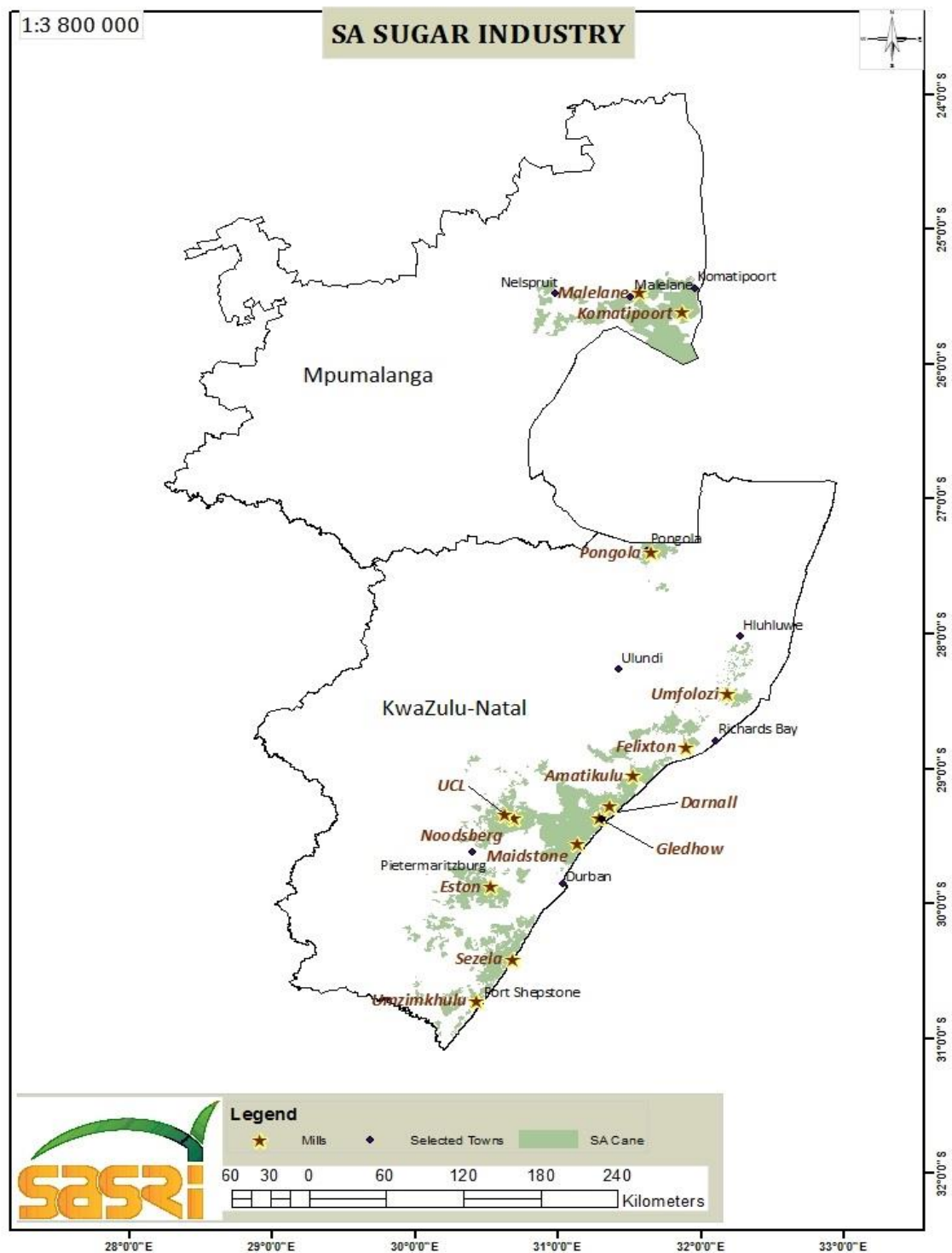


Figure 1.1: The South African sugarcane industry and mills located in the Mpumalanga and KwaZulu-Natal provinces. (Map produced by Mzimase Jalisa- SASRI GIS intern)

1.3 Layout and approach

This dissertation is divided into six chapters. The introduction (Chapter 1) provides the background to and rationale for the study. It highlights the sugar industry in South Africa and states the main aim and objectives of this study. The literature review (Chapter 2) considers the reproductive morphology of sugarcane and provides information regarding breeding programmes, leading to the need for biotechnology tools and the Biosafety laws guiding the production of GM crops. Chapters 3-5 present the results of the respective objectives. Each chapter first describes the technique used as an investigation method with the motivation thereof, the core findings and discussion of the results. Past hybridization of sugarcane with subtribe individuals (Chapter 3) presents the literature review results of the list of species that are reported to have hybridized with sugarcane. DNA barcoding for genetic relatedness of *Saccharum* species hybrids and wild relatives (Chapter 4) presents the phylogenetic analysis of relatedness in neighbor joining and parsimonious trees. Pollen viability of commercial sugarcane varieties (Chapter 5) presents the pollen viability percentages in relation to environmental gradients of temperature, soil water content, day length and relative humidity as possible explanatory variables. Synopsis and future prospects (Chapter 6) discusses the major findings and how the aims of the study have been met. It then concludes on the safety of GM cane and proposes further research.

CHAPTER 2: LITERATURE REVIEW

2.1 Origin and classification

Sugarcane is an important crop plant originating from South-east Asia and New Guinea (Cheavegatti-Gianotto *et al.*, 2011). It is classified in the grass family (Poaceae), subfamily Panicoideae, tribe Andropogoneae, and genus *Saccharum* L. (Dillon *et al.*, 2007; Hoang *et al.*, 2015). The genus name was derived from the Greek word “sakcharon” which means sugar and was first described by Linnaeus in 1753 (Cheavegatti-Gianotto *et al.*, 2011). The tribe consists of tropical and subtropical grass species, of which *Zea mays* L. and *Sorghum bicolor* L. Moench are the most cultivated members. According to Cheavegatti-Gianotto *et al.* (2011), there are numerous similarities in the genetic composition of *Saccharum* and *Sorghum*, as they are believed to have developed from a common ancestral lineage approximately 5 million years ago (Mya). In addition to this, Dillon *et al.* (2007) states that among cultivated crops, *Saccharum* and *Sorghum* are the closest relatives.

Closely related to *Saccharum* are another four genera (species previously classified under *Erianthus* section *Ripidum*, *Miscanthus* section *Diandra*, *Narenga* and *Sclerostachya*) that are known to interbreed readily, collectively forming the *Saccharum* complex (Cheavegatti-Gianotto *et al.*, 2011; Dillon *et al.*, 2007). The genus includes 36 species, of which *Saccharum spontaneum* L. and *Saccharum robustum* E.W.Brandes & Jeswiet ex Grassl are wild species of economic importance, while *Saccharum officinarum* L. and *Saccharum sinense* Roxb. are cultivated species (Cheavegatti-Gianotto *et al.*, 2011; Dillon *et al.*, 2007; Evans & Joshi, 2016). The modern sugarcane varieties cultivated throughout the world are interspecific hybrids from the *Saccharum* genus, mainly resulting from a cross between *Sa. officinarum*, widely known as noble cane, and *Sa. spontaneum* (Dillon *et al.*, 2007; Mohan, 2016). Due to this hybridization, *Sa. officinarum* contributes the ability to accumulate high sucrose levels in the stem, while *Sa. spontaneum* offers vigorous growth and resistance to environmental stress (Cheavegatti-Gianotto *et al.*, 2011; Zhou, 2013).

2.2 Morphology and developmental stages

2.2.1 Plant morphology

Sugarcane is a semi-perennial monocotyledonous plant that is cultivated mainly for its ability to store high concentrations of sucrose in the stem (Cheavegatti-Gianotto *et al.*, 2011). The unbranched stem is divided into joints comprising of a node and internode. Each node consists of a vegetative bud situated in the axil of the leaf, a band of root primordia, and a growth ring.

The internodes are made of sucrose storing parenchyma cells and vascular tissue (Moore & Botha, 2013). The high variability of stalk morphology (length, diameter, shape and colour) between genotypes is used for varietal characterization. Attached to the stem at the bases of nodes are leaves alternating on two opposite sides (Figure 2.1a). The leaves are hairy on the abaxial side and have a hard, white midrib. Leaves comprise of a lamina and a tabular shaped sheath that encircles the stem. Numbering of sugarcane leaves is from top to bottom, beginning with the uppermost leaf showing a visible dewlap as leaf +1 (Figure 2.1a) (Cheavegatti-Gianotto *et al.*, 2011). The root system consists of sett roots that develop from the root primordial band and adventitious roots (shoot roots) that develop on the base of the new shoot (Figure 2.1b). Sett roots are fine and highly branched roots, responsible for water and nutrient uptake within the first few days of planting before the shoot roots develop and take over (Smith *et al.*, 2005).

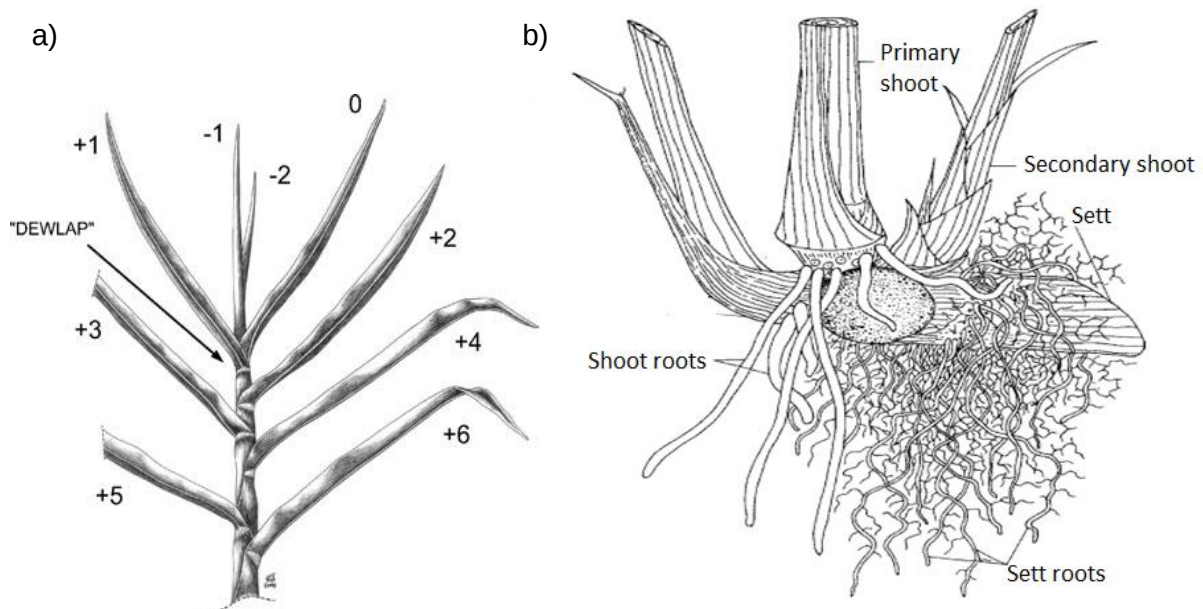


Figure 2.1: Morphology of sugarcane illustrated by a) the phylotaxy and numbering of the leaves (taken from Cheavegatti-Gianotto *et al.* 2011) and b) the root system indicating sett and shoot roots (taken from Smith *et al.*, 2005)

2.2.2 Reproductive morphology

The inflorescence (also known as a tassel or an arrow) is formed from the apical meristem when the sugarcane plant switches from a vegetative to a reproductive stage. This is usually caused by reduced growth due to shortening day lengths and cooler night temperatures (James, 2004). The continuation of the last stalk internode forms the rachis (anthoclinium), which is the main axis of the panicle (Figure 2.2a). It divides into secondary branches which

in turn hold the tertiary branches (Figure 2.2a) (Cheavegatti-Gianotto *et al.*, 2011). Attached to the branches are spikelets with individual hermaphroditic flowers. The pistil contains a purple coloured, single carpel with feathery stigma (Figure 2.2d), which explains the purple appearance of the panicles. The androecium has three stamens each with one pollen producing anther (Figure 2.2e) (Amaral *et al.*, 2013). Indehiscent anthers are usually yellow, while dehiscent anthers are brown or purple, producing yellow pollen grains (Figure 2.3.a) (James, 2004). Rings of silky and colourless trichomes arise at the base of each spikelet and are responsible for dispersal. Panicle maturation begins at the apex, progressing downwards and inwards, thus anther dehiscence only occurs in one third of the panicle at a time (Figure 2.3b) (Cheavegatti-Gianotto *et al.*, 2011). After flowering is completed, the secondary and tertiary branches contract, facing upwards (Figure 2.3c). Remnants of the inflorescence, caryopses, glumes and callus hairs are collectively known as the fuzz (Figure 2.3d) at the time of seed production.

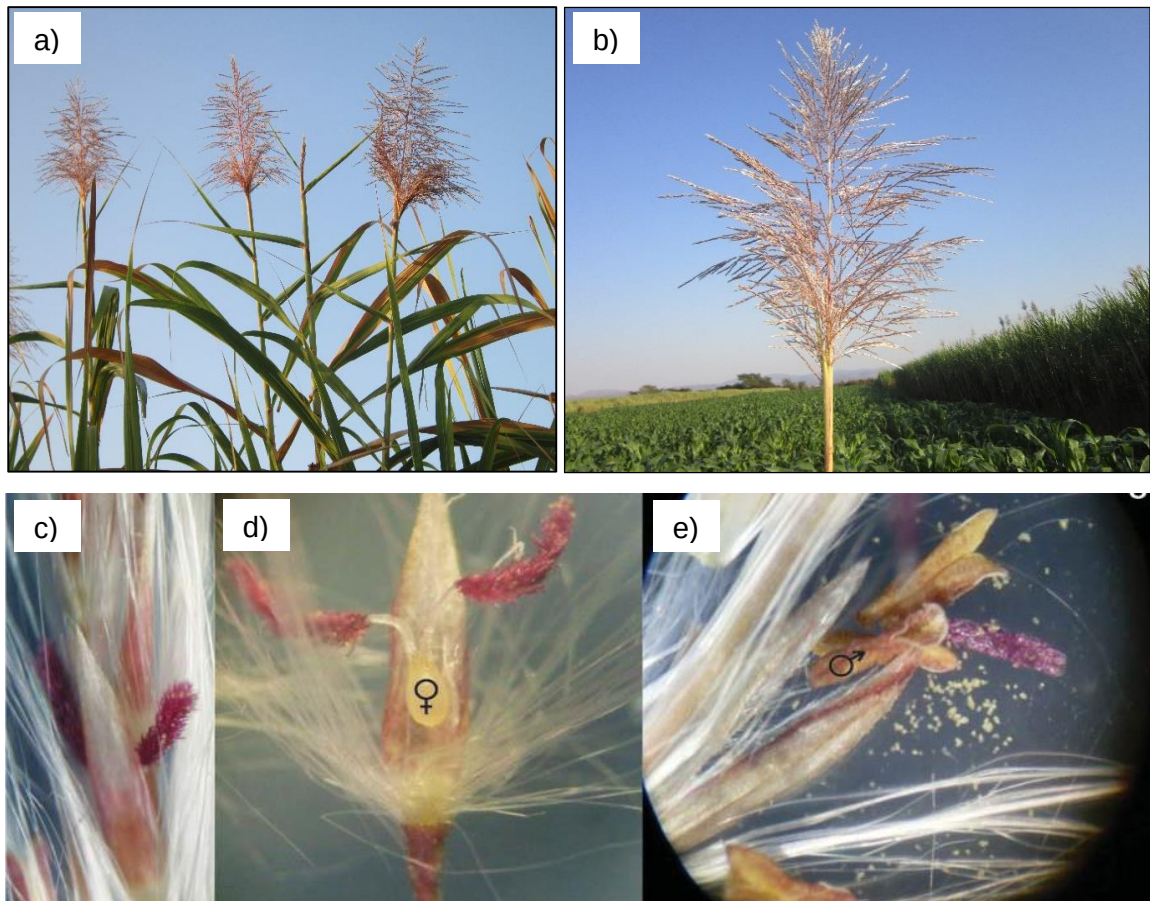


Figure 2.2: Reproductive morphology of a sugarcane crop. a) Rachis formed from the apex of the meristem (Photo: H. Khanyi), b) the panicle carrying secondary and tertiary branches of the inflorescence (Photo: H. Khanyi), c) sessile spikelets, d) purple and fluffy stigmas on the gynoecium and e) anthers in the androecium (Photos: Amaral *et al.*, 2013)

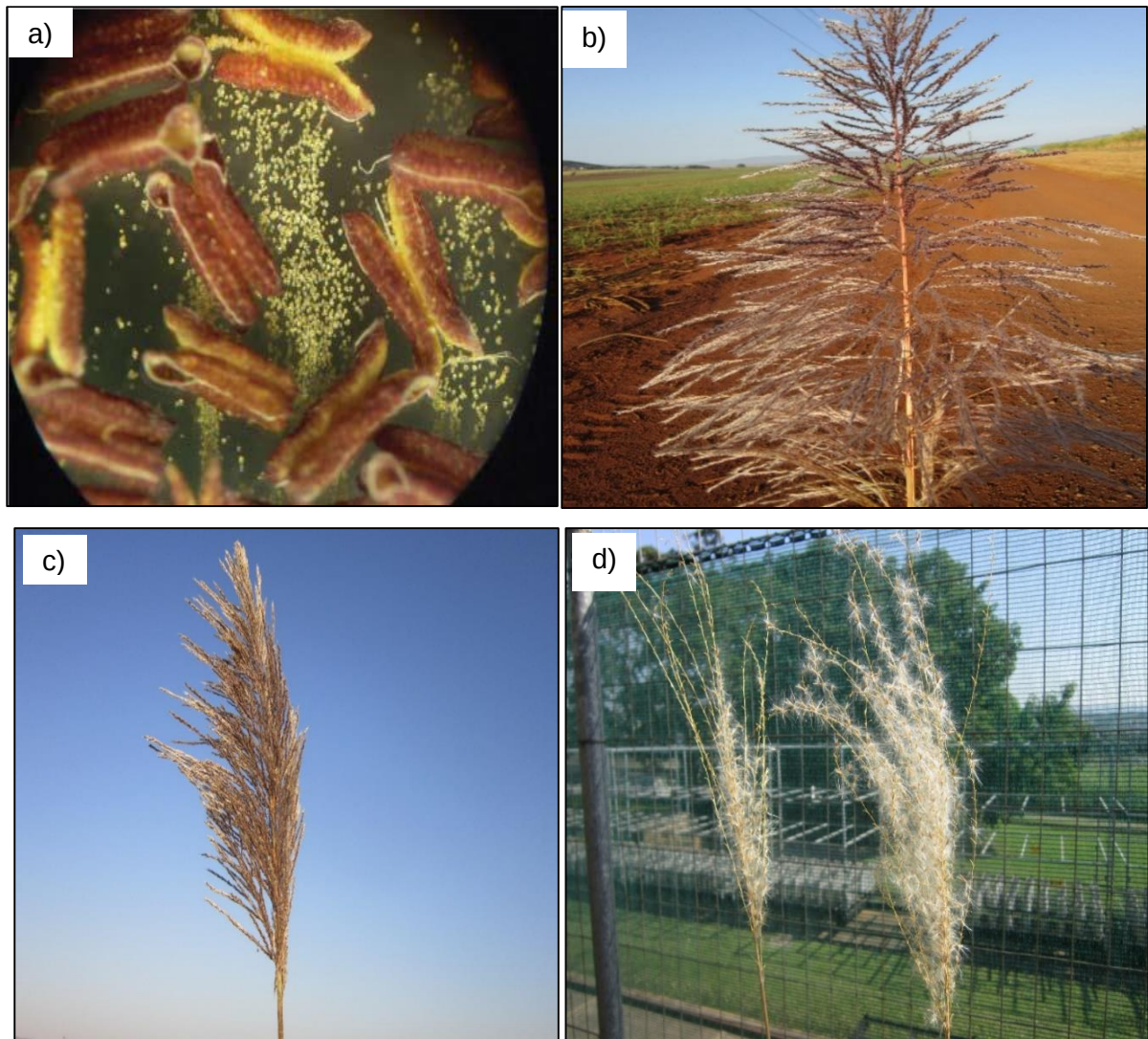


Figure 2.3: Reproductive morphology of a sugarcane crop. a) Mature anthers with pollen grains (Photo: Amaral *et al.*, 2013), b) flowering of the sugarcane inflorescence on the first two thirds with no flowering at the base (Photo: H. Khanyi), c) post flowering of the inflorescence with branches closed (Photo: H. Khanyi), and d) the formation of fuzz and subsequent seed production (Photo: D.M. Komape).

2.2.3 Environmental gradients affecting flowering and pollen viability

Low soil water content delays panicle development (Melloni *et al.*, 2015). For instance, Moore & Botha (2013) recorded the widest variance in sugarcane flowering from mean flowering date when rainfall was significantly lower. Since the panicle development is signalled by a gradual reduction in photoperiod, Bonnett *et al.* (2010) documented latitudes between 5° and 15° as suitable for flowering. An ideal situation for flowering and pollen viability needs to have a daily photoperiod ranging from 12 to 12.5 hours with a gradual reduction over 10-15 days during flower initiation period (Amaral *et al.*, 2013). The optimal daily temperatures lie between 18°C and 32°C, with temperatures outside this range being detrimental to panicle development and

pollen viability. To emphasize the role of optimum temperature, Brett *et al.* (1951 cited by Moore & Botha, 2013) obtained 158 seedlings under cool, natural temperature as opposed to 1574 seedlings under warmer, artificial temperature, from the same number of tassels of sugarcane cultivar NCo310. Temperature and photoperiod are correlated with latitude, thus flowering and pollen viability is expected to decrease with increasing latitude (Bonnett *et al.*, 2010).

Moore & Botha (2013) found that photoreceptors in the leaf blades trigger a flowering hormone that is transported into the shoot apical meristem to reprogram the formation of inflorescence instead of phytomers. Consequently, successive sheaths become more elongated, leaf blades shorten, leaf production by the apical meristem is stopped and inflorescence primordia develop around three months prior to the emergence of the panicle (Cheavegatti-Gianotto *et al.*, 2011). Relative humidity plays a major role from panicle development through to seed formation. When the relative air humidity is low during pollen development, an imbalance in osmotic pressure causes pollen grains to rupture and abort its contents. Thus, pollen grains are expected to be spherically shaped when viable and prismatic when sterile (Cheavegatti-Gianotto *et al.*, 2011). Flower opening (anthesis) (Figure 2.4) occurs a few hours before sunrise, when the plant is hydrated and relative air humidity is still at the optimum to maintain viability. From then, pollen grains have an estimated half-life of approximately 12 minutes and rapidly dry after dehiscence. They do not survive more than 35 minutes at a temperature greater than 26.5°C and relative air humidity lower than 67% (Amaral *et al.*, 2013).

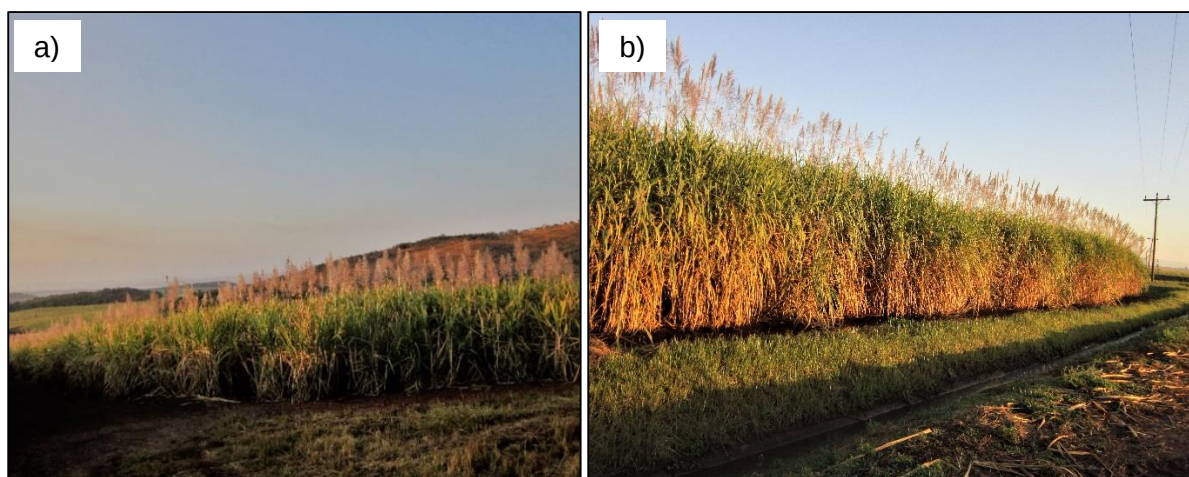


Figure 2.4: Flowering of sugarcane varieties a) N36 and b) N23 at Malelane, Mpumalanga (05 July 2017) (Photos: H. Khanyi)

2.2.4 Physiological and morphological changes throughout plant development

Ten (from 0-9) developmental stages have been identified for sugarcane (Moore & Botha, 2013) (Figure 2.5):

(0) Either a true seed (although rare), culm pieces containing auxiliary buds or underground auxiliary buds remaining after harvesting are used to cultivate new plants. In the latter cases, sett roots develop around 24 h after planting, followed by shoot roots at about 5-7 days after planting (Smith *et al.*, 2005).

(1) A node to which a leaf is attached, internode and axillary bud at the base of the first internode develop from phytomers (Moore & Botha, 2013). This repeated vegetative growth increases the number of leaves and length of the plant.

(2) Side shoots (primary tillers) emerge from the auxiliary bud to form additional culms, while secondary tillers are produced from primary tillers (Moore & Botha, 2013).

(3) Once the leaf at a base of each internode has fully stretched, intercalary meristem produces expanding cells, thus elongating the internodes. The elongation process proceeds and is only completed when the next four leaves have fully extended.

(4) While the new internodes are elongating at the top of the culm, sucrose accumulation begins from the basal internodes (Moore & Botha, 2013).

(5) The apex transitions from vegetative to reproductive growth by changing seasonal signals.

(6) Spikelets start to flower, either while the inflorescence is still emerging or only after it has fully emerged. Although anthers protrude from the spikelet during the night, they only dehisce at dawn when humidity drops.

(7) The few remaining anthers, stigmas and newly formed seed make up the fuzz (Moore & Botha, 2013).

(8) Mature spikelets break off from the branches of the inflorescence and the hair at the base aid in wind dispersal.

(9) After harvest at 12-18 months, small clumps are left behind as part of the ratoon crop. Old leaves from base of the stem dehisce throughout the developmental stages (Cheavegatti-Gianotto *et al.*, 2011).

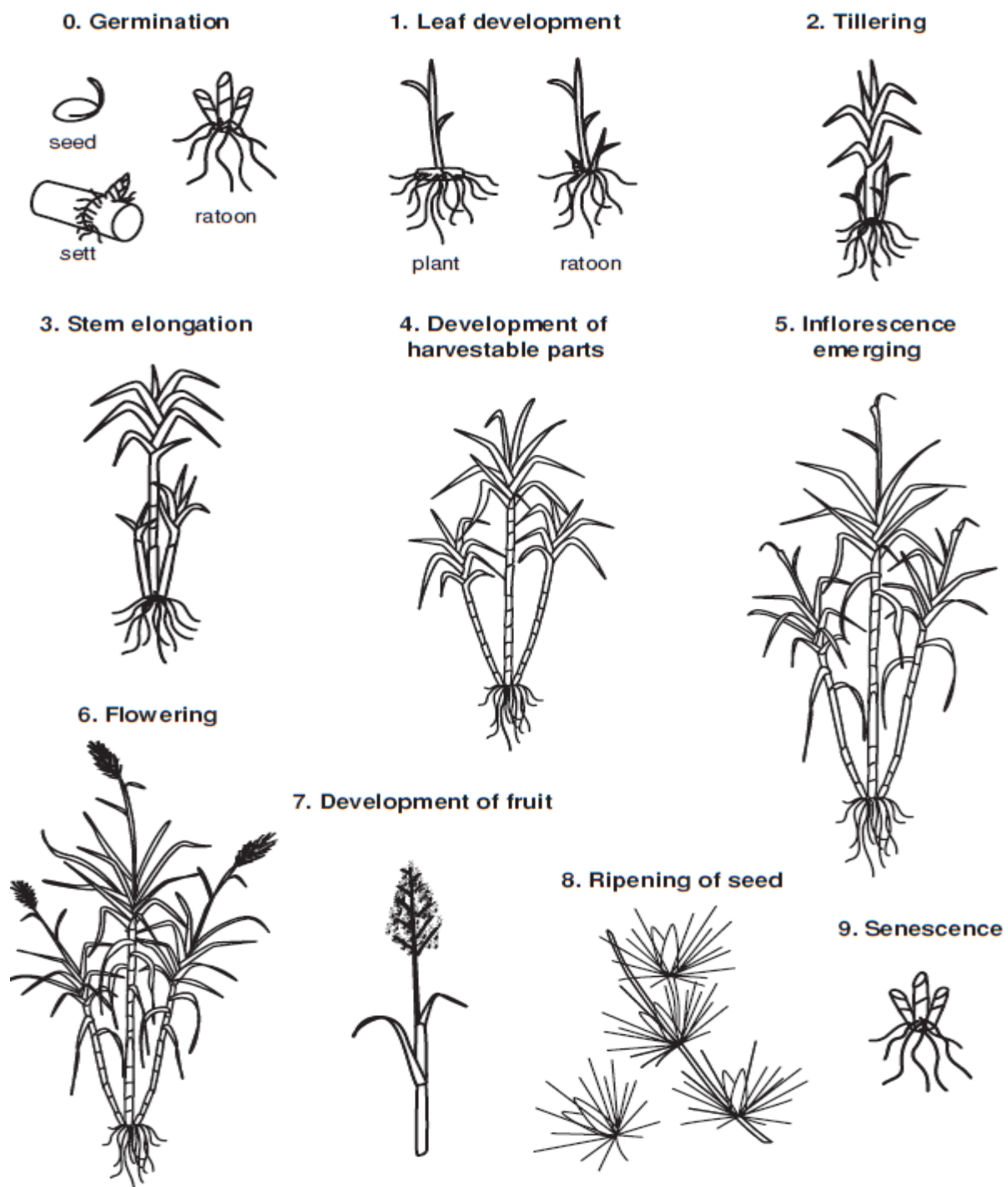


Figure 2.5: Developmental stages of sugarcane from planting to post harvest. (Taken from Moore & Botha, 2013).

2.3 Uses of sugarcane

2.3.1 Sugar production

In its natural form, unprocessed sugarcane produces a nutritious energy juice that is consumed by humans. However, since the primary objective of sugarcane cultivation is sugar production, industrial processing produces highly purified sugar through several phases of crushing, purification, concentrating and crystallization. The raw sugar obtained contains 99.8% sucrose (Cheavegatti-Gianotto *et al.*, 2011). According to Guo *et al.* (2015) and Zhou *et al.* (2016), sugarcane accounts for 92% of all sugar produced in China and 80% of that in the world. In South Africa, 60% of the sugar produced from cane is sold to the Southern African Customs Union, while the remainder is exported to Africa, Asia and the Middle East (Mnisi & Dlamini, 2012).

2.3.2 By-products: Fuel, energy and fertiliser

One of the world's most successful biofuel production systems is the bioethanol that is produced from sugarcane (Hoang *et al.*, 2015). In Brazil, approximately 23.4 billion litres of ethanol was produced in 2014. This product is obtained through the fermentation and distillation of sugar, and can be used directly as transportation fuel or dehydrated to produce anhydrous ethanol (Cheavegatti-Gianotto *et al.*, 2011). Bagasse is one by-product from sugar and ethanol production, and makes a significant contribution to energy supply. As a C4 plant, sugarcane is efficient in converting solar energy into stored chemical energy and biomass accumulation. In 2009, sugarcane bagasse contributed about 15% of the total electricity consumed in Brazil, and is projected to supply double this amount by 2020, which will be equal or greater than the electricity produced from hydropower (Hoang *et al.*, 2015). Cheavegatti-Gianotto *et al.* (2011) added that the high efficiency of this process is evident from the provision of electricity to nearby cities from the excess electric energy generated by the sugar mills. Even though this is unquantified, South Africa also generates electricity and produces industrial ethanol from bagasse and molasses (Mohlala *et al.*, 2016). Vinasse and filter cake are other residues of industrial sugarcane processing that are rich in minerals, thus are used as fertilisers or animal feed.

2.4 Overview of sugarcane breeding in South Africa

Breeding involves manipulating traits of a plant species in order to introduce desired characteristics. Sugarcane breeding programmes aim to develop commercial varieties with improved genetic makeup to positively contribute and sustain a competitive sugar industry (Bischoff & Gravois, 2004).

The cycle begins with the selection of parental genotypes based on desired characteristics and the collection of germplasm where flowering of the parental lines has been synchronised (Cheavegatti-Gianotto *et al.*, 2011). Crossing is achieved in a glasshouse by placing flowers of the female clone below the taller male clone flowers. Seeds obtained from these crosses are sown in trays and kept in a glasshouse from which the resultant seedlings are distributed across the agro-ecological regions of sugarcane production in South Africa i.e. irrigated, coastal and high altitude (Midlands) (Zhou, 2013). Field testing and selection of potential new varieties from experimental sugarcane clones is based on high sucrose content, agronomic characteristics, environmental adaptability, and resistance to diseases and pests (Hoang *et al.*, 2015). This takes 11-15 years through a five-stage selection programme (refer to Table 2.1) whereby the number of varieties are significantly reduced at each stage, and the survivors that meet the criteria are transferred to larger plots to be screened further (Cheavegatti-Gianotto *et al.*, 2011; Parfitt, 2005; Zhou, 2013)

Table 2.1: Selection stages of a new sugarcane variety by breeding programmes at SASRI (Taken from Parfitt, 2005)

Selection stage	No. varieties per station and total	Trial design	No reps	No crops	Selection rate (%)	Selection criteria	
Seedlings NURSERY	50 000 x 5 250 000	Potted seedlings	Family	1	70	Visual assessment of vigour. Freedom from diseases	
Stage 1 SINGLE STOOLS	35 000 x 5 175 000	Replication of families		1	11	Mean sucrose per family.	
Stage 2 SINGLE LINES	4 000 x 5 20 000	8 m row		2	10	Sucrose yield – plant and (estimated) ratoon Family performance from cross summary	Acceptable sucrose content Freedom from naturally occurring diseases and pests, and from agronomic defects
Stage 3 OBSERVATION TRIAL	400 x 5 2 000	2 x 8 m row	2	1	15	Plant cane sucrose yield Northern irrigated region – smut screening trial	
Stage 4 PRIMARY VARIETY TRIAL	60 x 5 300	5 x 8 m row (gross plot)	3	3	30	Plant cane sucrose yield (reselect in second ratoon) Smut/mosaic/eldana screening trials	
Stage 5 SECONDARY VARIETY TRIAL	20 x 5 100	As above - 3 off-station trials per region	3	3	5	Compilation of all yield and pest/disease results Yield stability or specific adaptability	
FINAL PROPAGATION	1 – 2 5	0.8 ha				Grower acceptance Resource Optimisation establish further trials	

From the initial 250 000 seedlings at the first stage to one or two hybrids that survive through the selection process to the final propagation stage (Zhou, 2013), the Variety Release Committee (VRC) decides on the approval of these varieties for bulking on co-operator farms. The Local Pest Disease and Variety Control Committee (LPD and VCC) will thereafter be informed of the new varieties approved for their area, from which they decide whether or not to bulk these varieties. While the varieties are in bulking, entomologists and pathologists inspect these varieties and deliver their results to the VRC for quality check, from which the

LPD and VCC then takes final decision on releasing the variety for commercial propagation (Bischoff & Gravois, 2004; Parfitt, 2005). Since the release of the very first varieties, NCo310 in 1945 and NCo376 in 1955 (Zhou, 2013), SASRI has maintained global competitiveness by selecting for increased sucrose yield and ratoon crops. Further contributing to this success, are the recently adopted 'N' varieties with increased disease resistance (Parfitt, 2005).

2.5 Challenges with sugarcane breeding

Artificial crossing is usually done in glasshouses with controlled environmental conditions due to various reproductive challenges e.g. flowering does not occur naturally thus has to be induced (Bischoff & Gravois, 2004), flowering in South Africa occurs during winter when minimum temperatures drop below 20°C thus promoting pollen sterility (Parfitt, 2005; Zhou, 2013), and flowering can be asynchronous (Melloni *et al.*, 2015). Variety development is a time-consuming, multi-disciplinary and costly process because of the high polyploidy (eight homology groups) and complexity in the sugarcane genome that brings a great challenge to genetic improvement (Dillon *et al.*, 2007). The varying set of chromosomes ($2n$: 80-140) and distinct alleles at each chromosome locus make the characteristics of offspring unpredictable, and thus demand evaluation of thousands of lines from numerous parents (Hoang *et al.*, 2015). The allopolyploid and aneuploid genetic background of sugarcane has made it very difficult to obtain varieties with high biomass, high sugar content and excellent pest and disease resistance solely from traditional breeding (Xue *et al.*, 2014). Mohan (2016) added that introgressive hybridization of multiple traits is impossible through traditional breeding, whereas the adoption and use of transgenic technology could make this achievable.

2.6 Genetically Modified sugarcane

2.6.1 Approaching biotechnology tools

Unlike traditional breeding whereby a plant acquires genes under natural conditions, genetic modification (GM) artificially inserts foreign transgene obtained either from the same or a different species (Tripathi, 2005). As a supplement to the breeding techniques, GM technology has the potential to deliver stable inheritance of traits, reduced cost and a shortened breeding period (Zhou *et al.*, 2016). Sugar industries globally are now investing in the development of transgenic sugarcane in order to adjust to the expanding sugar and biofuel market (Joyce *et al.*, 2014; Snyman & Meyer, 2012). One of the advantages that makes sugarcane an ideal candidate for genetic engineering lies in its limited ability to flower and produce viable seed naturally (Xue *et al.*, 2014), thus reducing the potential for genetic drift.

2.6.2 Progress with transgenic sugarcane

Much progress has been made with the biotechnological manipulation of sugarcane since the 1990s (van der Vyver *et al.*, 2013), especially with herbicide resistance and pest and disease resistance (Bonnett *et al.*, 2007). Worldwide, genetic modification of food crops has produced traits for herbicide tolerance (71%), insect resistance (28%) and quality (1%) (Tripathi, 2005). Genetically modified sugarcane has also been reported on, with the insertion of genes such as *pat* (Van der Vyver *et al.*, 2013) and *bar* (Zhou *et al.*, 2016) for herbicide resistance, and *cry1Ac* (Gao *et al.*, 2016; Ismail, 2013; Snyman & Meyer, 2012) and *cry1Ab* (Arencibia *et al.*, 1997) for insect resistance, and coat protein (Guo *et al.*, 2015; Yao *et al.*, 2017) for disease resistance.

Progress on herbicide-resistant GM sugarcane in the South African sugar industry was confirmed by stable gene expression over several ratoons in the field (Snyman & Meyer, 2012). GM projects at SASRI are ongoing and aimed at improving sucrose metabolism, drought tolerance and efficiency in nitrogen use (Snyman & Meyer, 2012). The advances of sugarcane through GM technology at SASRI are however not yet commercialised (James, 2016). The first GM sugarcane was released for commercial cultivation in 2013 by Indonesia (Xue *et al.*, 2014), with a drought tolerance transgene, while Brazil has projected that GM sugarcane will be on the market by 2017 (James, 2016).

2.7 International and national status of GM crops

The first global commercial cultivation of genetically modified (GM) crops (maize) was in 1996 (Snyman & Meyer, 2012). By 2016, commercial GM crops were cultivated over 185 million hectares in 26 countries by 18 million farmers (James, 2016). Globally, the four major crops genetically modified for herbicide and/or insect resistance are soybean, cotton, maize and canola (Mitchell, 2011). South Africa planted its first GM crop in 1998 with insect resistant cotton, followed by insect resistant maize and herbicide tolerant soybean in 2001 (James, 2016). South Africa cultivates two insect resistant and herbicide tolerant crops, i.e. cotton (*Gossypium hirsutum* L.) and maize (*Zea mays* L.), as well as herbicide tolerant soybean (*Glycine max* (L.) Merr.). In 2016 GM crops covered over 2.66 million hectares in SA, which comprised of maize (81.2%), soybean (18.6%) and cotton (0.2%) (James, 2016). Field trials with drought-tolerant maize are currently under way (Smyth, 2017).

2.8 Benefits of GM crops

Risk assessments of GM crops across 30 different countries have not detected any increased threats to humans, animals or the environment over the past 20 years (Smyth, 2017). Despite

this, environmental non-governmental organizations (eNGOs) continue to spread rumours about damages and risks posed by GM crops. Despite the negative publicity, GMOs have instead shown great potential in improving food production, crop quality and environmental aspects (Bothma *et al.*, 2010). Over a period of 18 years, close estimates of economic benefits amounted to US\$150 billion, with reduced pesticide poisoning, less time spent on manually weeding fields and a reduction in chemical based application to crops (Smyth, 2017). For instance, Bothma *et al.* (2010) reported a decrease by 53% of chemicals used on the herbicide tolerant canola. In addition to this, studies on GM crop yield reported 74% increased yields of canola due to reduced insect and weed population pressure in 2010.

South Africa has made progress with the development of GMOs since 1998 when the first GMO crop was approved for commercialization. In 2012, the commercialized GM maize, GM soybean and Bt cotton brought economic gains amounting to approximately R930 billion (Wafula *et al.*, 2012). Bt insect and herbicide tolerant maize is said to have decreased levels of cancer causing agents compared to conventional and organic maize. Consequently, no undesired effects to human health have been reported by consumers. The use of insecticides and herbicides decreased by 33%, thus reducing the amount of carbon dioxide that gets released into the atmosphere from fuel usage to spray crops (Bothma *et al.*, 2010). Smyth (2017) added that the adoption of GM corn decreased the requirement for hand weeding.

2.9 Gene flow

Since the cultivation of GM crops was commercialised in 1996, the potential contamination of non-GM crops and wild or weedy relatives growing outside cultivated areas by transgenes from GM crops became the main environmental concern (Cornille *et al.*, 2013; Goodman & Gepts, 2004; Haygood *et al.*, 2003; Rieben *et al.*, 2011; Sanchez *et al.*, 2016; Snow, 2002; Warwick *et al.*, 2009). Although Haygood *et al.* (2003) noted that crop to wild relative gene flow has been occurring from ancient times, Snow (2002) stated that GM technology intensifies the concern since it would not only introduce genes that confer fitness-related traits but also extend the introduction of these novel genes into many diverse crops. Haygood *et al.* (2003) added that consequences of gene flow can be problematic regardless of whether transgenes are involved. Stewart *et al.* (2003) stated that it is too simplistic to declare transgene introgression a threat to the environment without exploring the trait conferred by the transgene, the crop and the cropping system.

In an event of transgene escape and introgression, new phenotypic traits such as resistance to insects, diseases, herbicides and abiotic stress could potentially be introduced to a new population and lead to the development of super weeds (Sanchez *et al.*, 2016; Snow, 2002).

Ellstrand *et al.* (1999) investigated the potential of introgression of genes from 13 important crops (based on area harvested) to their wild relatives. They found that gene flow to wild relatives is implicated in the evolution of more aggressive weeds for seven of the world's 13 most important crops that have been proven to hybridize with their wild relatives. Crop to wild relative gene flow thus could have been more frequent than initially perceived based on the assumption that domesticated traits possibly reduce fitness in the natural environment (Cornille *et al.*, 2013; Haygood *et al.*, 2003).

Gene flow can occur via seed and pollen dispersal. Understanding pollen mediated gene flow is more crucial to ensuring co-existence of GM and non-GM crops without gene flow, since pollen does not only transfer genes within and between populations, but also between species (Rieben *et al.*, 2011). Amongst various factors that affect hybridization frequency between GM crops and wild relatives, Goodman & Gepts (2004) and Warwick *et al.* (2009) stated that individuals first need to belong to the same biological or closely related species (with sexual compatibility) in order to exchange genes. Furthermore, close spatial arrangement, overlapping flowering periods and availability of pollen vectors are needed to facilitate gene exchange (Rieben *et al.*, 2011).

2.10 Cases of transgene escape

Referring to the worldwide GM contamination register of incidents associated with GMOs between the period 1997-2013, Price & Cotter (2014) investigated the frequency of GM contamination events, crops responsible and countries linked with these events. They found 396 incidents of GM contamination across 63 different countries. The five countries with the highest incidences were Germany, USA, France, United Kingdom and The Netherlands. Rice, despite not being commercialised, accounted for 34% of these incidents, followed by maize at 25%, then canola and soybean, each at approximately 10% (Price & Cotter, 2014). While Stewart *et al.* (2003) placed an emphasis on the need to differentiate between hybridization and introgression of the transgene where gene flow from GMOs is concerned, Price and Cotter (2014) did not supply any evidence of introgression in their review. Ellstrand *et al.* (1999), in their assessment of gene flow and introgression from domesticated crops to their wild relatives, confirmed putative transgene introgression for seven crops. These crops were wheat (*Triticum aestivum* L.), rice (*Oryza sativa* L.), sunflower (*Helianthus annuus* L.), soybean (*Glycine max* (L.) Merr.), sorghum (*Sorghum bicolor* (L.) Moench), millet (*Pennisetum glaucum* (L.) R.Br.), and beans (*Phaseolus vulgaris* L.).

2.11 Biosafety

GM sugarcane, just like any other genetically modified organism, must be approved by regulatory authorities before it can be released for commercialisation. All GM activities in South Africa are regulated under the GMO Act 15 of 1997, revised in 2006 to incorporate elements of the Cartagena Protocol on Biosafety (Snyman & Meyer, 2012). The protocol lays emphasis on the evaluation of scientific evidence in risk assessment and making decisions in the best interest of environmental, human and animal health (Bothma *et al.*, 2010). Established under the Convention of Biological Diversity (CBD), the Cartagena Protocol on Biosafety aims to ensure the safe transfer, handling and use of living modified organisms resulting from modern biotechnology (Wafula *et al.*, 2012).

The GMO act legislates a body that consists of the registrar, two regulatory bodies and inspectors (Bothma *et al.*, 2010). The first regulatory body is the Advisory Committee (AC) comprised of scientists that conduct research to assess the risk of proposed GM activities. The second regulatory body is the Executive Council (EC) represented by government officials to decide on the approval of GMOs based on the scientific recommendations and socio-economic factors. The registrar then issues a permit to regulate approved GMO activities. The Department of Agriculture, Forestry and Fisheries (DAFF) monitors compliance to permit regulations prior the release of a GMO while post commercial monitoring is assigned to the South African National Biodiversity Institute (SANBI).

In fulfilling the requirements of the legislature, Snyman & Meyer (2012) confirmed that SASRI laboratories and glasshouses where GM research is conducted, acquired facility registration from DAFF. Permits for field trials are as compulsory and the sites are routinely inspected by DAFF officials. When the commercialisation of sugarcane phase is reached, a General Release Permit has to be obtained. In addition to this, SASRI scientists have compiled baseline documents to demonstrate the practical equivalence of all commercial GM sugarcane against non-GM varieties.

2.12 Environmental risk analysis and regulation

Risk assessment encompasses risk assessment, risk management and risk communication (Jansen van Rijssen *et al.*, 2015; Johnson *et al.*, 2007). The assessment identifies the probability of an adverse effect towards human health and the environment, its magnitude and consequential hazards. For GMOs, scientific risk assessments are required by law (Cartagena Protocol on Biosafety) to regulate activities that can potentially harm the environment (Johnson *et al.*, 2007). As a precautionary measure, GMOs are established under pre-market

regulatory requirements and risks are inspected at the different steps of development and production (Jansen van Rijssen *et al.*, 2015). In order to regulate potential gene flow from GM sugarcane, an understanding is required of the ecology of related species, the degree of relatedness, hybridization history of sugarcane with other species, pollen viability of commercial varieties (investigated by this study), and flowering synchronicity with wild related species (supplied as Addendum A). (Chandler & Dunwell, 2008; Nickson, 2008). Outcomes of such a study would then indicate whether risk assessment studies are needed and highlight aspects to investigate.

CHAPTER 3: SPECIES REPORTED TO HAVE HYBRIDIZED WITH *SACCHARUM*

3.1 Introduction

Gene flow is a two-step process involving pollination of sexually compatible species followed by subsequent introgression of the trait into the pollinated species. Spontaneous hybridization between crops and related species could transfer crop genes conferring fitness-enhancing traits to the receiving populations (Snow *et al.*, 2001). While transgenic crop breeding is intended to benefit crop production, the resulting potential economic and environmental damage could overshadow these benefits (Campbell *et al.*, 2009). However, little is known about how easily this takes place in various weed-crop complexes. Introgression of a crop transgene is a dynamic process that requires multiple years and successive generations before the transgene can be fixed into the genetic background of hybrids (Stewart *et al.*, 2003). Thus, first generation (F1) hybrids must persist for at least one generation and be adequately fertile to produce backcross generations. Artificial crossing with an aim of introgressing desirable traits of wild relatives into a crop species could be viewed similarly to the process of spontaneous selection of transgenes that could unintentionally be transferred from crops into wild populations (Stewart *et al.*, 2003). Therefore, one of the aims of this study was to review literature around hybridization of *Saccharum* species to identify species likely to spontaneously hybridize with sugarcane in the study area.

3.2 Methodology

A systematic literature review was conducted as per the guidelines provided in Khan *et al.* (2003). Several questions were formulated to assess the likelihood of spontaneous hybridization involving commercial sugarcane cultivars. These were:

- i) Has spontaneous hybridisation between cultivated sugarcane and wild species been reported?
- ii) If not spontaneous, which artificial crosses have been achieved?
- iii) What is the survival rate of the seedlings generated from artificial crosses?
- iv) Were offspring in both cases confirmed by molecular markers to be true hybrids?

Answers to these questions would allow us to assess the chances of sugarcane passing on its genes should all barriers of gene flow be crossed, and assess the probability of new hybrid populations establishing in the environment.

Various online databases of peer reviewed journals (Jstor, ScienceDirect, Scopus) were searched using keywords “sugarcane”, “hybrids”, “gene flow”, “*Saccharum* hybridization” and word chains like “sugarcane spontaneous hybridization”, “natural hybrids with *Saccharum* parent”, “sugarcane outcrosses with related species”, “offspring with sugarcane genes”, “successful introgression between *Saccharum* species” and “wild related species”. Google searches were also conducted using word phrases (“spontaneous/natural hybridization from *Saccharum* to wild species”). In both the word chain and word phrase approaches, the title followed by abstracts of hit articles were examined to check for relevance and irrelevant articles were excluded. The selected articles were those reporting on crosses involving a *Saccharum* species. A manual scan of references from the relevant articles was conducted to source additional sources. Articles that did not report on hybridization, but that only tested the efficiency of markers in detecting *Saccharum* hybrids were also excluded as their input plants were putative hybrids from an unspecified study with no details on the crosses. A total of 22 relevant studies were sourced.

Study designs that reported on the original hybridization and generation of offspring were of importance. Secondary sources and reviews of the original publications were also used where the latter could not be obtained. The original author who conducted the crosses was cited to avoid reporting the same hybridization incident obtained from two sources as different incidents. The findings from literature were then summarised in a table format.

3.3 Results

3.3.1 Type of hybridization involving *Saccharum*

Literature from published work (Government gazettes, OECD documents and peer reviewed articles) recorded 36 incidents of efforts towards hybridization (Table 3.3.1). Notably from these hybridization incidents, there were only three claims of spontaneous hybridization reported and none involved cultivated sugarcane (Ellstrand *et al.*, 1999; Parthasarathy, 1948). The first natural crosses were implicated in the origin of *Saccharum maximum* (Brongn.) Trin. (synonym: *Erianthus maximus* Brongn.) and other unspecified *Saccharum* cultivars in India (Ellstrand *et al.*, 1999). The last cross produced *Saccharum sinense* L., from a natural cross involving *Saccharum officinarum* L. and *Saccharum spontaneum* L. (Parthasarathy, 1948). All

33 of the other crosses were artificially conducted under controlled laboratory conditions through experimental procedures that maximise the rate of pollination and seedling survival.

3.3.2 Parental lines in artificial crosses with *Saccharum*

Commercial sugarcane cultivars (*Saccharum* spp. hybrids) were utilised as parental lines in 59.46% of the artificial crosses, followed by *Sa. officinarum* L. (29.73%), *Sa. spontaneum* L. (8.11%) and *Sa. robustum* E.W.Brandes & Jeswiet ex Grassl (2.7%) respectively. The latter parental lines (*Sa. officinarum*, *Sa. spontaneum* and *Sa. robustum*) will collectively be referred to as *Saccharum* species from here onwards, unless specific to the individual cross. *Saccharum* species and *Saccharum* spp. hybrids were utilised as female parental lines in 25 (71.43%) out of the 35 incidents, while for six cases the sex of neither parents was specified.

3.3.3 Motives for crosses

Artificial crosses were performed to integrate the beneficial traits of one species with another. These included breeding to enhance agronomic traits such as vigorous growth, good ratoonability and biomass accumulation (Aitken *et al.*, 2006; Brett, 1954; Gao *et al.*, 2014; Piperidis *et al.*, 2000). Additionally, cane improvement by increasing the resistance to biotic and abiotic stresses led to the search for genes that could confer such traits from related wild species (Piperidis *et al.*, 2000). For instance, *Saccharum longisetosum* (Andersson) V.Naray. ex Bor. (synonym *Erianthus rockii* Keng) is known to be drought and cold tolerant (Aitken *et al.*, 2006). *Saccharum arundinaceum* Retz. (synonym *Erianthus arundinaceus* (Retz.) Jeswiet) on the other hand, possesses great agronomic traits such as better ratooning ability, vigour, tolerance to environmental stresses and disease resistance, thus appearing to be a good source for germplasm (Cai *et al.*, 2005a; D'Hont *et al.*, 1995).

3.3.4 Artificial crosses achieved and seedling survival rate

Crosses have been made since 1935 (Bourne, 1935) up to as recently as 2014 (Gao *et al.*, 2014); these were mostly intergeneric and interspecific crosses, involving genera within the *Saccharum* complex. The genus previously known as *Erianthus* (now *Saccharum*) was used in 18 artificial crosses, predominantly with *Saccharum arundinaceum* (synonym *Erianthus arundinaceus*). Other genera which have been crossed with sugarcane include *Bambusa*, *Imperata*, *Miscanthus*, *Sorghum* and *Zea*. The highest amount of surviving seedlings in cultivation was 1 371, resulting from a cross of *Saccharum* spp. hybrid x *Sorghum bicolor* (L.) Moench, representing a 9.70% seedling recovery rate from 14 141 total seedlings produced from the crosses (Hodnett *et al.*, 2010). The lowest seedling survival was in a cross involving *Zea mays* L., where only one from more than a thousand seedlings survived (Bonnett *et al.*,

2008). One of the reported crosses involving *Sorghum bicolor* completely failed with no true seedlings obtained (Bourne, 1935). Seventeen out of 35 studies that attempted hybridization unfortunately did not indicate the number of initial seedlings, but only the amount of seedlings that survived. From the few hybrids that did survive from these crosses, only one study reported vigorous and strong hybrids (Gao *et al.*, 2014) while others reported weak seedlings with slow growth rate and poor establishment (Bonnett *et al.*, 2008; Bourne, 1935; Fukuhara *et al.*, 2012; Hodnett *et al.*, 2010; OECD, 2013).

3.3.5 Distribution of relative parent lines

Only one of the reported hybridization efforts was conducted in South Africa involving a non-indigenous species from tropical Africa, namely *Saccharum* spp. hybrid x *Miscanthus violaceus* (K.Schum.) Pilg. (Brett, 1954). China, Brazil and Australia hosted a greater portion of these hybridization attempts. The genera *Bambusa*, *Narenga*, *Saccharum* and *Sclerostachya*, for which approximately 72% of the hybridizations with *Saccharum* species have been recorded, do not occur naturally within South Africa. Only a few species within *Imperata*, *Miscanthus*, and *Sorghum* occur in the wild in South Africa and *Zea mays* occurs as a cultivated crop.

3.3.6 Using molecular markers to verify hybrids

With the exclusion of the attempt that completely failed in crossing *Saccharum* species with *Sorghum bicolor*, the remaining crosses claim to have obtained hybrids. However, only 44.44% used molecular markers to verify the presence of the maternal and paternal alleles from putative hybrids, whereas the remaining crosses (55.56%) relied on visual inspection of inherited morphological characteristics against those of parent lines as well as chromosome counts. Molecular markers such as amplified fragment length polymorphisms, SSR markers and 5S rDNA PCR were used to be able to distinguish self-fertilised progeny from true hybrids (Cai *et al.*, 2005a; Piperidis *et al.*, 2000). For instance, the 5S rDNA PCR confirmed 37 offspring to be true hybrids in a cross involving *Saccharum officinarum* x *S. arundinaceum*, whereas progeny of *Saccharum* spp. hybrid x *Sa. arundinaceum* were all verified to be self-fertilised in origin. Similarly, all tested *Sa. officinarum* x *Sa. longisetosum* (Andersson) V.Naray. ex Bor. progeny were confirmed to be true hybrids while no hybrids were produced from *Saccharum* spp. hybrid x *Sa. longisetosum* during the same study (Aitken *et al.*, 2006).

Table 3.1: Hybridization incidents reported in the literature to involve either *Saccharum* species or commercial sugarcane cultivars, with wild related species. (¹spontaneous and ²artificial hybridization)

Hybridization (¹ or ²) and <i>Saccharum</i> parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
¹ Cultivated <i>Sa.</i> <i>officinarum</i> L.	Unknown	Wild <i>Miscanthus</i> spp. Wild <i>Saccharum</i> spp.	N	<i>Saccharum</i> <i>maximum</i> (Brongn.) Trin. (Synonym <i>Erianthus</i> <i>maximus</i>)	Spontaneous hybridization produced <i>Saccharum maximum</i> (Brongn.) Trin., claims are based on chromosomal evidence Spontaneous hybridization between cultivated and wild <i>Saccharum</i> is implicated in the origin of 'many cultivars', based on morphology and genetic traits	^a Daniels & Roach,1987 ^b Ellstrand <i>et al.</i> , 1999
¹ <i>Sa.</i> <i>officinarum</i> L.	Unknown	<i>Saccharum spontaneum</i> L.	N	<i>Sa. sinense</i>	Supported by cytogenetic studies and morphological characteristics	^{a,b} Parthasarathy, 1948
² <i>Sa.</i> <i>spontaneum</i> L. /	Male	<i>Bambusa bambos</i> (L.) Voss. (Syn. <i>Bambusa</i> <i>arundinacea</i> Willd.)	N	4/960	Putative hybrids failed to germinate from seed/ died at seedling stage	^a Rao et al. 1967 ^b Australian Government,

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
<i>Sa. robustum</i> <i>L.</i>						2004; Bonnett <i>et al.</i> , 2008
² <i>Saccharum</i> spp. hybrid	Female	<i>B. bambos</i> (L.) Voss. (Syn. <i>B. arundinacea</i> Willd.)	Y	29 hybrids	Embryos were aborted during development	^a Rao <i>et al.</i> 1967 ^b OECD, 2013
² <i>Sa.</i> <i>officinarum L.</i>	Female	<i>B. bambos</i> (L.) Voss. (Syn. <i>B. arundinacea</i> Willd.)	N	2 seedlings	Putative hybrids did not show any <i>B.</i> <i>arundinacea</i> characteristics	^{a,b} Venkatraman, 1938 ^b Raghavan, 1952;
² <i>Sa.</i> <i>officinarum L.</i>	Female	<i>Saccharum</i> <i>arundinaceum</i> Retz. (Syn. <i>Erianthus</i> <i>arundinaceus</i> (Retz.) Jeswiet	N	19/>1000	All were sterile and weak with poor growth	^a Lee <i>et al.</i> , 1998 ^b Australian Government, 2004; OECD, 2013

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
² Saccharum spp. hybrid	Female	<i>Sa. arundinaceum</i> Retz. (Syn. <i>E. arundinaceus</i> (Retz.) Jeswiet	Y	Fertile F1 generation	All were self-fertilised progeny with no <i>Saccharum</i> traits	^a Janaki-Ammal, 1941 ^b OECD, 2013
² Sa. <i>officinarum</i> L.	Female	<i>Sa. arundinaceum</i> Retz. (Syn. <i>E. arundinaceus</i> (Retz.) Jeswiet)	Y	Not specified	43% of F1 generation were selfed progeny	^{a,b} Piperidis <i>et al.</i> , 2000 ^b OECD, 2013;
² Sa. <i>officinarum</i> L. / ² Saccharum spp. hybrid	Female	<i>Sa. arundinaceum</i> Retz. (Syn. <i>E. arundinaceus</i> (Retz.) Jeswiet)	Y	19/1328	37 hybrids confirmed, 18 died Cross with commercial sugarcane did not produce any hybrids	^{a,b} Piperidis <i>et al.</i> , 2000
² Saccharum spp. hybrid	Female	<i>Sa. arundinaceum</i> Retz. (Syn. <i>E. arundinaceus</i> (Retz.) Jeswiet)	Y	23 hybrids	Randomly selected progeny that exhibit maternal and paternal characteristics, were all fertile hybrids with vigorous growth	^{a,b} Gao <i>et al.</i> , 2014

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
² Saccharum spp. hybrid	Female	<i>Sa. arundinaceum</i> Retz. (Syn. <i>E. arundinaceus</i> (Retz.) Jeswiet)	Y	1/167	Five hybrids confirmed, 4 died during growth, leaving one weak hybrid	^{a,b} Fukuhara <i>et al.</i> , 2012
² Sa. <i>officinarum</i> L.	Female	<i>Sa. arundinaceum</i> Retz. (Syn. <i>E. arundinaceus</i> (Retz.) Jeswiet)	Y	Confirmed fertile hybrids	First report to confirm intergeneric fertile hybrids	^{a,b} Cai <i>et al.</i> , 2005a ^b Deng <i>et al.</i> , 2010; Wu <i>et al.</i> , 2014
² Saccharum spp. hybrid	Female	<i>Saccharum rufipilum</i> Steud. (Syn. <i>Erianthus fulvus</i> Nees ex Hack.)	Y		Surprisingly, F1 progeny with no maternal bands were considered to be true hybrids	^{a,b} Zhang <i>et al.</i> , 2009
² Sa. <i>officinarum</i> L. / ² Saccharum spp. hybrid	Female	<i>Saccharum</i> <i>longisetosum</i> (Andersson) V.Naray. ex Bor. (Syn. <i>Erianthus</i>	Y	10/10 hybrids No hybrids	Ten randomly selected F1 progeny of each of the two crosses showed that the progeny tested from the first cross (with <i>Sa. officinarum</i>) were all hybrids while	^{a,b} Aitken <i>et al.</i> , 2006

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
		<i>longisetosus</i> Andersson)			the second cross did not produce any hybrids First verification of <i>E. rockii</i> hybrids	
² Sa. <i>spontaneum</i> L.	Unknown	<i>Saccharum ravennae</i> (L.) L. (Syn. <i>Erianthus</i> <i>ravennae</i> (L.) P.Beauv.)	N	Fertile progeny	No evidence of hybridization from molecular markers	^a Janaki- Ammal, 1941 ^b OECD, 2013
² Saccharum spp. hybrid	Female	<i>Saccharum</i> <i>alopecuroides</i> (L.) Nutt. <i>Saccharum contortum</i> (Elliott) Nutt. <i>Saccharum giganteum</i> (Walter) Pers.	N	448 seeds	Did not verify whether off-spring were hybrids	^{a,b} Burner & Webster, 1994
² Sa. <i>officinarum</i> L.	Unknown	<i>Saccharum narenga</i> (Nees ex Steud.) Hack.	N	Progeny capable of producing seed	Progeny not known to be true hybrids	^a Janaki- Ammal, 1942

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
		(Syn. <i>Narenga porphyrocoma</i> (Hance) Bor)				^b Bonnett <i>et al.</i> , 2008; OECD, 2013
² Saccharum spp. hybrid	Unknown	<i>Imperata cylindrica</i> (L.) Raeusch.	N	35 seedlings	Not known whether progeny were true hybrids Cytological analysis suggested that 5 plants contained chromosomes from both parents	^a Janaki- Ammal, 1942 ^b Bonnett <i>et al.</i> , 2008; OECD, 2013; Price, 1959
² Saccharum spp. hybrid	Male	<i>Sorghum bicolor</i> (L.) Moench	N	Attempt failed	Hybrids were genetically weak	^{a,b} Bourne, 1935 ^b Hodnett <i>et al.</i> , 2010
² Saccharum spp. hybrid	Female	<i>So. bicolor</i> (L.) Moench	N	Not specified	Hybrids verified by morphological characteristics and chromosome count Some were progeny of selfed origin, with no <i>Saccharum</i> chromosomes	^{a,b} Gupta <i>et al.</i> , 1978

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
² Sa. <i>officinarum</i> L.	Male	<i>So. bicolor</i> (L.) Moench	Y	Four hybrids obtained from 3670 florets heavily loaded with pollen	After the 4 th generation during backcross, <i>Saccharum</i> chromosomes were eliminated from the intergeneric hybrids	^a Nair <i>et al.</i> , 2005 ^b Hodnett <i>et al.</i> , 2010; OECD, 2013
² Saccharum spp. hybrid	Male	<i>So. bicolor</i> (L.) Moench	N	1371/14141 seedlings 23 confirmed hybrids	Only assessed seedling morphology for sugarcane characteristics and chromosome counts	^{a,b} Hodnett <i>et al.</i> , 2010
² Sa. <i>officinarum</i> L.	Unknown	<i>Miscanthus fuscus</i> (Roxb.) Benth. (Syn. <i>Sclerostachya fusca</i> (Roxb.) A.Camus)	N	74 offspring	Successful backcross	^{a,b} Parthasarathy, 1948 ^b OECD, 2013;
² Sa. <i>officinarum</i> L.	Female	<i>Zea mays</i> L.	Y	1/>1000	Hybrid embryos were aborted during development	^a Jamaki <i>et al.</i> , 1941

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
						^b Bonnett <i>et al.</i> , 2008; OECD, 2013
² Saccharum spp. hybrid	Unknown	<i>Imperata cylindrica</i> (L.) Raeusch., Unspecified <i>Saccharum</i> spp. (previously <i>Erianthus</i>), <i>Zea mays</i> L.	N		Weak first and second generation descendants	^{a,b} Anderson, 1949
² Saccharum spp. hybrid	Female	<i>Sa. rufipilum</i> Steud. (Syn. <i>E. fulvus</i> Nees ex Hack.)	Y	Hybrids obtained		^a Foreman <i>et al.</i> , 2007 ^b OECD, 2013
² Saccharum spp. hybrid	Female	<i>Sa. longisetosum</i> (Andersson) V.Naray. ex Bor. (Syn. <i>E. rockii</i> Keng)	Y	Hybrids obtained	Unquantified number of hybrids	^a Foreman <i>et al.</i> , 2007 ^b OECD, 2013
² Saccharum spp. hybrid	Female	<i>Miscanthus violaceus</i> (K.Schum.) Pilg.	Y	12/>400	Seedlings had pithing thus couldn't be of any commercial value	^{a,b} Brett, 1954

Hybridization (¹ or ²) and Saccharum parent	Parent sex Female, Male or Unknown	Second parent	Verified by molecular markers (Y/N)	Outcome (seedlings that survived over total seedlings)	Comments	^a Original source/ conducted by and ^b As cited by source/s for this study
		(Syn. <i>Erianthus violaceus</i> K.Schum.)				^b OECD, 2013

3.4 Discussion

3.4.1 Spontaneous hybridization between *Saccharum* species and hybrids with wild relatives

A report by the Australian government (2004) found that horizontal gene transfer is the only way in which genes could be transferred from commercial sugarcane hybrids to related species. Although Daniels & Roach (1987; cited by Ellstrand *et al.*, 1999) suggested that spontaneous hybridization between cultivated *Sa. officinarum* and wild *Miscanthus* species led to the origin of *Sa. maximum*, wild hybridization has not been reported with current sugarcane hybrids. Cheavegatti-Gianotto *et al.* (2011) conducted a literature study for the regulation of GM sugarcane in Brazil and found no evidence of spontaneous hybridization involving sugarcane. Likewise, Bonnett *et al.* (2008) assessed the risk of transgene escape from GM sugarcane with no success on finding literature on natural hybridization. Consequently, their study was rather based on hybridization attempts by breeders. The only well-documented spontaneous hybridization event (Babu *et al.*, 2009; Bonnett *et al.*, 2008; Cheavegatti-Gianotto *et al.*, 2011; Grivet *et al.*, 2004) is the origin of *Sa. sinense* from natural hybridization between *Sa. officinarum* and *Sa. spontaneum*, first described by Parthasarathy (1948).

The effects of climate change have led to the expansion of drought-prone areas and ineffective irrigation methods, thus the development of drought tolerant sugarcane is necessary (Fukuhara *et al.*, 2012). Due to the restricted genetic base of sugarcane varieties, wild species present a good source for introgression of genes into these varieties. For instance, the well-developed root system and polyphenols contained in root tips of *Erianthus* makes the taxon tolerant to drought and various stresses (Fukuhara *et al.*, 2012; Gao *et al.*, 2014). Thus, 20% of studies were on hybridization involving *Sa. arundinaceum*. In agreement with these findings, Bonnett *et al.* (2008) found that the intergeneric hybridizations involving sugarcane were carried out to increase the genetic diversity and purposefully integrate targeted traits. Likewise, crop wild relatives have been useful as a source of genetic diversity in attempts to improve crop species like wheat and rice (Zhang *et al.*, 2017).

3.4.2 Artificial crosses achieved and seedling survival rate

Cheavegatti-Gianotto *et al.* (2011) found that even when the barriers to hybridization were eliminated in artificial crosses (flowering synchronised, male pollen viability increased, numerous florets hand pollinated), the success of progeny was really low. Attempts to produce plants containing *Erianthus* chromosomes were unsuccessful for 50 years (Gao *et al.*, 2014),

with Cai *et al.* (2005a) being the first study to verify the introgression of *Erianthus* genes into sugarcane. The difficulty was brought on by the incompatibility of pollen and pistils of the two genera (Gao *et al.*, 2014). Similarly with *Sorghum* (female parent) as one of the closest relatives to sugarcane, hybrid production success ranged from 0.13% in 1999 to 9.70% in 2008 (Hodnett *et al.*, 2010). The study indicated that the prezygotic pollen tube growth from foreign species is terminated before reaching the ovule. More than 80% of studies reported sterile progeny, difficulty in maintaining the seedlings and subsequent high mortality. This could be due to genetic incompatibilities amongst parents and/or unbalanced meiotic products (Campbell *et al.*, 2009). Therefore, it is clear that even with the intervention of breeders, generating hybrids from sugarcane is not easy.

While the reduced fitness of first generation progeny could indicate a lower probability of genes dispersing and persisting in natural areas, Snow *et al.* (1998) concluded that this reproductive barrier could be quite permeable. In their study of hybrids from wild and cultivated *Helianthus annuus*, they found that the fecundity rate of the initially weak F1 generation was not significantly different from that of purely wild plants after the second cross. In another study, Campbell *et al.* (2009) found that weak hybrids (*Raphanus raphanistrum* x *R. sativus*), when grown under favourable conditions, evolved to flower as early as the wild lineages and recovered wild-type pollen viability, thus showing rapid adaption of crop-wild populations.

3.4.3 Using molecular markers to verify hybrids

In attempts towards the improvement of sugarcane, progeny must be verified to be hybrids prior to use in introgression studies (Cai *et al.*, 2005a). Approximately 55% of studies relied on the use of morphological and cytological characteristics to verify the introgression of genes in hybrids. For instance, Gupta *et al.* (1978) suggested that the variation in morphology of progeny from a *Saccharum* x *Sorghum* cross indicated that their genomes were comprised of different combinations of parental chromosomes. Cai *et al.* (2005a) declared morphological characteristics unreliable as they do not distinguish between genuine hybrids, self-fertilised progeny or progeny resulting from pollen contamination. In addition to the unreliability, identification of hybrids from morphological traits was found to be difficult (D'Hont *et al.*, 1995).

While $2n$ chromosome counts on progeny has been found to correlate to some extent with the $n+n$ chromosome transmission from parents, distinguishing hybrids from selfed progeny poses a challenge. Selfing in the heterozygous 80-chromosome sugarcane with normal meiotic division produces morphologically variable offspring with the same number of chromosomes as the parent (Price, 1959). In addition, due to the varying number of unpaired and lagging

chromosomes, some seedlings resulting from self-fertilisation have been found with either fewer or extra numbers of chromosomes than the parental clone (Price, 1959). As a result, molecular markers are said to be the reliable method for identifying genuine hybrids (Cai *et al.*, 2005a). McIntyre and Jackson (2001) found that levels of selfing in sugarcane ranged from 0 to 17.6%, thus concluding that they were not a significant problem in the Australian sugarcane breeding program.

3.5 Conclusion

The attempts to produce hybrids with *Saccharum* species from artificial crosses highlighted some dissimilarities when compared to the undesired spontaneous transfer of a transgene in the wild. For instance, *Saccharum* has mostly been regarded as a female parent while, for a transgene to escape spontaneously, it is the species of the receiving population that would have to be female. The choice of *Saccharum* as female indicates its low ability to act as a pollen donor. Nearly all of the crosses were carried out in a controlled environment with maximised chances of successful hybridization. Some putative hybrids were not verified by molecular markers but by unreliable morphology and cytology based methods. However, even from the portion confirmed by molecular markers to be true hybrids, there was poor growth and low survival of hybrid seedlings. The reduced fitness in first generation progeny indicates that it would not be easy for crop-to-wild introgression to occur, as transmission of genes to subsequent generations would be low, if occurring. Therefore, there is a reduced likelihood of hybrid generation if any introgression occurs between commercial sugarcane and its wild relative species.

CHAPTER 4: DNA BARCODING

4.1 Introduction

DNA barcoding is an extensively used molecular tool that enables instant and precise identification (Li *et al.*, 2015) of unknown materials from known species using short DNA sequences of standard genome regions (Dong *et al.*, 2015; Su *et al.*, 2016). The principal aim of DNA barcoding is to build a shared communal resource bank of DNA sequences that can be utilised to identify organisms and for taxonomic clarification (Hollingsworth *et al.*, 2011). The requirements for DNA barcoding are easy amplification in the polymerase chain reaction (PCR), cost effectiveness, high throughput sequence and a short sequence length to allow bidirectional reading from a single-primer pair with less necessity for manual editing of sequence traces. Most importantly, an ideal barcode should be variable enough to provide maximal discrimination between closely related species, while short enough to allow for experimental manipulation (CBOL Plant Working Group, 2009; Dong *et al.*, 2015; Li *et al.*, 2015; Su *et al.*, 2016).

The mitochondrial cytochrome *c* oxidase subunit 1 (*CO1*) DNA barcode fits the criteria of an ideal barcode and is thus accepted as a standard barcode for animals (Su *et al.*, 2016). In plants, however, mitochondrial DNA substitution rates are low, therefore requiring alternative barcoding regions. At present, plant DNA barcoding technology relies mostly on chloroplast loci because of their relatively low evolutionary rates compared to nuclear loci (Li *et al.*, 2015). Based on a comprehensive evaluation of four portions of coding genes (*matK*, *rbcL*, *rpoB*, and *rpoC1*), and three noncoding spacers (*atpF–atpH*, *trnH–psbA*, and *psbK–psbI*), the CBOL Plant Working Group (2009) formally recommended the combination of two chloroplast loci, *matK* + *rbcL*, as the core barcode for land plants. In addition to this, the nuclear ribosomal ITS was set as a complementary marker to the core barcode during this study.

The *rbcL* gene, as the best characterised gene of the plastid region, is highly universal with good but not strong discriminating power. The *matK* gene on the other hand, is among the most rapidly evolving plastid regions, and offers high discrimination power, but has poor primer universality. The higher discriminatory power of ITS over the plastid regions makes it a powerful phylogenetic marker at species level. However, the multiple copies with considerable differentiation at inter- and intra-species level in this genome raised concern (Li *et al.*, 2015). Although there has been an ongoing search for a universal plant barcode, none of the available loci were found to work across all species. As proposed by the CBOL Plant Working Group (2009), a plant barcode should combine more than one locus, namely a

phylogenetically conservative coding locus (*rbcL*) with a rapidly evolving region (*matK*). This study explored the discriminatory power of the *matK*, *rbcL* and ITS barcodes on selected species of the Andropogoneae tribe.

4.2 Material and methods

4.2.1 Selection of *Saccharum* wild relatives

In order to select close relatives of sugarcane for phylogenetic analysis to determine the genetic relatedness with *Saccharum* species hybrids, a literature review was consulted Snyman *et al.*, (2018). Findings from a recent phylogenetic classification (Welker *et al.*, 2015) of the grass family indicated that Germainiinae and Sorghinae are the closest subtribes to Saccharinae. The phylogeny of *Saccharum* species was also studied (Skendzic *et al.*, 2007; Soreng *et al.*, 2015) and a list of closely related Andropogoneae species was generated (Table 4.1). Species names were checked against The Plant List (2013) to conform with the latest nomenclature.

4.2.2 Parental lines and outgroups

Several other species were selected for inclusion into the phylogenetic analysis, which included distantly related outgroups for tree rooting, closely related outgroups, parental lines and commercial sugarcane varieties as reference of relatedness with wild relatives (Table 4.2). Multiple outgroups (*Bothriochloa insculpta* (A.Rich.) A.Camus, *Chrysopogon zizanioides* (L.) Roberty, *Trachypogon spicatus* (L.f.) Kuntze and *Zea mays* L.) were included to increase the resolving power in the tree nodes and to improve bootstrap support values.

4.2.3 Collection of plant material

Leaf samples from *Saccharum* parental lines were collected from the nursery at SASRI (23 May 2016) as they do not occur naturally in the study area or southern Africa. Three sugarcane varieties were collected from commercial plantations (4-7 July 2016) immediately after testing for pollen viability (Appendix A-1). Herbarium vouchers from various herbaria and iSpot (www.ispotnature.org) were used to pinpoint localities and habitat types where the chosen wild relatives of *Saccharum* species have been collected in the past and are known to occur. Samples of plant leaf material were then collected from those localities (Mpumalanga and KwaZulu-Natal) for the selected wild relative species between 27 February and 9 March 2017. Fresh, young leaves were selected from each species as they contain the best quality and quantity of DNA. Leaf midribs were removed to eliminate the xylem and phloem. The leaves

were sprayed with 70% ethanol to remove potential dirt, then placed in 50ml falcon tubes filled with 15g silica gel. The falcon tubes were closed tightly and leaf samples were allowed to dry. Gloves and secateurs were also cleaned with 70% alcohol after each collection to avoid cross contamination between the collected species. Preserved plant specimens (vouchers) were deposited in the AP Goossens Herbarium (PUC) as future reference material (Appendix A-2).

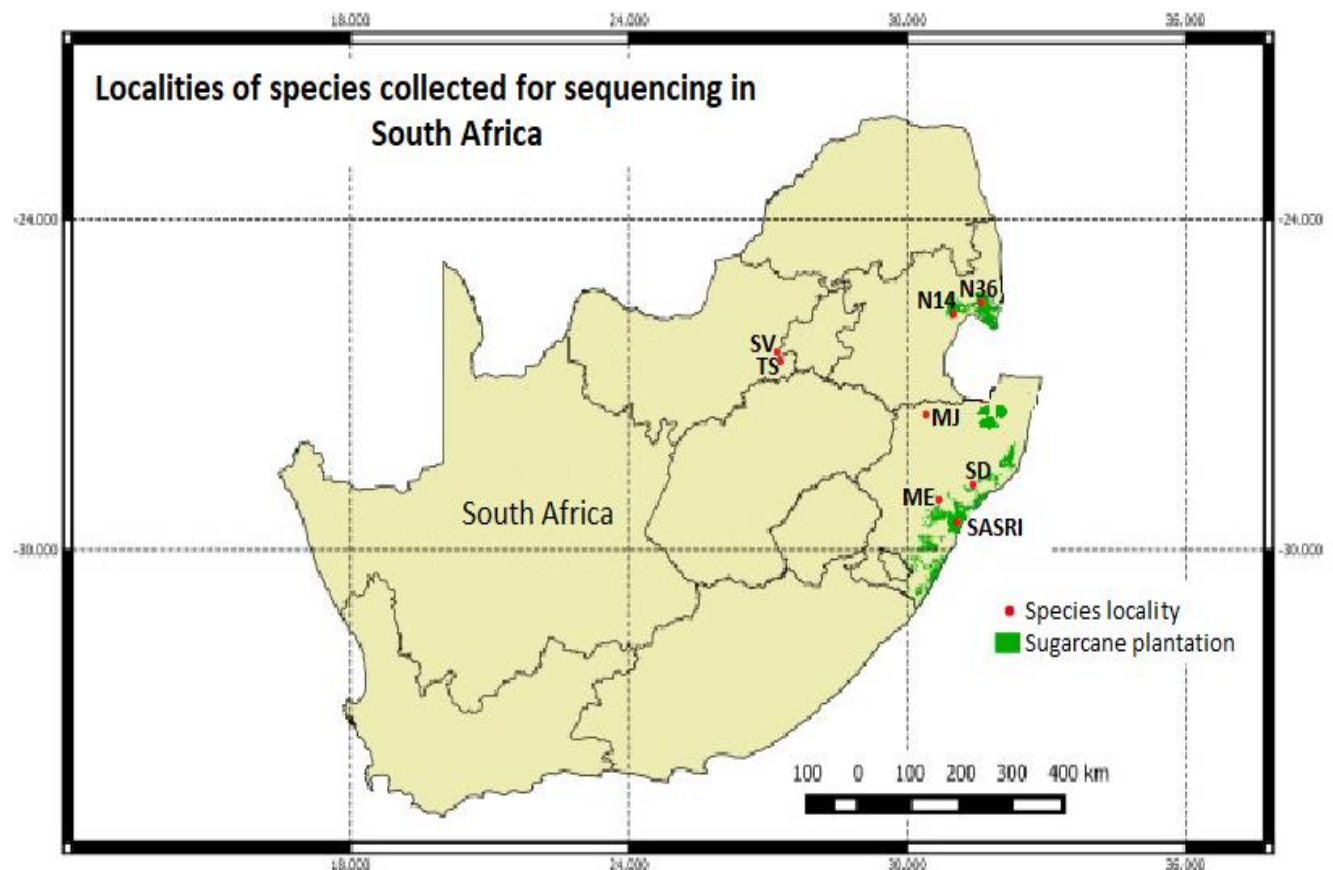


Figure 4.1: Localities where species were collected. N14 & N36: sugarcane commercial varieties; SV: *Sorghum versicolor*; TS: *Trachypogon spicatus* MJ; MJ: *Miscanthus junceus*; SD: *Sorghum x drummondii*; ME: *Miscanthus ecklonii*; SASRI: Site where NCo376 (sugarcane commercial variety), *Saccharum arundinaceum*, *Saccharum ravennae*, *Saccharum officinarum*, *Saccharum sinense*, *Saccharum robustum*, *Saccharum spontaneum*; *Sorghum arundinaceum* and *Imperata cylindrica* were collected. Names of these localities can be found in Appendix A-1.

Table 4.1: Targeted wild relatives of sugarcane from the Andropogoneae, their life form and habitat type.

Subtribe	Species	Common name	Life form	Habitat (Fish <i>et al.</i> , 2015)
Saccharinae	<i>Imperata cylindrica</i> (L.) Raeusch.	Cotton wool grass	Perennial	Poorly drained, damp soil such as wetlands and riverbanks. Grows in other habitat types in areas with high rainfall.
	<i>Microstegium nudum</i> (Trin.) A.Camus *	-	Annual	Moist shady places in forests.
	<i>Miscanthus ecklonii</i> (Nees) Mabb.	Daba grass	Perennial	Damp places e.g. riverbanks and forest edges.
	<i>Miscanthus junceus</i> (Stapf) Pilg.	Wireleaf daba grass	Perennial	Riverbanks and wetlands, often in standing water. Mostly in coarse sandy soil.
Graminiinae	<i>Trachypogon spicatus</i> (L.f.) Kuntze	Giant spear grass	Perennial	Open undisturbed grassland. Bushveld areas with a relatively high rainfall. Sandy and gravelly soil types.
Sorghinae	<i>Cleistachne sorghoides</i> Benth.**	-	Annual	Riverbanks and wetlands.
	<i>Sorghastrum nudipes</i> Nash*	Indian grass	Perennial	Wet areas.
	<i>Sorghastrum stipoides</i> (Kunth) Nash**	Needle Indian grass	Perennial	Wet areas.
	<i>Sorghum arundinaceum</i> (Desv.) de Wet & J.R. Harlan	Common wild sorghum	Annual	Weed of disturbed, damp areas such as water courses and besides irrigated land (often in sugarcane fields). Under natural conditions grows on riverbanks and near wetlands, usually in clay soil.

Subtribe	Species	Common name	Life form	Habitat (Fish <i>et al.</i> , 2015)
	<i>Sorghum x drummondii</i> (Nees ex Steud.) de Wet	Shatter cane		
	<i>Sorghum halepense</i> (L.) Pers. *	Johnson grass	Perennial	Disturbed places usually in damp clay or sandy soil.
	<i>Sorghum versicolor</i> Andersson ***	Black-seed sorghum	Annual	Disturbed places in damp black clay soil. Along roadsides, old cultivated fields or open veld. Black turf soil

Plant material for some wild relative species was not freshly collected for sequencing in this study: * Species for which sequences were obtained from GenBank; ** no sequences obtained from GenBank or in the wild and thus excluded from analysis; ***plant material was sourced from AP Goossens Herbarium.

Table 4.2: Species and hybrids included for phylogenetic tree construction within the Andropogoneae tribe.

Subtribes	Genus	Species	Motivation for inclusion
Saccharinae	<i>Saccharum</i> L.	<i>Saccharum arundinaceum</i> Retz.	Source of germplasm
		<i>Saccharum ravennae</i> (L.) L.	
		<i>Saccharum officinarum</i> L.	Parental lines
		<i>Saccharum robustum</i> E.W.Brandes & Jeswiet ex Grassl	
		<i>Saccharum sinense</i> Roxb.	
		<i>Saccharum spontaneum</i> L.	
		<i>Saccharum</i> sp. Co745	One of the first imported sugarcane hybrids
		<i>Saccharum</i> sp. N14	Commercial canes with most pollen viability
		<i>Saccharum</i> sp. N36	
		<i>Saccharum</i> sp. NCo376	Commercially cultivated mother variety to almost all commercial varieties
		<i>Saccharum</i> sp. 'Rowan Green'	Domestic chewing cane
Sorghinae	<i>Bothriochloa</i> Kuntze	<i>Bothriochloa insculpta</i> (A.Rich.) A.Camus	Sister group
	<i>Chrysopogon</i> Spreng.	<i>Chrysopogon zizanioides</i> (L.) Roberty	
Germainiinae	<i>Trachypogon</i> Nees	<i>Trachypogon spicatus</i> (L.f.) Kuntze	Closely related outgroup
Tripsacinae	<i>Zea</i> L.	<i>Zea mays</i> L.	Distantly related outgroup

4.2.4 DNA sequences sourced from genetic databases

ITS, *matK* and *rbcL* DNA sequences for closely related species and outgroups that could not be collected in the field were sourced from GenBank (National Centre for Biotechnology Information (NCBI) Bethesda, Maryland, United States of America). Accession numbers are presented in Table 4.3.

Table 4.3: Taxa selected as sugarcane related wild and outgroup species not collected in the field during this study. DNA nucleotide sequences (ITS, *matK*, *rbcL*) were obtained from GenBank.

Species name	GenBank and BOLD accession numbers		
	ITS	<i>matK</i>	<i>rbcL</i>
<i>Sorghastrum nudipes</i>	HIMS1931-12	FR821326	FR821342
<i>Microstegium nudum</i>	EU489067	KF163832.1	AB125343.1
<i>Panicum maximum</i>	AY129712.1	GU135005.1	GU135169.1
<i>Chrysopogon zizanioides</i>	DQ005089	KU556663.1	KP996860.1
<i>Bothriochloa insculpta</i>	AF190756.1	LN906656.1	KR737308.1
<i>Sorghum halepense</i>	HQ600488.1	HE967003.1	HQ600392.1
<i>Zea mays</i>	NC 001666.2	KU587494.1	KU695813.1

4.2.5 DNA Extraction

For each sample, 0.1 – 0.15g of dry plant leaf material was deep frozen at -196°C with liquid nitrogen in a mortar and then ground with a pestle. Through this process the cells were disrupted and the plant tissue became a fine powder. DNA was then extracted using GeneJET Genomic DNA Purification kit (ThermoFisher Scientific) following manufacturer's protocol. The elution volume was modified, with 100µl of buffer AE used instead of 200µl. This was done to increase the concentration of extracted genomic DNA. The extracted genomic DNA was stored at -20°C to be used in subsequent applications.

4.2.6 Quantity and quality measurements

The purity as well as the concentration of the DNA was assessed by using a NanoDrop ND-1000 (Nanodrop Technologies, US) spectrophotometer.

Some of the samples had low DNA concentration (5.7ng/µl) and no peak at 260nm (Figure 4.2a). This was the case for *Sorghum versicolor* for which plant material was sourced from a mounted voucher specimen collected in 1996. The bad quality of DNA obtained as indicated

by the low 260/280 ratio of 1.3, was most probably caused by a very low concentration (<10ng/μl) from which genomic DNA was accepted to have been degraded by the preservation process that involves repeated specimen heating and cooling over the years at the herbarium. For the other extracted genomic DNA, accepted samples to be used in downstream applications were those with a DNA concentration >10ng/μl and a 260/280 ratio between 1.6-2.2 (Figure 4.2b).

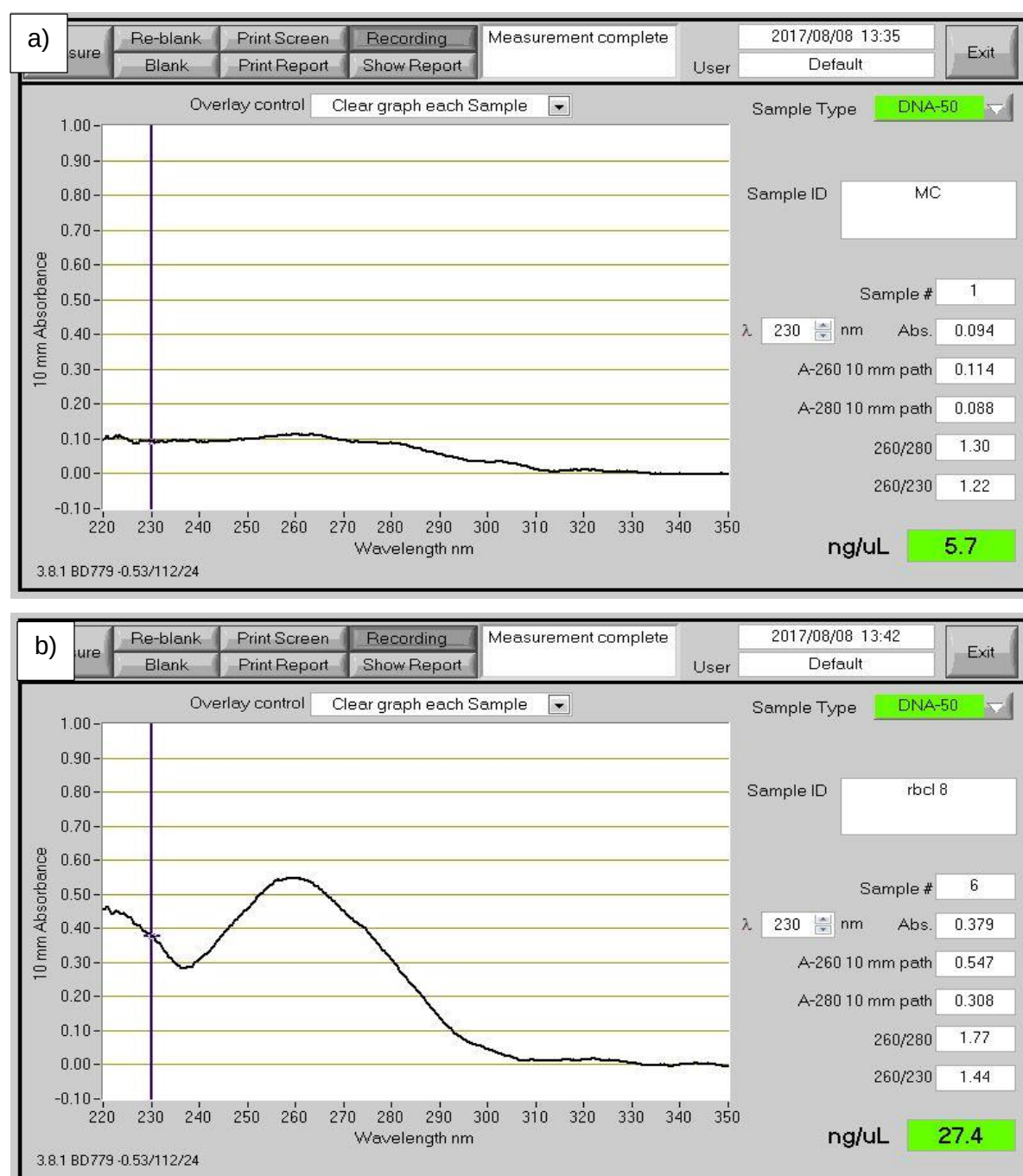


Figure 4.2: Nanodrop measurement window: a) Bad DNA quality with low concentration; b) Acceptable DNA quality and concentration to be used in downstream applications.

4.2.7 Polymerase Chain Reaction (PCR)

Amplification of a fragment of the target DNA barcodes (ITS, *matK* and *rbcL*) was done through PCR. The primer pairs used (synthesized by Integrated DNA Technologies, USA), as shown in Table 4.4, were chosen based on their universality, amplification and sequencing success rates for the Poaceae (Welker *et al.*, 2015). Brief centrifugation at 10 000 x g for 10 seconds was performed on a Minispin (Eppendorf, Germany) prior to opening to dislodge the oligo pellets and then re-suspended using TE buffer (10mM Tris pH 8.0, 1mM EDTA). The dilutions to create the stock and working solutions were per the yield information contained on IDT product specification sheets and oligo tube, as shown in Table 4.4 for stock solution. Working solution (10µM primer) was created by diluting the 100µM stock solution in a ratio 1:10 with TE buffer.

Table 4.4: The primers used for the amplification and sequencing of the ITS, *matK* and *rbcL* gene fragments.

Primer	For 100µM	5' to 3' primer sequence	Size (bp)	Reference
<i>rbcLa</i> -F	245µl	ATG TCA CCA CAA ACA GAG ACT	±550	Kress <i>et al.</i> , 2009
<i>rbcLa</i> -R	193µl	GTA AAA TCA AGT CCA CCR CG		
ITS 4F	211µl	TCC TCC GCT TAT TGA TAT GC	±650	Barnard <i>et al.</i> , 2013
ITS 5A	250µl	CTT TAT CAT TTA GAG GAA GGA		
<i>matK</i> 472F	255µl	CCC RTY CAT CTG GAA ATC TTG	±750	Yu <i>et al.</i> , 2011
<i>matK</i>	197µl	GCT RTR ATA ATG AGA AAG ATT		

A temperature gradient for PCR thermocycling conditions was run to determine the optimum annealing temperatures for the ITS, *matK* and *rbcL* primers. The temperature gradient range were as follows for each barcode: ITS, 48-52°C; *matK*, 48-55°C; and *rbcL*, 48-52°C each with an interval of 0.5°C. The optimal annealing temperatures to be used during PCR were then chosen based on the absence of other bands and primer-dimers, the fragment size as was compared to the commercially available DNA size ladder (1 kb, O'GeneRuler, Fermentas, US) and intensity of the generated fragment. The optimal annealing temperatures were 50°C for ITS and *rbcL*, and 48°C for *matK*.

PCR amplification was carried out using the C1000 Thermal Cycler (BioRad, US). The reaction mixture included 2X KAPA Taq readyMix (1x PCR buffer, 2 U Taq DNA polymerase, 0.2mM of each dNTP, 1.5mM MgCl₂ and stabilisers), 0.5µM forward and reverse primers, 5-50ng DNA template and PCR nuclease-free water. Three rounds of PCR optimization were done to successfully amplify the DNA fragments (except for ITS); these included the addition of 1.5mM

MgCl₂, 5% (v/v) DMSO as well as modifying the quantity of DNA template as shown in Table 4.5.

The thermocycling conditions were used with different annealing temperatures. For each primer set the initial denaturation step was at 94°C for 3 minutes, followed by denaturation at 94°C for 60 seconds. Annealing temperatures varied depending on the primer set, 50°C for 30 seconds for ITS and *rbcL*, and 48°C for 40 seconds for *matK*. The extension was at 72°C for 30 seconds for ITS and *rbcL*, and 60 seconds for *matK*. These thermocycling conditions were followed for 35 cycles for ITS and *rbcL*, and 40 cycles for *matK*. The last stage was a final extension at 72°C for 10 minutes.

Table 4.5: Amount (µl) of PCR reactants for each DNA fragment for the three rounds that were run during the optimising stages. Annealing temperatures remained at 50°C for ITS and *rbcL*, and 48°C for *matK* throughout these rounds.

Reactants (µl)	PCR Round 1			PCR Round 2			PCR Round 3		
	ITS	<i>matK</i>	<i>rbcL</i>	ITS	<i>matK</i>	<i>rbcL</i>	ITS	<i>matK</i>	<i>rbcL</i>
Ready mix	10	10	10	10	10	10	-	12.5	12.5
F+R primer	1.6	1.6	1.6	1	1	1	-	1.6	1.6
DD water	6.4	6.4	6.4	4.75	6.75	5.75	-	6.65	5.65
DNA template	2	2	2	3	1	2	-	2	3
MgCl ₂	-	-	-	-	-	-	-	1	1
DMSO	-	-	-	1.25	1.25	1.25	-	1.25	1.25
Total volume	20	20	20	20	20	20	-	25	25

4.2.8 Agarose gel electrophoresis

PCR products were visualised on a 1% (w/v) agarose gel to assess the success of DNA barcode amplification. The gel consisted of 1g agarose dissolved in 100ml of 1X TAE buffer (40mM Tris, 20mM Acetic acid, and 100mM EDTA, pH 8.0) and ethidium bromide (1µg/ml) (Bio-Rad, UK). Five microliters (µl) DNA molecular weight marker (ThermoFisher Scientific, Fermentas-US) mixed with 3µl of 6X Orange loading dye (250-10000bp, O'GeneRuler) was used for the comparison and confirmation of the molecular weight of the amplified DNA barcodes. The remainder of the wells were loaded with 2µl of PCR product and 3µl of 6X Orange loading dye. Gel electrophoresis was performed at 75 volts for 40-60 minutes in a Mini Sub-cell GT and a Power-Pac (Bio Rad).

ITS, *rbcL* and *matK* DNA barcode agarose gel images were captured using a Bio-Rad ChemiDoc MP gel imaging system (Figures 4.3-4.5). Whenever an intense band of DNA fragment appeared at the right molecular length, it was considered as a successfully amplified DNA region (Figures 4.3-4.5). Wearing UV-resistant glasses, these bands were cut out above UV-light and put into labelled tubes, then stored at -4°C until use in downstream applications.

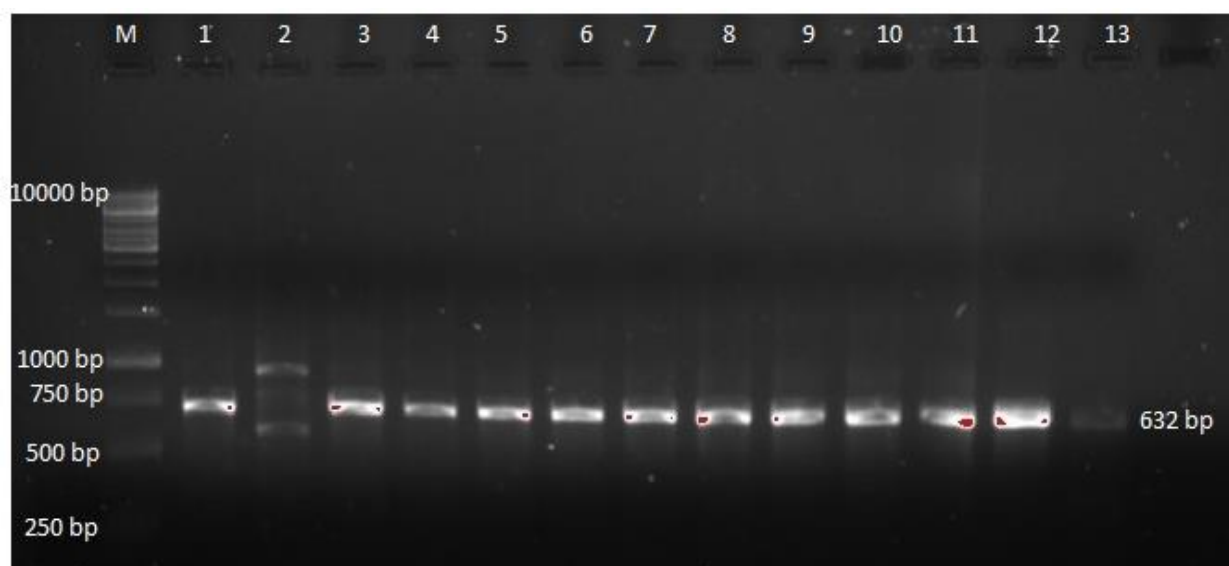


Figure 4.3: Successful DNA amplification of the ITS DNA barcode with length of ± 650 bp for samples 1,3 -12. Sample 2 produced multiple non-specific bands while sample 13 had a weak band, these were used for the second round of PCR. The first well marked 'M' is the DNA molecular weight marker, ranging from 250-10000 bp. 1: *Saccharum robustum*; 2: *Saccharum spontaneum*; 3: *Saccharum ravennae*; 4: *Saccharum arundinaceum*; 5: *Saccharum officinarum*; 6: *Saccharum sinense*; 7: *Saccharum* sp. 'Rowan Green'; 8: *Saccharum* sp. Co745; 9: *Saccharum* sp. N14; 10: *Saccharum* sp. N36; 11: *Saccharum* sp. NCo376; 12: *Imperata cylindrica* and 13: *Miscanthus ecklonii*

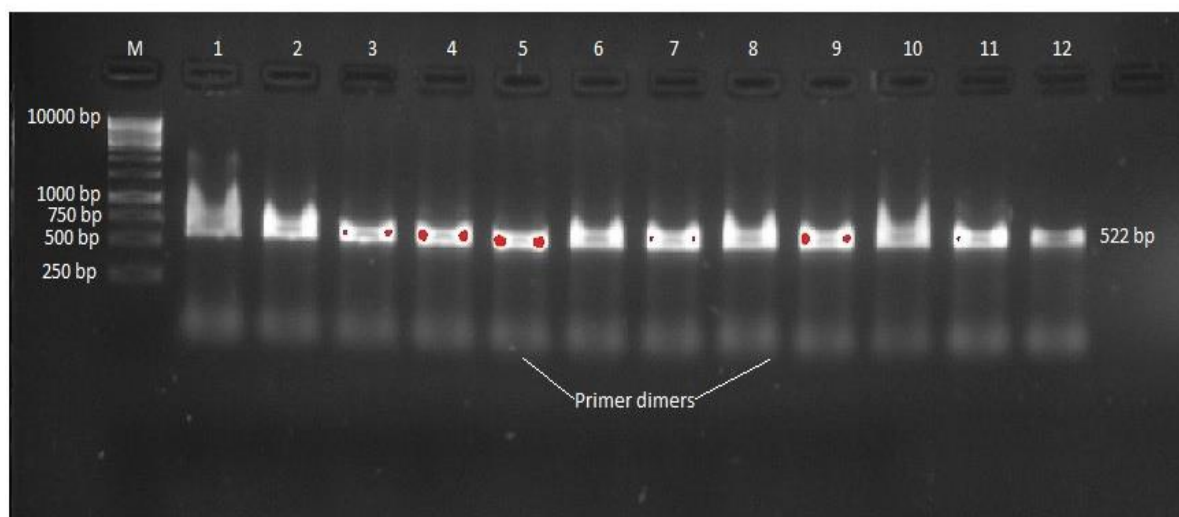


Figure 4.4: Successful DNA amplification of the *rbcL* DNA barcode with length of ± 550 bp for 12 samples. Primer dimers also formed from this reaction. 1: *Saccharum robustum*; 2: *Saccharum spontaneum*; 3: *Saccharum ravennae*; 4: *Saccharum arundinaceum*; 5: *Saccharum officinarum*; 6: *Saccharum sinense*; 7: *Saccharum* sp. Rowan green; 8: *Saccharum* sp. Co745; 9: *Saccharum* sp. N14; 10: *Saccharum* sp. N36; 11: *Saccharum* sp. NCo376 and 12: *Imperata cylindrica*

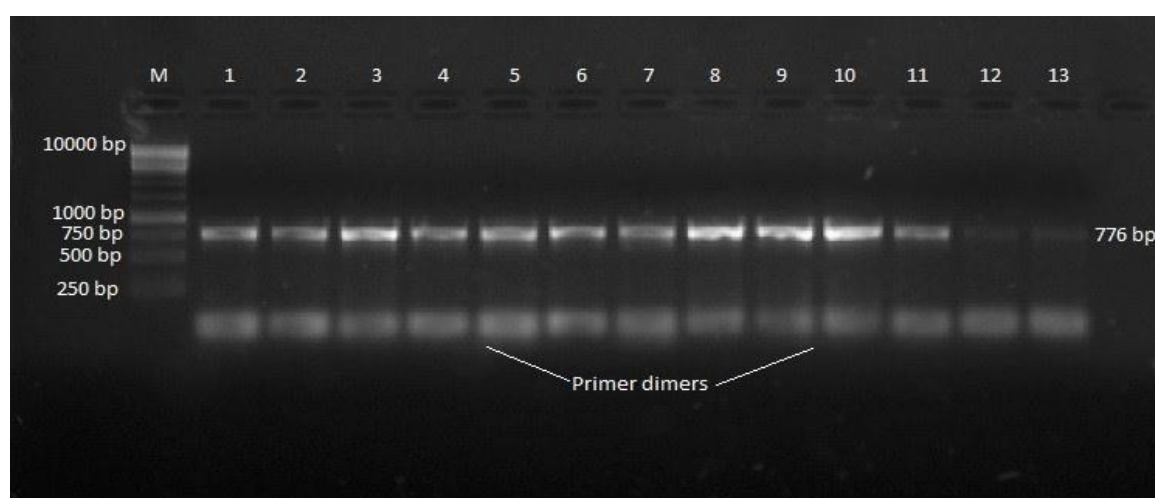


Figure 4.5: Successful DNA amplification of the *matK* DNA barcode with length of ± 750 bp for samples 1-11. Samples 12 and 13 produced weak bands thus were used for a second round of PCR. Primer dimers also formed from this reaction. 1: *Saccharum robustum*; 2: *Saccharum spontaneum*; 3: *Saccharum ravennae*; 4: *Saccharum arundinaceum*; 5: *Saccharum officinarum*; 6: *Saccharum sinense*; 7: *Saccharum* sp. Rowan green; 8: *Saccharum* sp. Co745; 9: *Saccharum* sp. N14; 10: *Saccharum* sp. N36; 11: *Saccharum* sp. NCo376; 12: *Imperata cylindrica* and 13: *Miscanthus ecklonii*

4.2.9 Post PCR Purification

PCR clean-up was done to purify the amplified DNA fragment. Short primers, unincorporated dNTPs, enzymes, short-failed PCR products, and salts from PCR reactions were removed using two purification techniques. The first involved purifying the bands that were cut out from the agarose gel using the Zymoclean gel DNA recovery kit (Zymo Research) following the manufacturer's protocol. The second was purifying of the original PCR product after assessing the success of its amplification on the agarose gel. This was done using GeneJET PCR purification kit. The amount of recovered DNA from the original PCR product was better compared to the first method, but this could be attributed to the fact that only 2µl of the PCR product was loaded into the gel, leaving 18/20µl or 23/25µl for direct purification. Nanodrop measurements were repeated to measure the DNA concentration of the purified products and only concentrations $\geq 16\text{ng}/\mu\text{l}$ were acceptable for use in sequencing.

4.2.10 Sequencing

All sequencing reactions were performed by the South African Sugarcane Research Institute (SASRI), Mount Egdecombe, South Africa. The sequencing reactions and post-PCR purification were performed with BigDye Terminator V1.3 cycle sequencing kit (Applied Biosystems, 2014). This was followed by fluorescence-based DNA analysis using capillary electrophoresis technology on the 3500 Genetic Analyser (Hitachi, Applied Biosystems). Sequences were analysed and trimmed using Sequencing Analysis V5.3.1 (Applied Biosystematics).

All DNA fragments were successfully amplified and sequenced. The country of origin, their herbarium accession numbers, as well as the collecting localities are described in Appendix A-1. The description also includes species for which sequences were sourced from GenBank. Combined, it provides 24 sequences that were in the phylogenetic analysis for each gene region.

4.2.11 Editing and alignment of sequences

Phylogenetic analysis was performed using MEGA 7 software. The ITS, *rbcL* and *matK* matrices were verified manually and corrected to eliminate ambiguous bases. Missing data and gaps were eliminated. Sequences were trimmed and then aligned for each DNA barcode using Muscle within MEGA 7.

4.2.12 Results analysis

Statistical values for each gene region such as the guanine-cytosine (GC) content, coverage, percentage conserved sites, parsimonious informative sites as well as the pairwise % identity were obtained on Geneious R11 (Kearse *et al.*, 2012). Since the GC pair is bound by three hydrogen bonds compared to two hydrogen bonds between the AT (adenine-thymine) pair in DNA, higher GC content is associated with stability.

4.2.12.1 Tree-based analysis

Initially, the distance method, Neighbor-Joining (NJ) (Saitou & Nei, 1987), as well as a character-based approach, Maximum Parsimony (MP) (Tamura *et al.*, 2011) were used. However, MP had poor bootstrap values and was thus replaced with another character-based approach, maximum likelihood (ML) (Tamura *et al.*, 2011). To determine the best substitution model for each barcode region, the “find best DNA/Protein Models” function was chosen using default settings. Kimura 2-parameter (K2) was calculated to be the best DNA substitution model for all the barcodes (Appendix B1, B2 & B3). The K2 model was however incalculable for the *matK* barcode using the NJ method; therefore, a P-distance model was used instead. The Markov Chain Monte Carlo (MCMC) settings were kept on default values. *Zea mays* as part of the Andropogoneae, is believed to have diverged before sugarcane diverged from sorghum (Guimaraes *et al.*, 1997), thus all the trees were rooted with *Zea mays*. A bootstrap percentage (based on 1000 bootstrap replicates) $\geq 80\%$ was considered a high bootstrap support, a bootstrap support of 50%-79% as moderate and a bootstrap support $< 50\%$ as weak (De Vienne *et al.*, 2007). For ML, initial tree(s) for the heuristic search (NNI) were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. All positions containing gaps and missing data were eliminated.

4.2.12.2 Congruency

A newick tree for each DNA fragment and phylogenetic tree analysis method was exported in a *.nwk format. Species names were sorted alphabetically and each assigned an alphabetic letter in an ascending order (e.g. letter A for *Bothriochloa insculpta* and X for *Zea mays*). A test for topological congruence based on maximum agreement subtrees (MAST) was conducted using the Icong index. Subsequently, the congruency index and *p*-value for the matrices were calculated (de Vienne *et al.*, 2007). Matrices for the core barcode (*rbcL*+*matK*) and all the barcodes (core barcode + ITS) were concatenated using Geneious R11 software

(Kearse *et al.*, 2012) for trees that were more congruent than expected by chance. NJ trees were re-constructed using both MEGA 7 (K2 model) (Kumar *et al.*, 2016) and Geneious R11 (Jukes-Cantor model) (Jukes & Cantor, 1969) in order to determine which method provided the highest bootstrap support for the monophyletic groups. Where Geneious R11 is concerned, the Jukes-Cantor model was used, as this programme does not have the option of the K2 model as in MEGA 7. Furthermore, Jukes-Cantor is usually a preferred substitution model due to its assumptions of equal base frequencies and mutation rates (Erickson, 2010). ML trees were however constructed using MEGA 7, since Geneious R11 does not contain any character-based tree building methods.

4.3 Results

4.3.1 Sequence statistics of each gene fragment

From the 24 sequences obtained for each of the three barcodes, the aligned nucleotide length for each fragment varied from 548-616 base pairs (Table 4.6). The *rbcL* fragment contained the highest percentage (96.27%) of conserved sites between species sequences, with only 2.80% of parsimonious informative sites. This low variation of nucleotide bases between the species, coupled with the lack of informative sites, indicated that the *rbcL* fragment had lower ability to discriminate between the related species used in this study. The ITS fragment on the other hand, had the lowest percentage of conserved sites (24.82%) with the highest percentage of informative sites (52.19%). In addition to this, the GC content of 62.52% showed ITS to be the most stable DNA fragment compared to *rbcL* (44.91%) and *matK* (34.47%) respectively.

Table 4.6: Statistics for each gene fragment from 24 DNA sequences after alignment and trimming

DNA fragment	Aligned length (bp)	Nucleotide positions used in final data-set (bp)	Conserved sites (%)	Informative sites (%)	GC content (%)
ITS	548	59	24.82	52.19	62.52
<i>matK</i>	616	210	33.12	19.48	34.47
<i>rbcL</i>	536	471	96.27	2.80	44.91
<i>matK+rbcL</i>	1152	681	62.50	11.72	40.4
<i>matK+rbcL+ITS</i>	1700	740	50.35	24.76	47.3

4.3.2 Nucleotide identity and pairwise differences between wild relatives and *Saccharum* species

The variation in percentages of mean identical nucleotides across wild relatives and sugarcane were obtained from the difference between highest and lowest identity. *matK* had the highest range (60.96%), followed by ITS (37.17%) and *rbcL* (2.05%). The amount of identical nucleotides for the ITS and *rbcL* DNA fragments indicated that *So. arundinaceum* had the highest mean nucleotide match (Table 4.7) with the commercial sugarcane cultivars (NCo376, N36 and N14). *matK* in contrast showed the highest mean identity with *Imperata cylindrica*. The ITS and *matK* fragments had similar results, all showing *Sorghum*

arundinaceum, *Imperata cylindrica* and *Miscanthus ecklonii* to have the highest mean nucleotide identity with commercial cultivars and the parental lines (*Sa. spontaneum*, *Sa. sinense*, *Sa. robustum* and *Sa. officinarum*). *rbcL* however indicated that *Miscanthus junceus*, *So. arundinaceum* and *So. halepense* were equally identical (99.87%) to commercial sugarcane varieties.

Estimates of evolutionary divergence between sequences by pairwise differences were in support of the results of nucleotide identity. The number of base substitutions per site between sequences is attached as Appendix C. There was no consistency across the DNA fragments and species computed to have no nucleotide substitution with the commercial varieties and *Saccharum* parental lines. However, the most common species amongst these fragments were *So. arundinaceum*, *Miscanthus ecklonii* and *Imperata cylindrica*. The pairwise distances ranged from zero to 1.365, 0.024, 1.204 for ITS, *rbcL* and *matK* respectively. ITS was better at determining the pairwise substitution (average 0.345), followed by *matK* (0.174) and *rbcL* (0.006).

Table 4.7: The percentages of identical nucleotides from pairwise comparison of DNA sequences between wild relative species and commercial sugarcane species (Com.), as well as the parental lines (Parent).

Species name	% Identity							
	ITS		<i>matK</i>		<i>rbcL</i>		Average	
	Com.	Parent	Com.	Parent	Com.	Parent	Com.	Parent
<i>Imperata cylindrica</i>	94.68	91.00	99.84	99.27	99.69	99.81	98.07	96.69
<i>Microstegium nudum</i>	62.80	44.66	98.21	97.81	99.69	99.81	86.90	80.76
<i>Miscanthus ecklonii</i>	98.11	93.11	99.51	99.11	98.19	98.32	98.60	96.84
<i>Miscanthus junceus</i>	95.90	91.13	38.88	38.83	99.87	100	78.22	76.66
<i>Sorghastrum nudipes</i>	87.61	80.31	98.54	98.18	99.31	99.44	95.15	92.64
<i>Sorghum arundinaceum</i>	98.11	93.29	99.51	99.11	99.87	100	99.16	97.47
<i>Sorghum x drummondii</i>	93.09	87.33	58.77	58.81	97.82	97.95	83.23	81.36
<i>Sorghum halepense</i>	60.94	42.36	99.28	98.88	99.87	100	86.70	80.41
<i>Sorghum versicolor</i>	94.56	89.53	99.34	99.01	99.69	99.81	97.86	96.12
<i>Trachypogon spicatus</i>	92.95	86.97	98.54	98.14	99.50	99.63	97.00	94.91

Red boxes highlight the three highest percentages of identical nucleotides per column

4.3.3 Maximum Likelihood method

4.3.3.1 ITS

The tree with the highest log likelihood (-401.42) is shown in Figure 4.6. The phylogenetic tree for the ITS region consisted of two clearly separated groups composed mainly of *Saccharum* species and outgroup species (Figure. 4.6), each supported by bootstrap values of 94% and 96% respectively.

Miscanthus junceus branched off from the *Saccharum* clade to form a sister lineage with *Saccharum arundinaceum*, separating from *Miscanthus ecklonii* and *Sorghum arundinaceum* as descendants of a common ancestor with all the commercial sugarcane cultivars. The descendants of this *Saccharum* clade formed a polytomy supported by a weak bootstrap of 36%. *Sorghum versicolor*, just like *Saccharum officinarum* appears to have diverged from the rest of the in-group. Also, *Sorghum x drummondii* formed a sister clade with *Trachypogon spicatus* (closely outgroup species) having bifurcated from the first order internal node, supported by a high bootstrap of 98%. *Microstegium nudum* and *Sorghum halepense* placed within the clade of closely related outgroup with high support (84%) as a sister lineage.

4.3.3.2 rbcL

The tree with the highest log likelihood (-802.15) is shown in Figure 4.7. Distinctive grouping into three clades was observed in the *rbcL* ML method, two clades had a moderate bootstrap support while one clade had a high bootstrap support. Similarly, to the ITS ML tree, the first clade with the *Saccharum* species contained polytomies. *Sorghum arundinaceum* was amongst this clade whereas *Miscanthus ecklonii* descended into a clade containing a sister group species (*Chrysopogon zizanioides*) with a high bootstrap support (92%). *Sorghastrum nudipes* and *Trachypogon spicatus* as a sister lineage appear to have be the distantly placed from in-group species.

4.3.3.3 matK

The tree with the highest log likelihood (-1083.22) is shown in Figure 4.8. A single clade grouping containing all the *Saccharum* species as descendants was observed in the ML tree for *matK* barcode. Only *Sorghum x drummondii* and *Miscanthus junceus* of the closely related species placed out from the *Saccharum* polytomy, with the latter forming a sister lineage with *Chrysopogon zizanioides*.

4.3.3.4 Summary of the ML method

ITS, *rbcL* and *matK* each had seven, four and two good bootstrap values (>50%) respectively. *Sorghum x drummondii* always descended off from the *Saccharum* species clade to be the outmost descendant of the in-group or form sister lineage with outgroup species. *Sorghum arundinaceum* on the other hand, was always sister lineage or descendant sharing a common ancestor with all the sugarcane varieties. *Miscanthus ecklonii* and *Imperata cylindrica* were also descendants sharing a common ancestor with sugarcane varieties for the *rbcL* and *matK* ML trees. There was no consistency between the groupings of species across the barcodes. For instance, *Miscanthus ecklonii* was a descendant from a common ancestor with outgroup species with a good support (62%) for *rbcL*, while ITS and *matK* showed this species to belong with the *Saccharum* clade.

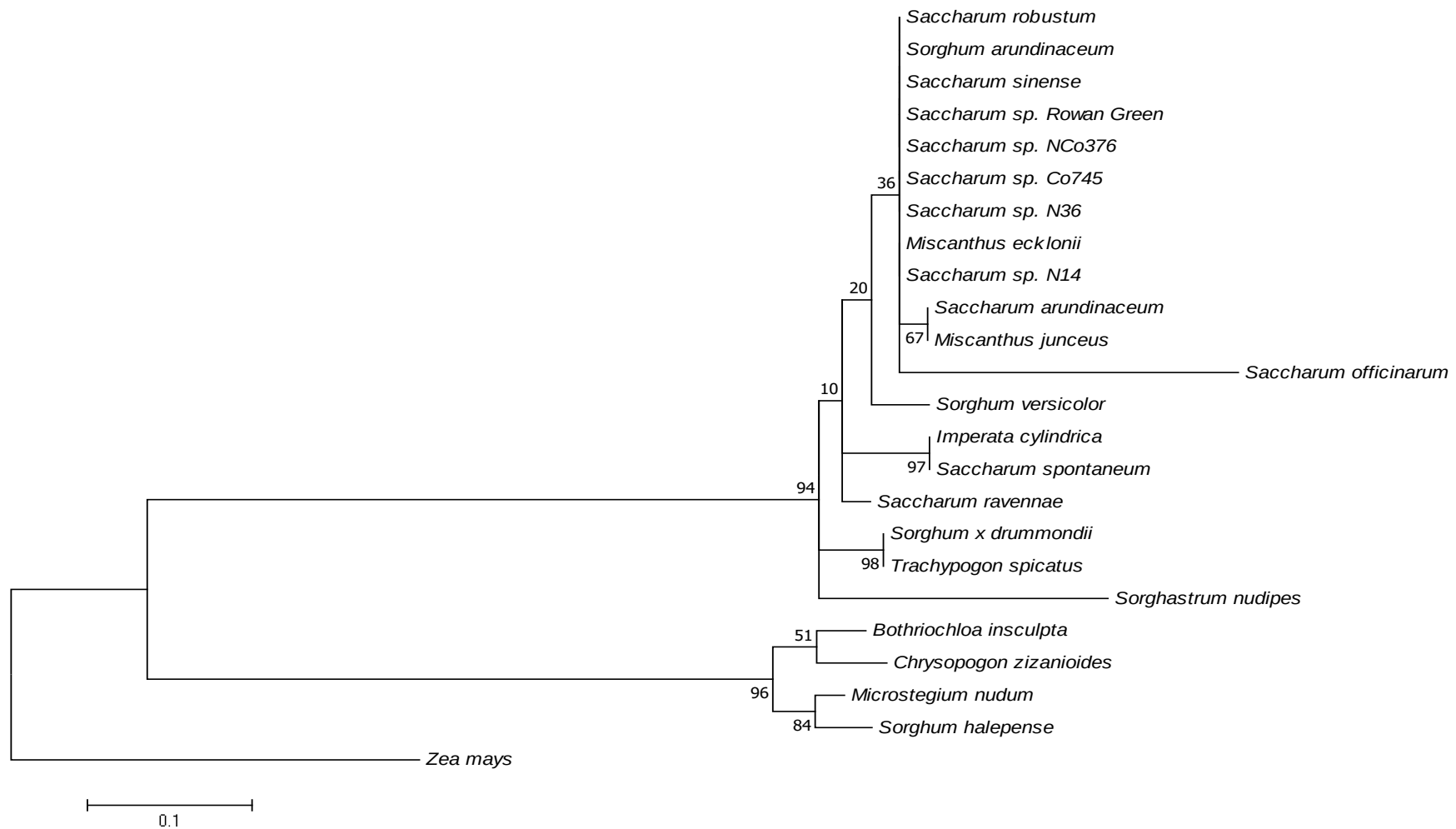


Figure 4.6: Maximum Likelihood phylogeny (K2 model, 1000 bootstrap replicates) based on a DNA fragment of ITS. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.

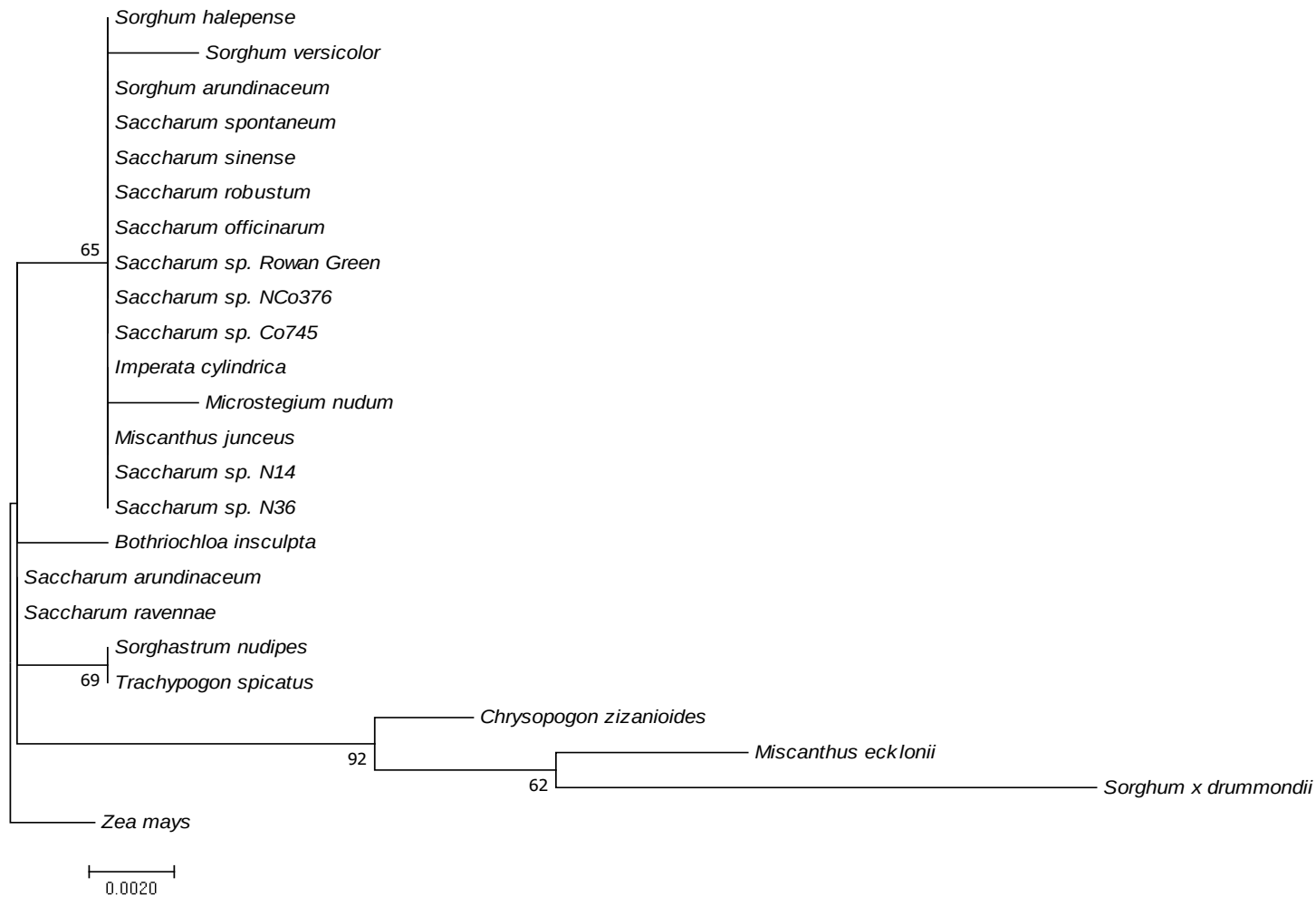


Figure 4.7: Maximum Likelihood phylogeny (K2 model, 1000 bootstrap replicates) based on a DNA fragment of *rbcL*. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.

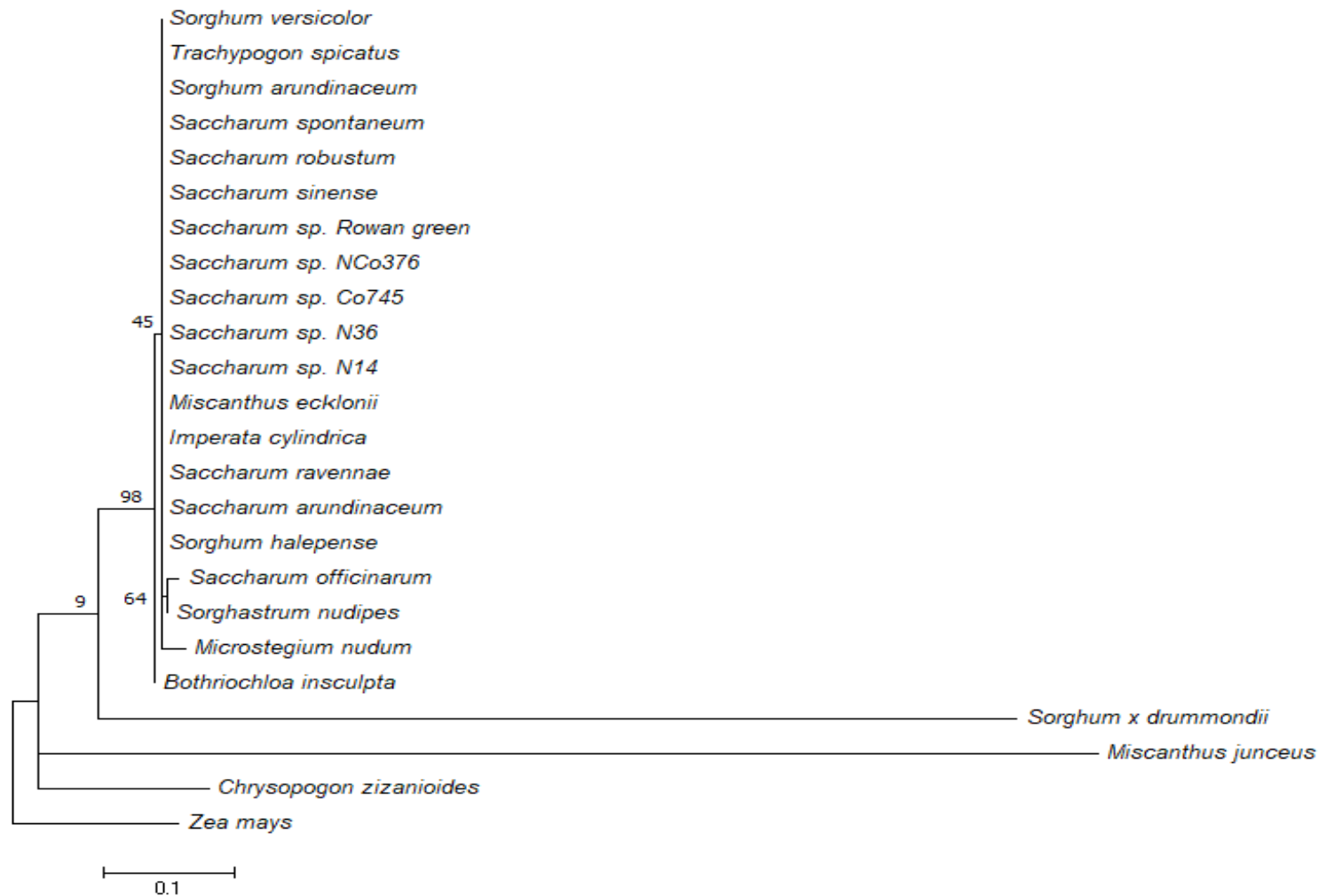


Figure 4.8: Maximum Likelihood phylogeny (K2 model, 1000 bootstrap replicates) based on a DNA fragment of *matK*. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated.

4.3.4 Neighbor Joining method

4.3.4.1 ITS

The optimal tree (Figure 4.9) with the sum of branch length =1.92741578 is shown. Two clades supported by bootstrap >80% were evident on the ITS NJ tree (Figure 4.9). The first clade contained polytomies, from which all the *Saccharum* species and wild relatives (excluding *Sorghum halepense* and *Microstegium nudum*) formed a monophyletic group. Sister lineages of *Sorghum x drummondii* with *Trachypogon spicatus*, *Imperata cylindrica* with *Saccharum spontaneum*, and *Miscanthus junceus* with *Saccharum arundinaceum* suggest that these taxa were the last to diverge from this clade. The second clade, of a monophyletic outgroup, had evolved from the most primitive ancestor with *Zea mays*, from which *Sorghum halepense* and *Microstegium nudum* have placed as direct descendants.

4.3.4.2 rbcL

The optimal tree (Figure 4.10) with the sum of branch length =0.042490827 is shown in Figure 4.10. With a clade support of 93%, *Sorghum x drummondii* was a member of the second divergent group from the core Andropogoneae after *Zea mays*, (Figure 4.10), from which it formed a sister lineage with *Miscanthus ecklonii*.

4.3.4.3 matK

The optimal tree (Figure 4.11) with the sum of branch length =1.06331845 is shown (210 nucleotide positions). Polytoomy from a monophyletic group of *Saccharum* descendants showed *Saccharum officinarum* to have recently evolved in a sister lineage with *Sorghastrum nudipes* (Figure 4.11). *Sorghum x drummondii* and *Miscanthus junceus* grouped distantly from the in-group.

4.3.4.4 Summary of the NJ method

The ITS barcode had seven good bootstrap support values (four strong and three moderate), while *matK* (two strong and three moderate) and *rbcL* (one strong and four moderate) each had five. *Saccharum* species always grouped together into a single clade, from which *Sorghum arundinaceum* always diverged from a common ancestor with the commercial sugarcane varieties. Branching of out groups from the in-group was always well supported.

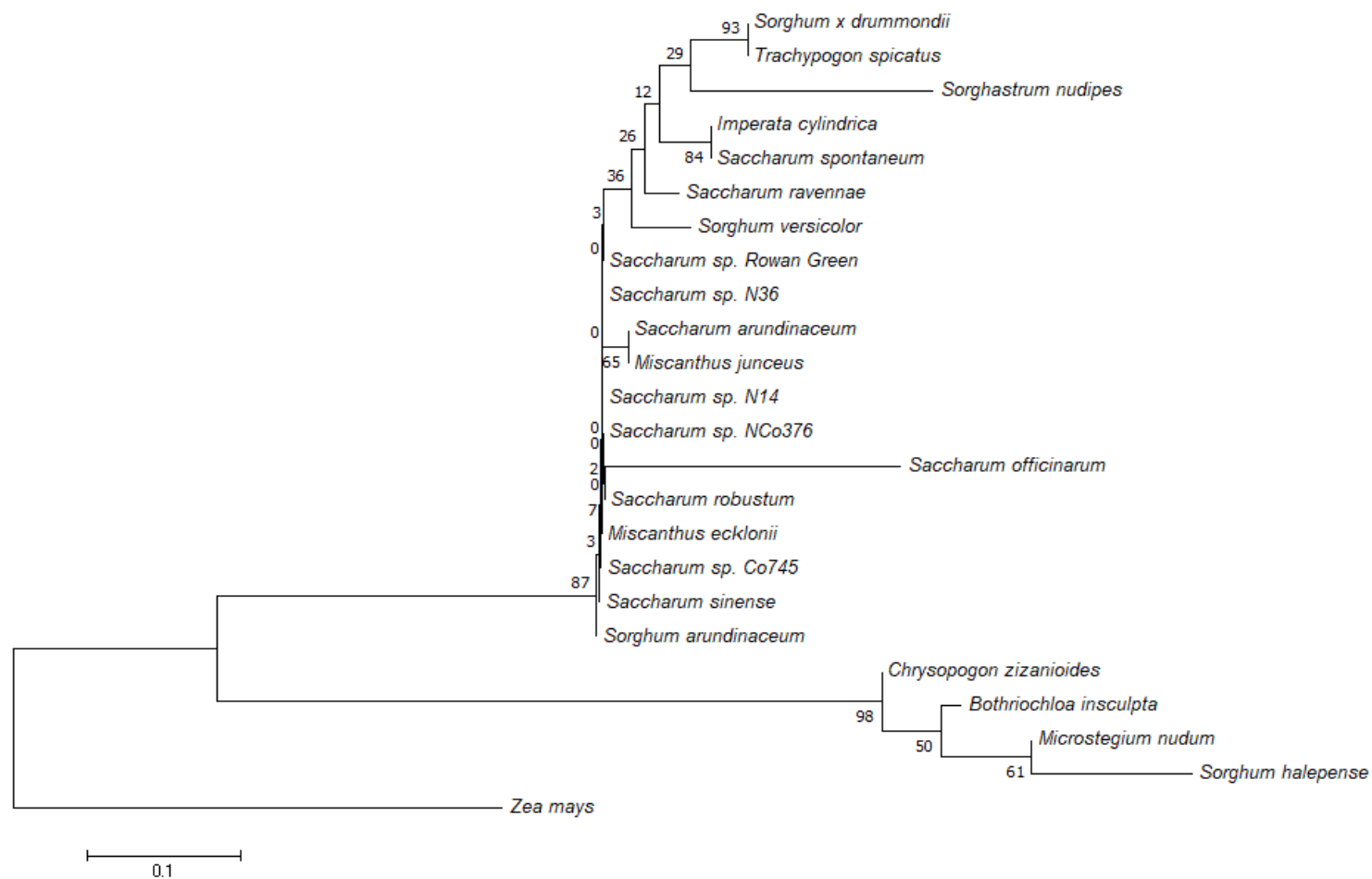


Figure 4.9: The evolutionary history was inferred using the Neighbor-Joining method based on a DNA fragment of ITS. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site

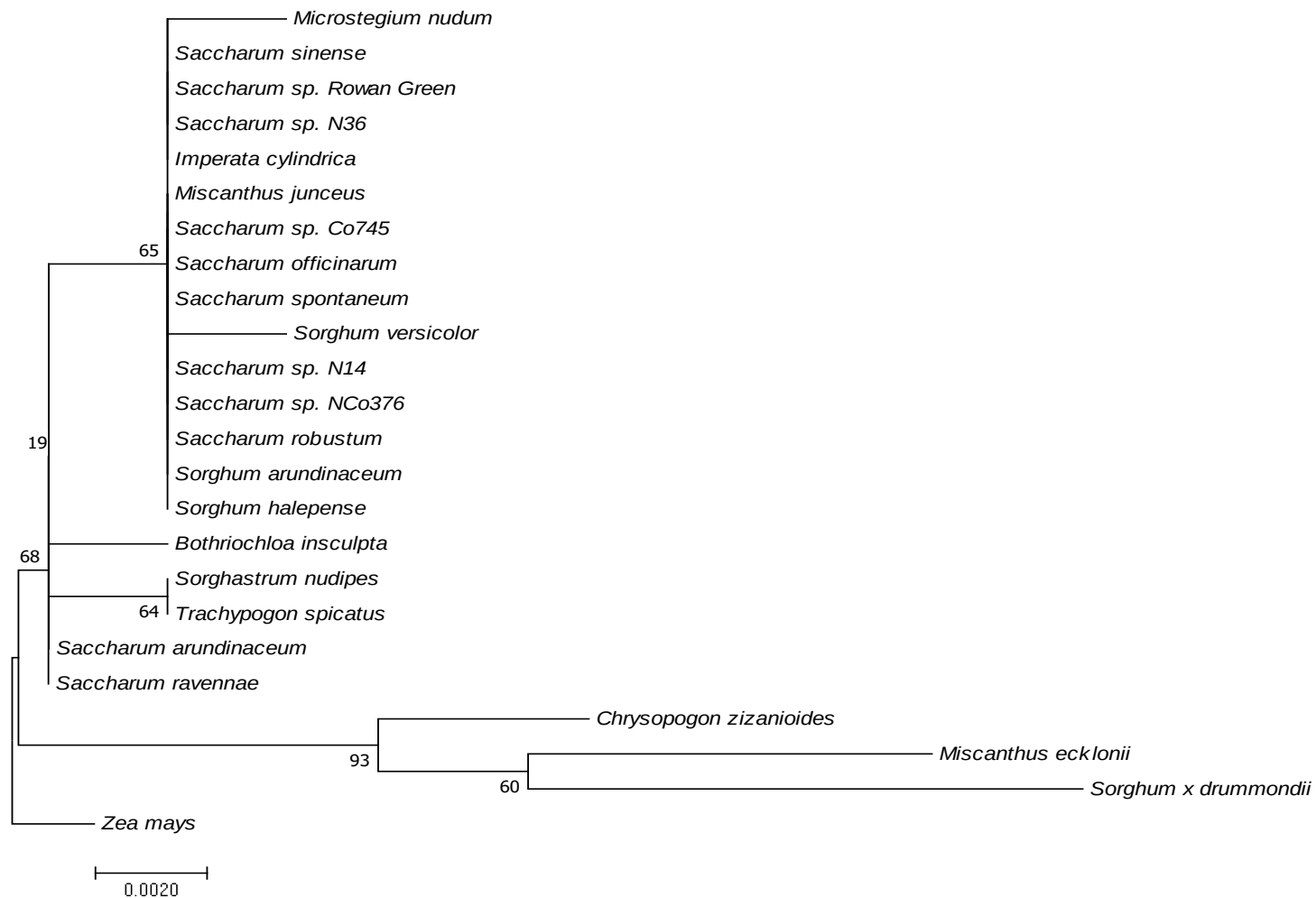


Figure 4.10: The evolutionary history was inferred using the Neighbor-Joining method based on a DNA fragment of *rbcL*. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site

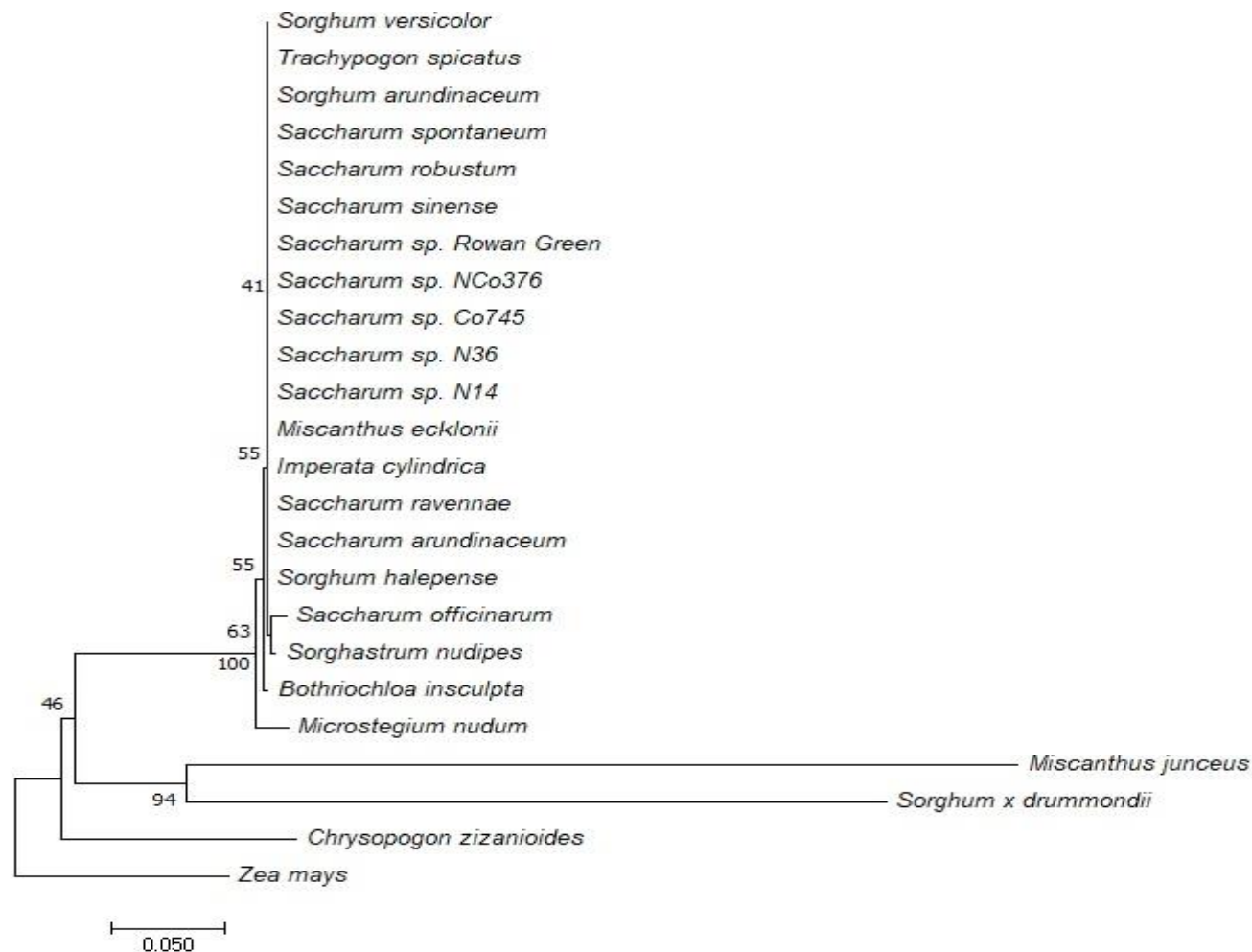


Figure 4.11: The evolutionary history was inferred using the Neighbor-Joining method based on a DNA fragment of *matK*. The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site.

4.3.5 Congruency

4.3.5.1 Maximum Likelihood method

The maximum agreement subtrees (MAST) between *rbcL* and *matK* had 12 out of 24 leaves thus indicating that the trees were more congruent than expected by chance (Table 4.8). Similarity for the core barcode and ITS, the MAST had 12 out of 24 leaves thus indicating again that the trees were more congruent than expected by chance (Table 4.9). The newick tree of the core barcode and ITS were used for the three loci barcode.

Table 4.8: The congruency analysis of the *matK* and *rbcL* gene fragments. Analysis showed trees having 24 leaves and MAST having 12 leaves.

Tree 1 (*matK*): (((((E,K,N,J,M,S,L,O,P),(G,F)),V,R,(C,Q),I,(T,W),H),(A,B,(D,U)),X);

Tree 2 (*rbcL*): (((F,U,O,H,C,S,N,P,L,K,M,J,Q,V,D),(R,W),G,I,A),(B,(E,T)),X);

Results: $I_{cong} = 1.68055702161744$ $P\text{-value} = 5.13842807191919e-05$

Table 4.9: The congruency analysis of the core barcode and ITS gene fragments. Analysis showed trees having 24 leaves and MAST having 12 leaves.

Tree 1 (*matK+rbcL*): (((((M,U,Q,S,N,P,C,O,L,J,K,V),H,D),G,I,(R,W),E,A),F,T,B,X);

Tree 2 (ITS) :((((M,O,P,J,E,K,N,L,S),(G,F)),V,(T,W),I,R,(C,Q),H),((A,B),(D,U)),X);

Results: $I_{cong} = 1.68055702161744$ $P\text{-value} = 5.13842807191919e-05$

The concatenated matrices of the core barcode did not provide any better bootstrap support for the ML tree (Figure 4.12). Instead, bootstrap support was weaker and good values (>50%) were fewer compared to the individual ML trees of these fragments. The three loci combined barcode had better bootstrap support (Figure 4.13). The branching of taxa in this method confirmed that *Sorghum x drummondii* and *Miscanthus junceus* shared a last common ancestor with the outgroups (Figure 4.12), whereas *Sorghum arundinaceum* formed a polytomy sister lineage with all the commercial sugarcane cultivars (Figure 4.13).

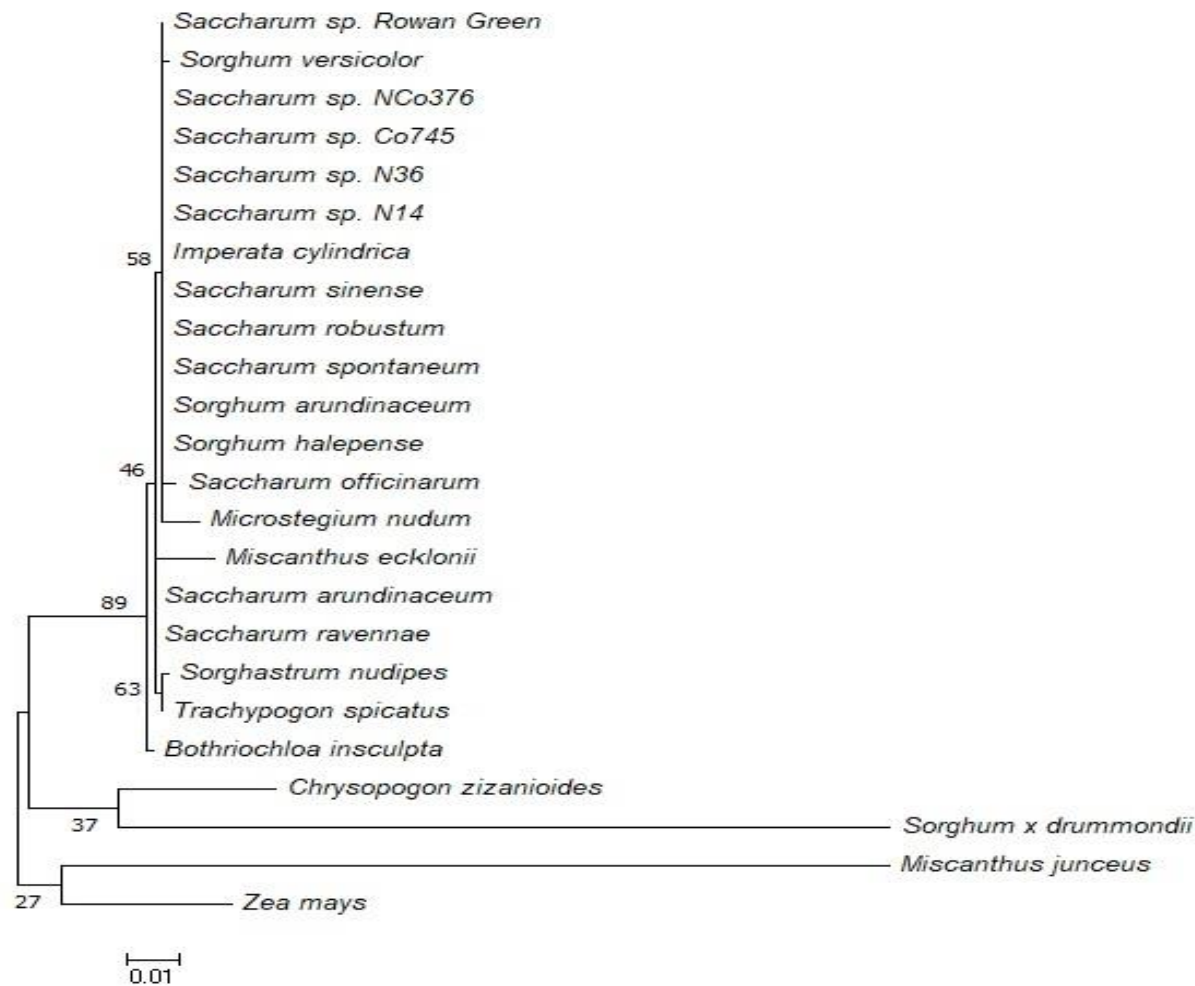


Figure 4.12: Maximum Likelihood phylogeny for *rbcL+ matK*. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches.

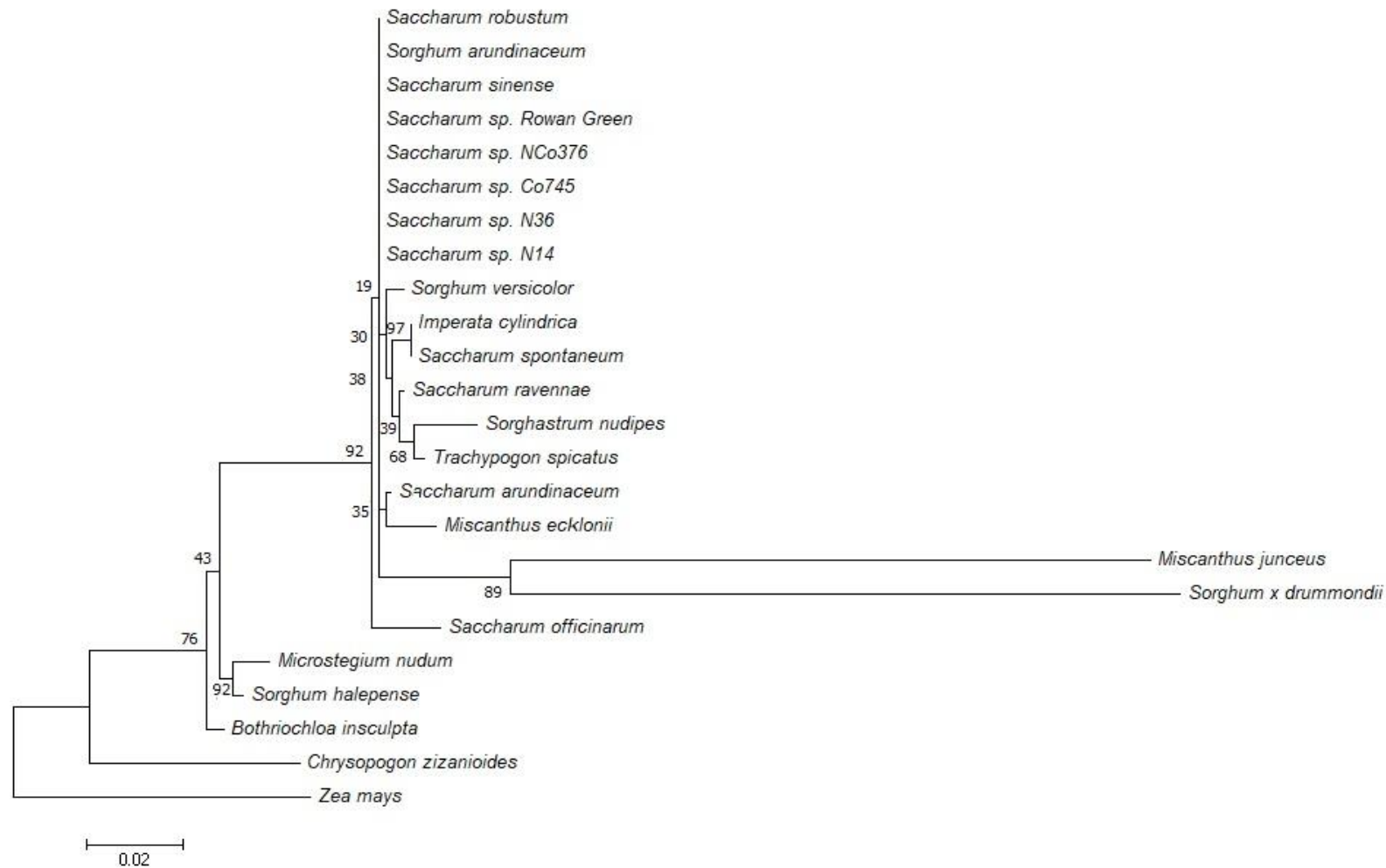


Figure 4.13: Maximum Likelihood phylogeny for *rbcL* + *matK* + ITS. The evolutionary history was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches

4.3.5.2 Neighbor joining method

The MAST between *rbcL* and *matK* had 16 out of 24 leaves thus indicating that the trees were more congruent than expected by chance (Table 4.10). Similarly with the core barcode and ITS, the MAST had 13 out of 24 leaves thus indicating that the trees were more congruent than expected by chance (Table 4.11). The newick tree of the core barcode and ITS were used for the three loci barcode. As mentioned under section 4.2.12.2, the study here onwards reports phylogeny trees constructed using Geneious R11.

Table 4.10 Congruency analysis of the *matK* and *rbcL* gene fragments. Analysis showed trees having 24 leaves and MAST having 16 leaves.

Tree 1 (*matK*): (((((E,V,M,L,O,K,C,N,P,S,G,I,J,Q,W,U,(H,R)),A),D),(F,T),B,X);

Tree 2 (*rbcL*): (((D,H,J,C,U,N,Q,L,K,F,V,O,S,M,P),A,(R,W),I,G),(B,(E,T)),X);

Results: $I_{cong} = 2.24074269548992$ $P\text{-value} = 1.05150223267933e-08$

Table 4.11 Congruency analysis of the core barcode and ITS gene fragments. Analysis showed trees having 24 leaves and MAST having 13 leaves.

Tree 1 (*matK+rbcL*): (((((V,U,S,L,K,M,P,O,J,C,N,Q,D,H),G,I,(R,W),A),E),B,(F,T),X);

Tree 2 (ITS): ((P,H,S,K,L,O,J,(G,F),E,M,N,V,(C,Q),I,(T,W),R),(B,(A,(D,U))),X);

Results: $I_{cong} = 1.82060344008556$ $P\text{-value} = 6.14575525268025e-06$

Higher bootstrap values were observed on the NJ trees constructed on the Geneious software compared to MEGA 7 (Figure 4.14.- 4.15). The consistency observed between the core barcode and three loci barcode was that *Sorghum x drummondii* and *Miscanthus junceus* both placed outside the clade grouping of the wild species and sugarcane. *Imperata cylindrica* on the other hand was either a sister taxon to the sugarcane parental lines or formed a polytomy with the commercial hybrids, both supported by high bootstraps. *Miscanthus ecklonii* and *Sorghum arundinaceum* on the three loci barcode (Figure 4.15) had strong support (97%) that they were descendants of a common ancestor shared with sugarcane cultivars, having directly branched off to form sister taxa.

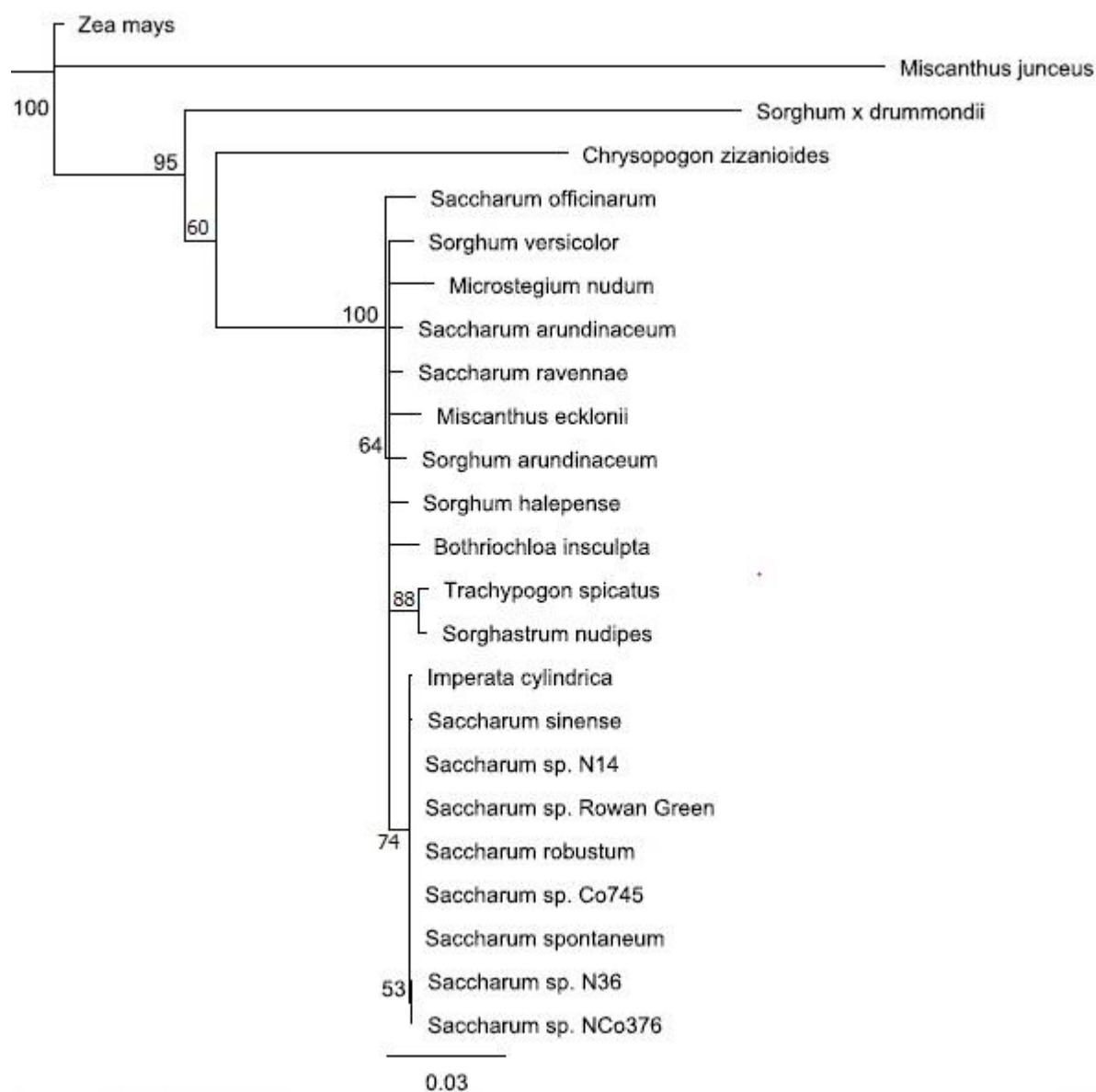


Figure 4.14: Neighbor-joining tree (Jukes-Cantor model) based on core barcode of the *matK* and *rbcL* region using Geneious R11. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

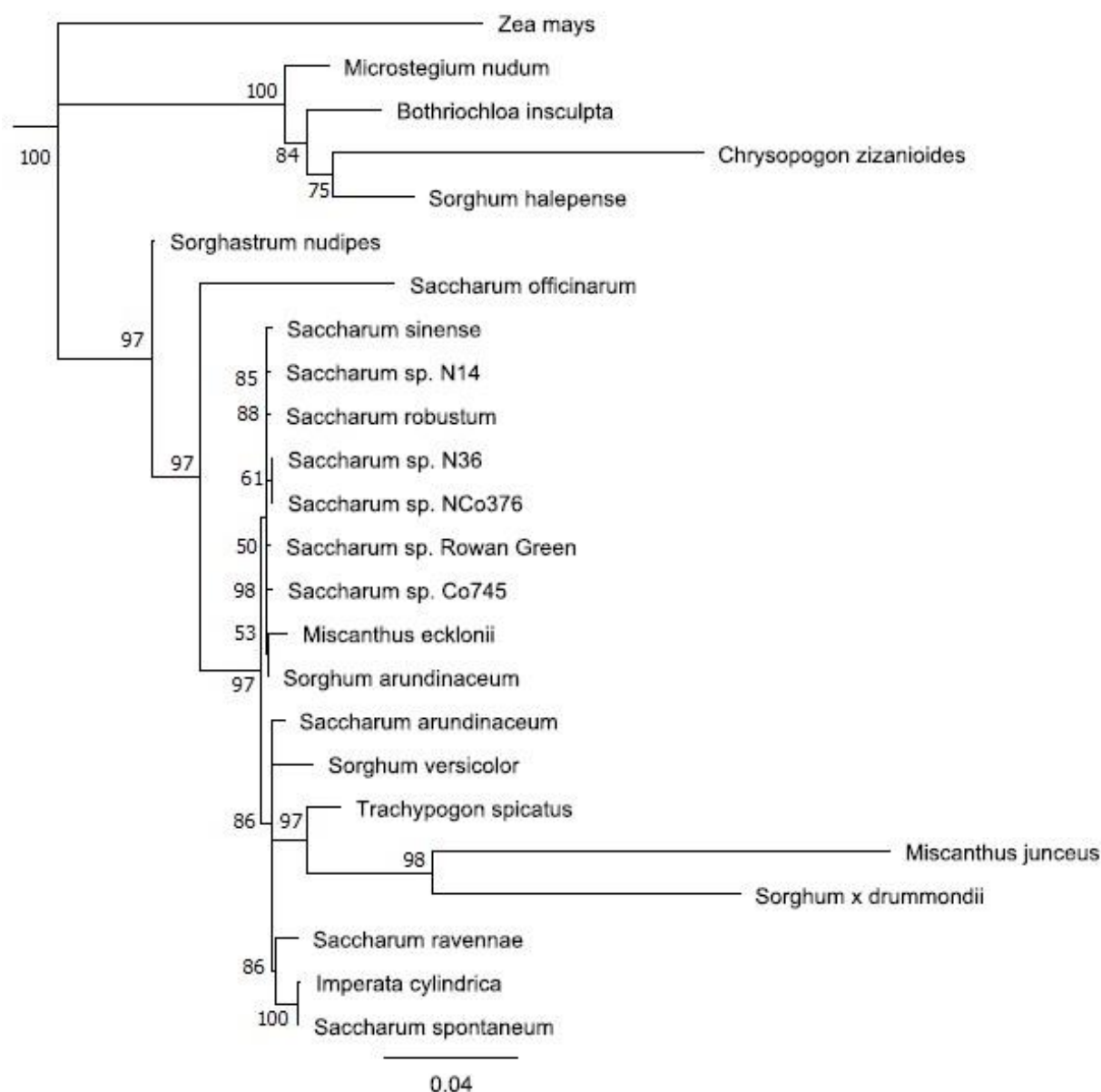


Figure 4.15: Neighbor-joining tree (Jukes-Cantor model) based on the core barcode and ITS region using Geneious R11. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. There were a total of 1,105 positions in the final dataset.

4.4 Discussion

4.4.1 Efficiency of barcodes in phylogeny construction

Based on the amount of conserved and informative sites, ITS proved to be the best barcode, followed by *matK* and *rbcL* respectively. *rbcL* was too conservative (96.27%) with sequence identity showing only a range difference of 2.05%, and thus cannot be trusted to correctly identify the closely related species within the same tribe. ITS on the other hand, was only 24.82% conservative, with sequence identity showing a range difference of 37.17%. Investigations within the Andropogoneae recommended the use of ITS regions because of their ability to incorporate changes at comparatively high rates (Bacci Jr. *et al.*, 2001). Because of the difficulty in determining phylogenetic relationships between sugarcane and the Saccharinae, Bacci Jr. *et al.* (2001) recommended supplementing ITS with other DNA regions of high evolutionary rates. The use of the core barcode as recommended by the Consortium for the Barcode of Life Plant Working Group (2009) did not improve bootstrap support nor the node resolving power compared to that observed from the use of a single locus. The integration of ITS into the core barcode however achieved this. Hence, discussion of the relatedness results was based mainly on the three locus barcode (Figure 4.15), unless stated otherwise.

4.4.2 Relatedness of the *Saccharum* complex with sugarcane varieties

The topology of both the ML and NJ tree building methods were consistent in that the *Saccharum* hybrids (N14, N36 and NCo376) frequently formed a monophyletic in-group with the parental lines (*Sa. sinense*, *Sa. robustum* and *Sa. spontaneum*). Welker *et al.* (2015) suggested that the short branches in the backbone of Andropogoneae trees are caused by rapid evolutionary radiation. Unlike distinct species with no overlap in genetic variation, closely related species present a challenge in phylogeny construction (Dai *et al.*, 2012). The lack of phylogenetic resolution (hard polytomy) in these taxa, even when a considerable amount of data was used by concatenating the three barcodes, has been interpreted as evidence of closely spaced cladogenetic events within *Saccharum*. Likewise, Arro *et al.* (2006) analysed the genetic diversity of the *Saccharum* complex and also found no distinct clustering amongst the members. Interestingly, *Saccharum officinarum* was not part of this clade, proving to be an outgroup to the entire *Saccharum* clade. This was unexpected as sugarcane varieties are said to have been developed from *Sa. officinarum* (contributing >80% genome) and *Sa. spontaneum*, followed by a backcross to *Sa. officinarum* (Arro *et al.*, 2006; Cai *et al.*, 2005b; Nair *et al.*, 2005). Meeting this expectation, Cai *et al.* (2005b) and Nair *et al.* (1999), found that

Sa. spontaneum, from which *Sa. officinarum* and *Sa. robustum* are believed to have evolved, was instead the most divergent taxon of *Saccharum*. However, in agreement with the findings of this study, Evans and Joshi (2016) concluded that *Sa. officinarum* is sister taxon to modern sugarcane varieties, with both lineages having diverged from a common ancestor 640 000 years ago, while *Sa. spontaneum* is younger than previously thought.

The contrasting groupings obtained by these abovementioned studies could be associated with the different *Sa. officinarum* varieties used in the phylogenetic analysis. For instance, the current study used *Sa. officinarum* variety IJ.76-429 which is closely related to variety IJ.76-514 from the study of Evans & Joshi (2016). Cai *et al.* (2005b) and Nair *et al.* (1999) instead used *Sa. officinarum* variety Badila, also seen forming a monophyletic group with the *Saccharum* taxa in the study done by Evans & Joshi (2016). Arro *et al.* (2006) used various varieties of *Sa. officinarum* and found that sugarcane cultivars were mostly found in clusters with *Sa. spontaneum*, which is in agreement with the findings of this study. The monophyletic relationship of variety Badila with modern cane cultivars is ascribed to it being one of the oldest examples of modern cane used for the breeding of new varieties of sugarcane (Evans & Joshi, 2016), while IJ.76-514 is a hybrid from *Sa. officinarum* and an unknown parent. Evans & Joshi (2016) pointed out that the considerable amount of evolutionary distance between sugarcane varieties and IJ.76-514 indicates a large genetic diversity within *Sa. officinarum* that has not been explored in the breeding programmes.

4.4.3 Relatedness with the *Sorghum* genus

Divergence of *Sorghum x drummondii* and *Sorghum halepense* to form a paraphyletic group with the closely related outgroups was well supported, while *Sorghum arundinaceum* instead was a descendant with the clade of sugarcane varieties as well as some parental lines. Placement of *So. arundinaceum* was further supported by a pairwise distance of <0.001 (ITS and *matK*) and sequence identity greater than 98% with the commercial lines. According to Teshome and Feyissa (2013), *So. arundinaceum* is the closest wild relative and a progenitor to the cultivated *Sorghum* species, while Dillon *et al.* (2007) states that sorghum and sugarcane are each other's closest relatives among cultivated crops. These crops are believed to have diverged from a common ancestor about 7.2 million years ago (Evans & Joshi, 2016; Tavares de Oliveira Melo, 2016). The great genetic divergence between *So. x drummondii* and *So. arundinaceum* is in agreement with Liu *et al.* (2017), who stated that the patchy signals from the chromosomes of these two species indicate high genomic divergence. *So. halepense* has been hypothesised to have originated from a spontaneous cross between *So. arundinaceum* and *So. propinquum* (Hill, 1983) while *So. x drummondii* is a hybrid between

So. arundinaceum and *So. bicolor* (Liu *et al.*, 2017; Dillon *et al.*, 2001; Skendzic *et al.*, 2007). *Sorghastrum nudipes* branching off from the rest of the in-group was well supported (97%).

4.4.4 Relatedness with the *Miscanthus* genus

The clustering of *Miscanthus ecklonii* with the *Saccharum* complex and cultivars was well supported. These findings support the conclusion of Hodkinson *et al.* (2002) and Snyman *et al.* (2018) that *Miscanthus* is a closely related species to the *Saccharum* complex. The study further states that there is high likelihood that there has been introgression between these two species. In agreement, Arro *et al.* (2006) compared the genetic diversity of *Sorghum* and *Miscanthus* against the *Saccharum* complex, and concluded that these species shared some level of synteny (conservation of the same genes) with the *Saccharum* complex. *Saccharum* and *Miscanthus* are believed to have shared a whole genome repetition before the genera diversified. A pairwise distance test applied to these two genera showed that they only differ by 1-5% (Bacci Jr. *et al.*, 2001). Furthermore, Evans & Joshi (2016) showed that *Miscanthus* was sister to *Saccharum*, while *Sorghum* was sister to both *Miscanthus* and *Saccharum*. *Miscanthus junceus* often closely aligned with the outgroup species, proving to be more divergent in this study.

4.4.5 Relatedness with the *Saccharum* genus (previously classified as *Erianthus*)

Sa. arundinaceum and *Sa. ravennae* (previously known as *Erianthus* species) formed a sister group with *Sa. spontaneum*, and a paraphyletic relationship with a closely related outgroup species, *Trachypogon spicatus*. Arro *et al.* (2006) found *Erianthus* to be closely aligned with *Sa. robustum* in a separate group from other *Saccharum* species. Likewise, a study by Sobral *et al.* (1994, cited by Nair *et al.*, 2005) showed *Erianthus* to be highly divergent from the *Saccharum* complex. Nair *et al.* (2005) further stated that this divergence was due to the independent maternal lineage of *Erianthus* compared to the whole *Saccharum* complex. With the separation of *Erianthus* from *Saccharum* in 1803 based on a morphological trait, Hodkinson *et al.* (2002) found no evidence to support this division using the ITS barcode. The intergeneric hybridization between *Saccharum* and *Erianthus* is suggested to be a reflection of their close relationship (Nair *et al.*, 2005). This hybridisation is now considered to be inter specific.

4.4.6 Relatedness with the *Imperata* genus

Imperata cylindrica was found to either group closely with *Saccharum spontaneum* or form part of a clade with sugarcane varieties. It also showed high nucleotide identity (98.07% and

96.69%) with both the *Saccharum* parent lines and sugarcane varieties respectively. Likewise, the *rbcL* and *matK* barcodes showed no pairwise distance (no base substitution) between *I. cylindrica* and sugarcane varieties. Soreng *et al.* (2015) placed *I. cylindrica* as an *incertae sedis* (cannot be assigned with confidence) within the Andropogoneae tribe. The findings of this study do not support this and show that it fits within the tribe and sister to *Saccharum*. Burke *et al.* (2016) also found strong evidence that *Imperata* was sister to a clade of *Saccharum officinarum*.

4.5 Conclusion

Generally, sugarcane was found to be closely related to its wild relatives, as the average nucleotide identity (ITS, *rbcL* and *matK*) ranged from 78-98%. The *Saccharum* species frequently formed a monophyletic clade, supporting their strong overlap in genetic variation and common ancestry. *Sorghum x drummondii*, *Miscanthus junceus* and *Sorghum halepense* constantly had significant bootstrap support to group distantly from the rest of the in-group that contains the *Saccharum* clade. Based on the three locus barcode clustering, pairwise distances and sequence identity of related species against the sugarcane varieties, *Sorghum arundinaceum*, *Miscanthus ecklonii* and *Imperata cylindrica* are the closest related species to commercial sugarcane varieties.

CHAPTER 5: POLLEN VIABILITY

5.1 Introduction

Pollen plays a vital role in facilitating gene flow in plants, especially in plants that are out-crossing. Knowledge on the ability of pollen grains to germinate upon reaching stigmas of flowers of a related species is highly useful when transgenic plants are concerned. Wind pollinated grasses like sugarcane may be able to pass on their genes to nearby populations, thus making an understanding of pollen-mediated gene flow a crucial issue prior to field release of transgenic grasses (Ge *et al.*, 2011). Sugarcane flowers are hermaphroditic with varying quantities of viable pollen within the inflorescences. This variation is dependent on the genotype and the environmental conditions prevailing during different stages of pollen development, from early development in the anther until it reaches the stigma (Melloni *et al.*, 2015; Melloni *et al.*, 2012)

Studies (Horsley & Zhou, 2013; Melloni *et al.*, 2015) have shown that environmental factors such as temperature (between 18°C and 32°C), photoperiod (between 12.5 and 13 hours), and relative humidity (>67%) have a positive influence on the emergence of inflorescences and favour pollen viability. These environmental factors have a more direct influence on pollen grain viability after pollen has been released from the anther. Soil moisture content is another important factor that affects sugarcane flowering by suppressing tassel formation when inadequate (Melloni *et al.*, 2015). According to Horsley & Zhou (2013), while environmental gradients may have the most direct influence, external stresses prior to anthesis also affect pollen viability. Therefore, this part of the study assessed pollen viability of sugarcane varieties in commercial fields and relates these to the above mentioned variables.

5.2 Materials and methods

5.2.1 Plant material

According to Amaral *et al.* (2013), sugarcane pollen grains have a half-life of only 12 minutes, making it difficult for them survive more than 35 minutes at a temperature greater than 26°C and humidity lower than 67%. The inflorescence successively matures in every one third from the apex to the base, thus pollen was collected from mature anthers in dehiscence. Three replicates of inflorescence from each variety were collected per sugarcane commercial hybrid per site during the flowering season month of July, in the years 2016 and 2017. Pollen is known to be very sensitive to dehydration by exposure to high temperatures and low humidity, and the inflorescence collection was thus done between 6:00 and 8:30 a.m.

5.2.2 Sampling sites

Pollen viability tests took place at various sites in sugarcane production regions where the commercial varieties were flowering. A total of nine sites were visited; of these, two fall in Mpumalanga and seven are in KwaZulu-Natal (Figure 5.1). Sampling sites were visited based on observed sugarcane flowering by SASRI extension specialists and biosecurity officers. Personal communication with a SASRI extension specialist (Mr Tshifhiwa Rhadzilani,) indicated that there was no official commercial cultivation in Limpopo. In addition, a crop production officer at the Department on Agriculture and Rural Development in Limpopo, Tzaneen area (Miss Dinah Shikwambana), confirmed regular small scale cultivation of sugarcane on plots less than one hectare. However, the Limpopo sugarcane was not in flower during July 2017 and was already being harvested at the time of the survey.

In the first year of sampling (2016), six sites were sampled i.e. - Malelane (first site) and Komatipoort (second site) in Mpumalanga, and Jozini (third site), Pongola (fourth site), Empangeni (fifth site) and Mount Edgecombe (sixth site) in KwaZulu-Natal (Figure 5.1). To enhance repetitions of the previous sampling year and to gather more pollen viability data in the 2017 follow-up survey, three new sites were added. In 2017, six of the sampling sites were the same as those of the previous year and the new sites included Mtubatuba (seventh site), Umhlali (eighth site) and Port Shepstone (ninth site) in KwaZulu-Natal (Figure. 5.1). Numbering of these sites was arranged according to their geographical location, starting from the northern inland regions to the southern coastal regions (1 to 9) so as to aid in the analysis of environmental gradients across the study area. Sites 1-5 are irrigated crop regions while sites 6-9 are rain fed sugarcane crops.

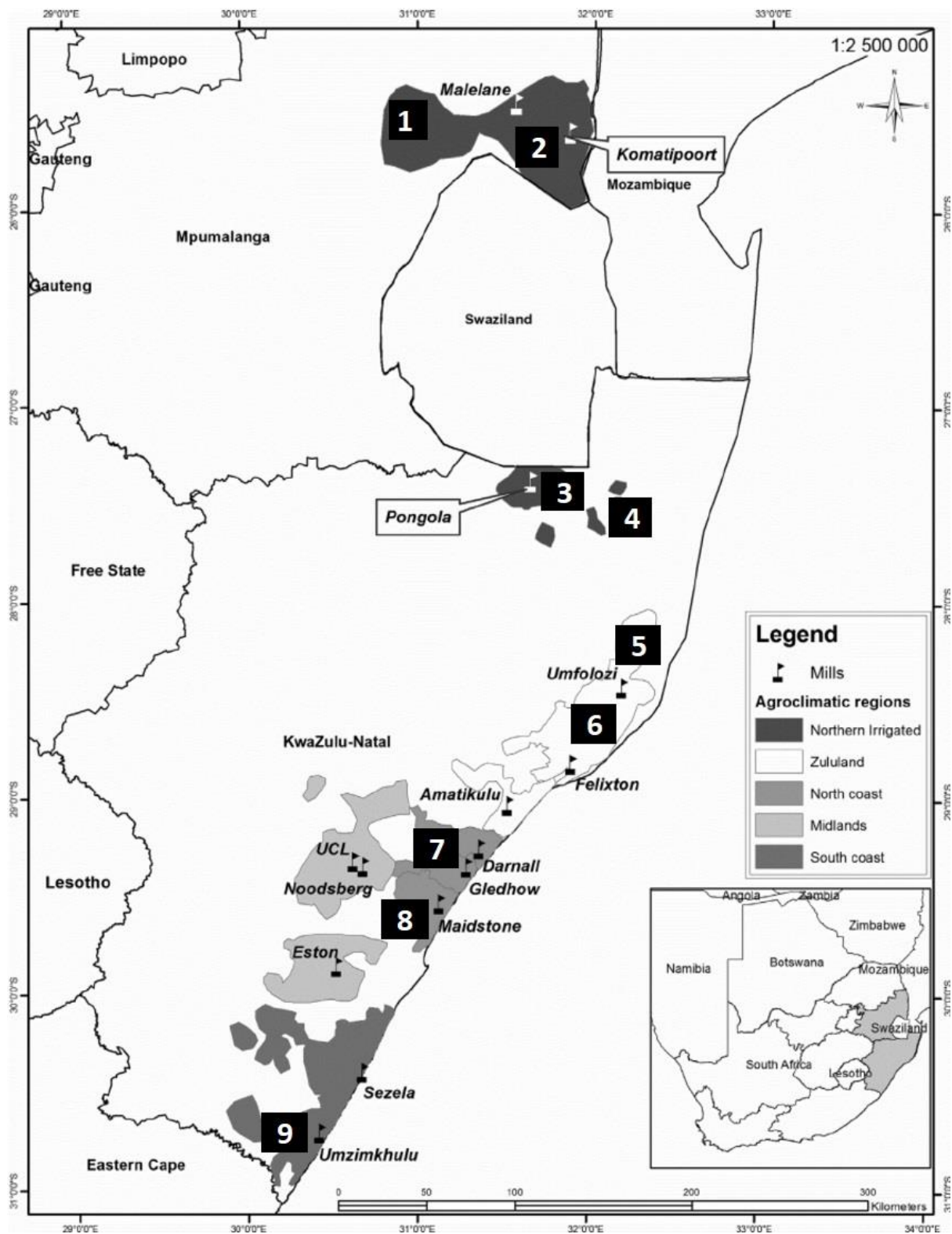


Figure 5.1: Sugarcane production regions and locations of sugar mills in the Mpumalanga and KwaZulu-Natal Provinces. See text for explanation of sites 1-9. (Map produced from data provided by SASRI)

5.2.3 Stains used to test for pollen viability

Amaral *et al.* (2013) recommended that panicle transportation from the field to the lab should not exceed 30 minutes considering the half-life of pollen to be 12 minutes, and therefore it was not possible to conduct germination tests in this study. Pollen viability was determined using two different staining techniques in the field, TTC (2,3,5- triphenyl tetrazolium chloride) and IKI (iodine potassium iodide). TTC cell viability test detects the activity of a dehydrogenase enzyme during respiration (Soares *et al.*, 2013b). If an active enzyme is present it will reduce the tetrazolium chloride to formazan, a reddish, water insoluble compound. It is based on the principle that a decrease in cell viability is always accompanied by a decrease in tetrazolium salt reduction (Dhanda & Munjal, 2006). Thus, viable pollen grains stained pink (Figure 5.2).

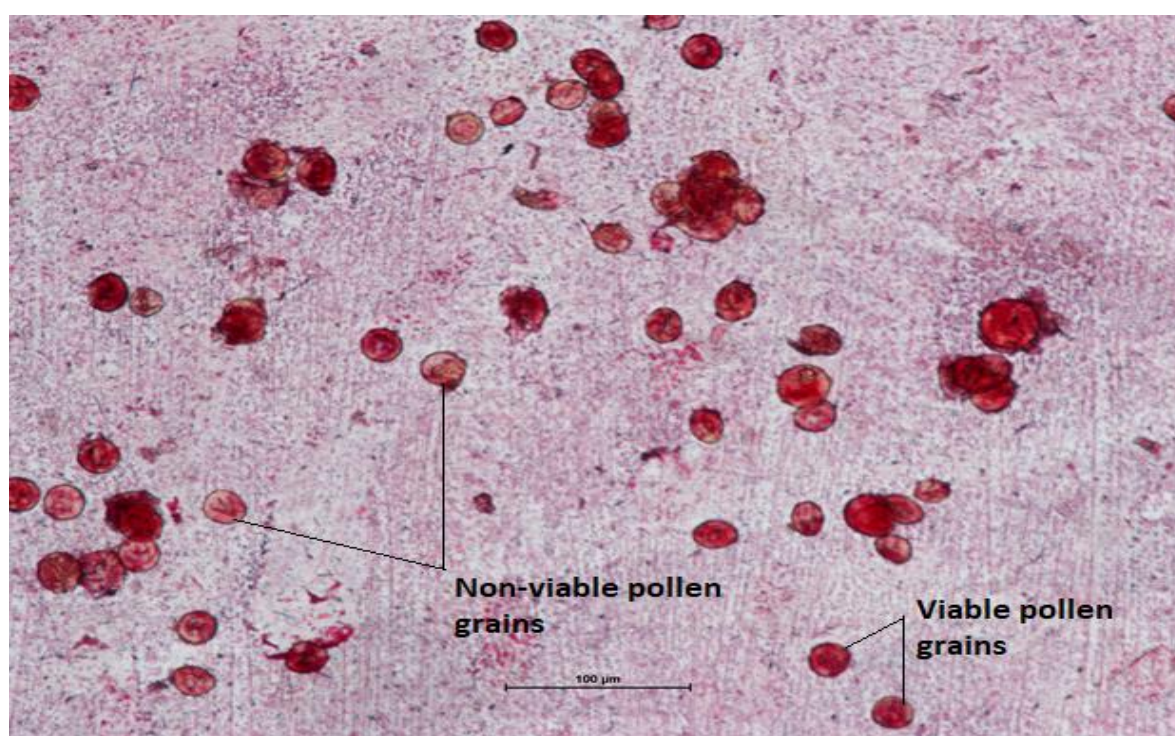


Figure 5.2: Viable and non-viable pollen grains of sugarcane observed using the TTC staining method. Nikon AZ100 M Stereoscopic microscope (Nikon Instruments Inc.) 100X magnification.

On the other hand, the IKI stain detects the presence of starch. Starch is recognised as a metabolite on reserve within a pollen grain, which is to be later used as source of energy for germination and formation of the pollen tube (Oliveira *et al.*, 2015). When a starchy pollen grain is treated with the IKI solution, the iodine molecule will insert inside the coiled amylose structure. This results in the formation of polyiodide chains which produce an intense brown-dark colour. Thus viable pollen grains stain black (Figure 5.3a) while non-viable pollen grains remained pale yellow/brown (Figure 5.3b).

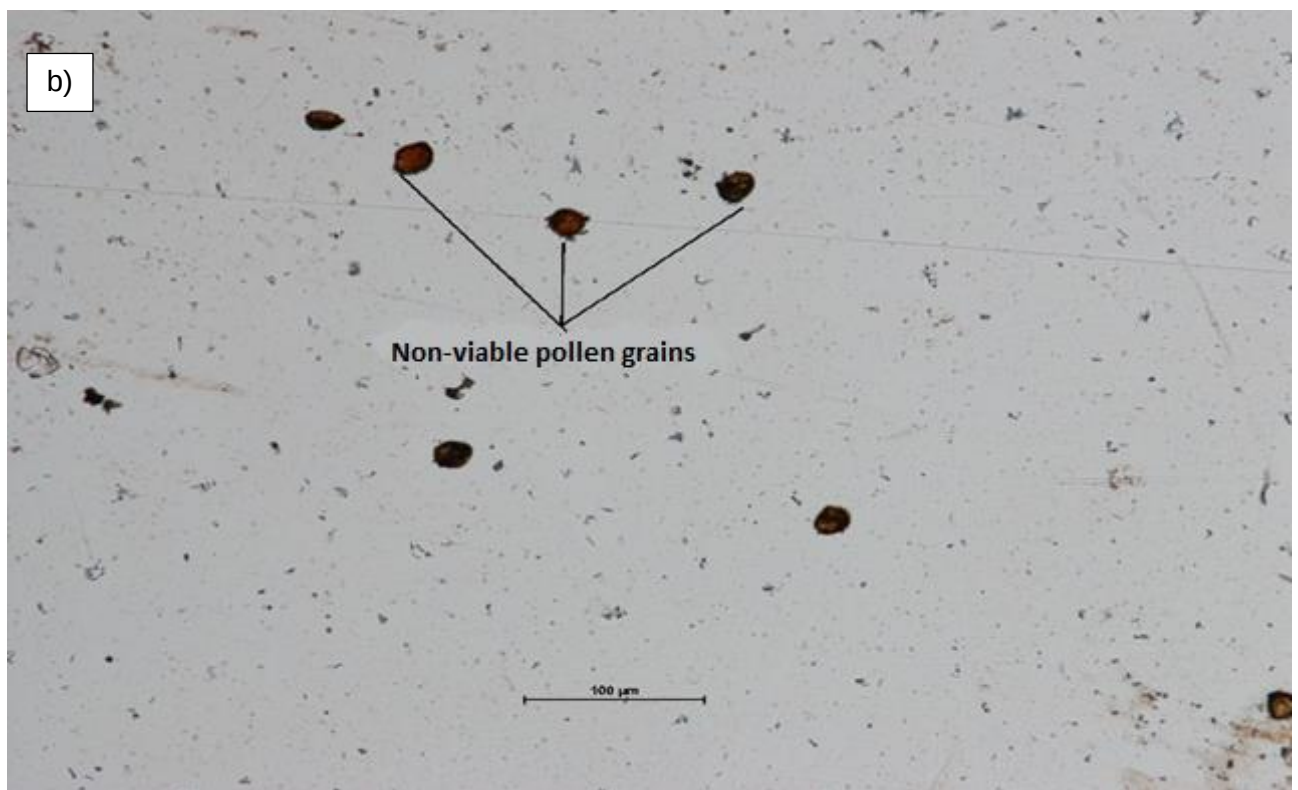
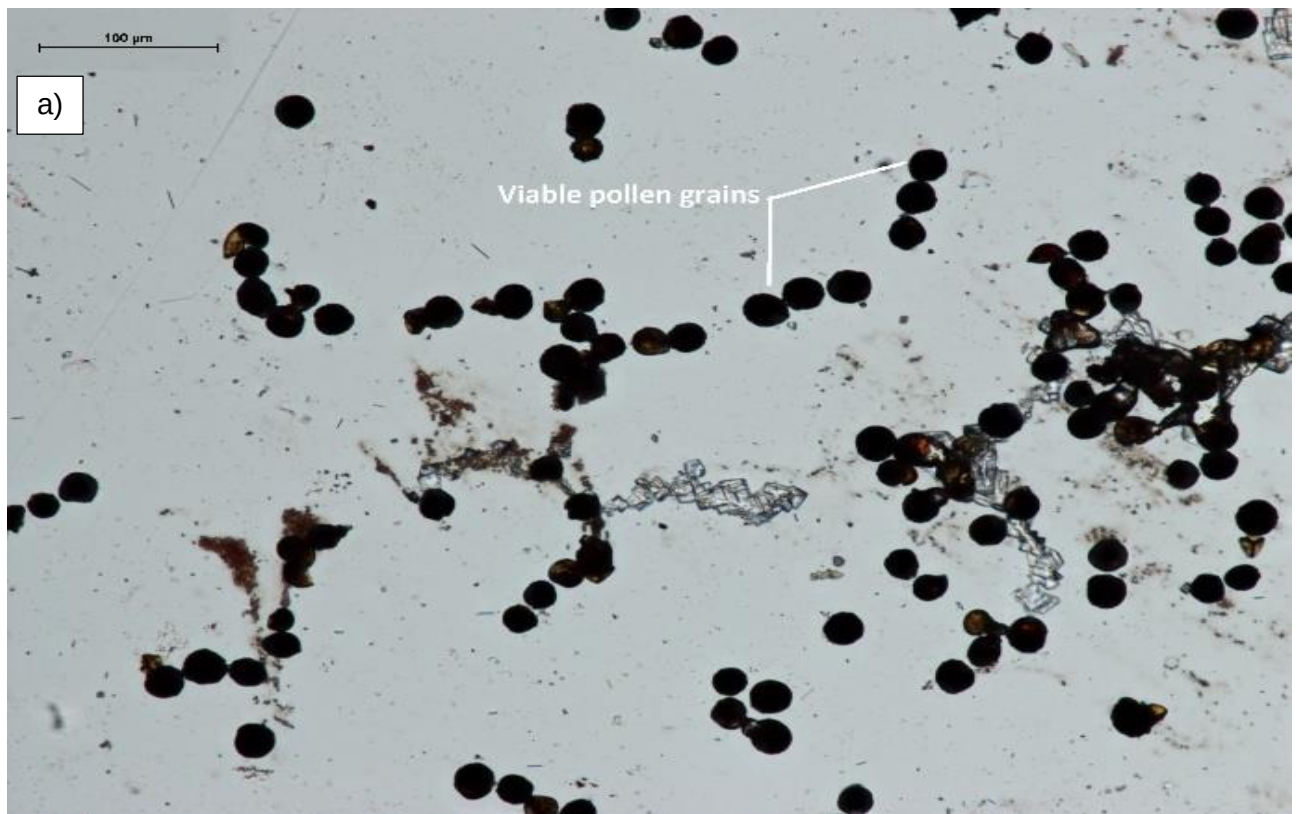


Figure 5.3: Starch test using the IKI staining method, where (a) viable and (b) non-viable pollen grains of sugarcane were observed. Nikon AZ100 M Stereoscopic microscope (Nikon Instruments Inc.) 100X magnification.

5.2.4 Testing pollen viability using staining methods

Freshly harvested mature anthers were dissected and pollen was released onto a glass slide with grids. Two to three drops of either IKI (1% (w/v) iodine and 2% (w/v) potassium iodide) or TTC (1% (w/v) TTC and 5% (w/v) sucrose) (Huang *et al.*, 2004; Soares *et al.*, 2013b) stains were added to the pollen grains and thereafter a coverslip was placed on top. Both methods consisted of the addition of the stain on pollen and observation under compound microscope (Wild, Heerbrugg Switzerland Model 11) at 100X magnification. Pollen stained with IKI was observed after ± 5 minutes of incubation, while pollen grains stained with TTC were examined after ± 15 minutes of incubation in direct sunlight (Huang *et al.*, 2004) to allow the reduction of tetrazolium chloride. The magnification was set at 100x to allow the viewing of slides at less detail but greater proportion, in order to have a clear view when counting viable against non-viable pollen grains. A haemocytometer grid was fixed under the glass slide from which the arrangement of the squares allowed easy tracking during the counting of pollen grains. A random count of a minimum of 100-150 pollen grains was done per each variety replicate, and the viability percentage was determined as the ratio of viable pollen grains (intense dark for IKI and deep pink for TTC) over the total number of grains for the count. An average was calculated from the three replicates per variety per study site and then used as pollen viability percentage. The whole viability test procedure (staining, visualisation and counting) was completed in the field immediately after each inflorescence sample was collected.

5.2.5 Weather data

Weather data including relative humidity (RH), soil water content at 100mm depth (SWC 100), minimum (TMN) and maximum temperature (TMX) was extracted from the SASRI weather web, accessible from the SASRI home page (<http://www.sasa.org.za/sasri>). Each of the automatic weather stations was situated at each of the experimental sites. Time resolution was set to be daily for the period from 1st of May up to the day at which sampling took place for each of the sites visited (e.g. 4 July 2016 for Malelane). The weather data was downloaded in an excel format, from which mean values were computed for each environmental variable. Day length data (DL) with the same time resolution and period was downloaded online from a reliable time zone-related website, namely [Time and Date AS \(www.timeanddate.com\)](http://www.timeanddate.com).

5.2.6 Data analysis

All statistical analyses were carried out using Statistica (version 13; Dell Inc., USA). The Kolmogorov-Smirnoff and Lilliefors test for normality showed that the data was not normally distributed. The Kruskal-Wallis non-parametric test was therefore selected to analyse the variance by mean value and standard deviation. This test does not require the fulfilment of assumptions of normal distribution, interval data and homogeneity of group variance (Ostertagová

et al., 2014). A p -value <0.05 indicates significant statistical difference in the group means (Appendix D1-5).

An association between the environmental variables and staining methods was performed by utilizing Spearman's rank correlation coefficient. This technique assesses how well a random monotonic function can describe the association between two variables, without assumptions about the frequency distribution of the variables (Hauke & Kossowski, 2011). A correlation was determined between each stain and RH, DL, SWC 100mm, as well as the TMN and TMX as an environmental gradient. A Spearman correlation of +1 or -1 is obtained when one of the variables is a perfect monotone function of the other. According to Hauke & Kossowski (2011), it is important to view the Spearman's correlation together with results of significance.

Dissimilarities in pollen viability between sampled sites were analysed using ordinations to analyse the contribution of environmental variables towards variation in pollen viability in a single analysis. Principal Component Analysis (PCA) (Software package CANOCO 4.5) is a multivariate linear transformation method, in which the variance of the variables is distributed along derived ordination axes (Kenkel, 2006). Unlike other ordination methods, PCA does not reduce dimensionality by filtering non-dominant axes (Kenkel, 2006; Van den Brink *et al.*, 2003). All the response variables were centred and standardised to compensate for unit differences in the PCA. Ordinations were interpreted using the following rationale: Parameters are i) positively correlated with each other if their vectors subtend a small angle, ii) not correlated if their vectors are arranged at an angle of greater than 90° apart, iii) negatively correlated if their vectors are directed oppositely (180°) and iv) parameters with the longest vector relative to an axis have the greatest influence on that axis. The same principle applies for the clustering of sampling sites, with respect to each other as well as the variables.

5.3 Results

5.3.1 Pollen viability of sugarcane varieties

A total of 11 different sugarcane varieties were tested for pollen viability during 2016, from six sites in the study area. Three varieties were from Mpumalanga and ten from KwaZulu-Natal, of which three out of the ten were the same varieties sampled from Mpumalanga. With the TTC staining method, the highest mean viability percentage (62.5%) was determined for the commercial sugarcane variety N36 at site 1 in Mpumalanga (Figure 5.4). With the IKI staining method, the highest mean viability percentage (55.4%) was determined for the commercial sugarcane variety N36 at site 4 in northern KwaZulu-Natal. Additionally, variety N36 was also found to have the most viable pollen in all the sites it was studied at, i.e. sites 1, 2 and 4. This was followed by the variety N14 which showed maximum mean pollen viability of 55.0% and 46.2% at site 2 for IKI and TTC stains respectively. All the other varieties (N19, N23, N25, N27, N28, N41, N42, N43 and NCo376) had lower mean viability percentages ranging from 0 to 11.6% with IKI and 0 to 7.6% for the TTC staining method.

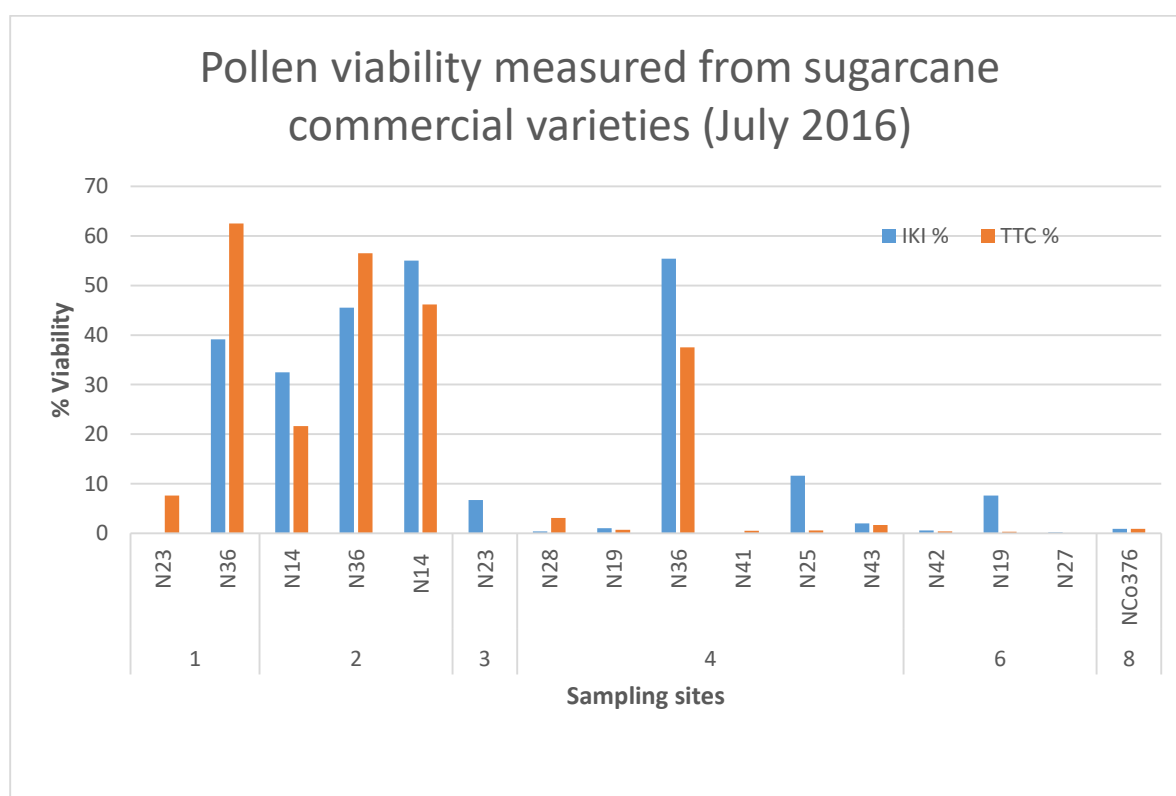


Figure 5.4: Mean pollen viability of 11 sugarcane varieties measured across six study sites in July 2016. Site 1: Malelane; Site 2: Komatipoort (Mpumalanga); Site 3: Pongola; Site 4: Jozini; Site 6: Empangeni; Site 8: Mount Edgecombe (KwaZulu-Natal).

Pollen viability tests during 2017 added an additional two varieties (N39 and N58) from Site 9, making it 13 varieties tested during this study. With TTC staining, the 2017 study showed that the highest mean viability of 84.6% was again observed for N36, closely followed by N14 at 83.8% (Figure 5.5). In contrast, with IKI staining the highest mean viability of 64.6% was measured for N14, closely followed by N36 at 63.5%. Variety N14 and N36 were found to have the highest pollen viability at sites 1, 2, 4 and 5. Varieties N23, N42, N58 and NCo376 were generally found to have low viability (less than 30%). Pollen viability showed a gradual decrease from the northern to the southern regions of the study area.

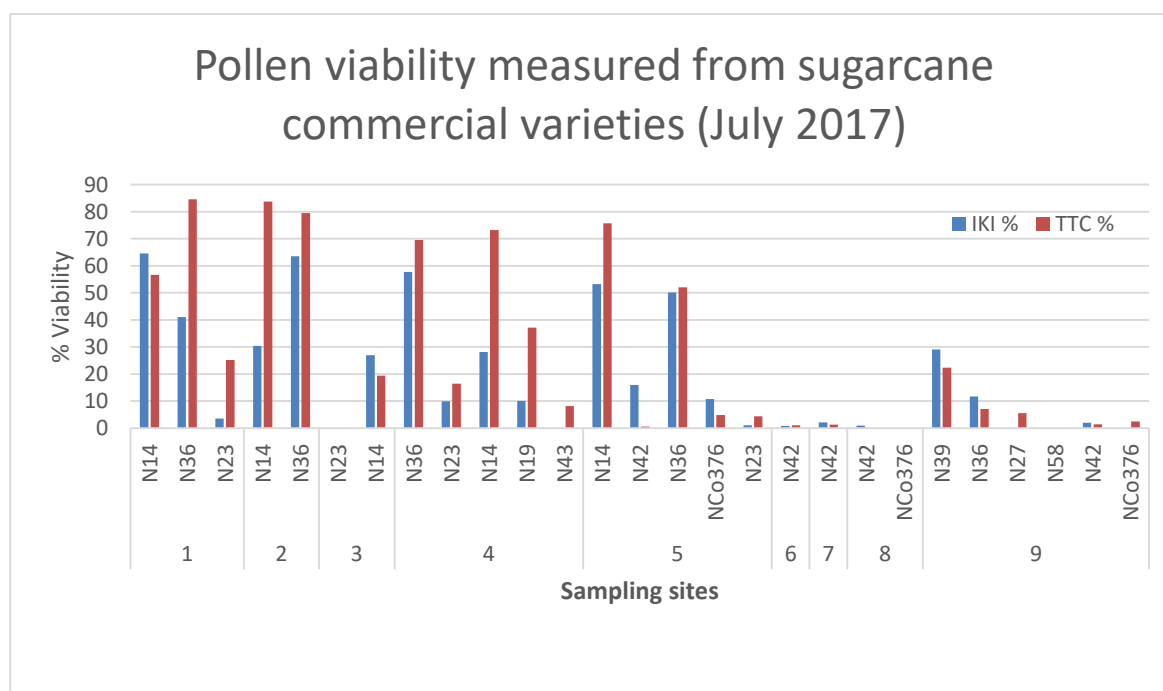


Figure 5.5: Mean pollen viability of eleven different sugarcane varieties measured across nine study sites during the month of July 2017. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

5.3.2 Variance between the staining methods

Out of the 43 individual tests carried out for each sugarcane variety, the TTC stain measured higher pollen viability than IKI in 17 test results (39.5%), while the IKI stain measured higher pollen viability in 16 (37.2%) instances. Pollen viability percentage from the two stains was equal for ten (23.3%) test results. A Kruskal–Wallis ANOVA test for statistical difference revealed that there was no significant difference between varying pollen viability percentage obtained from the two staining methods, with p -value of 0.622 (Figure 5.6). The study will thus only report results obtained from the TTC staining method from here onwards.

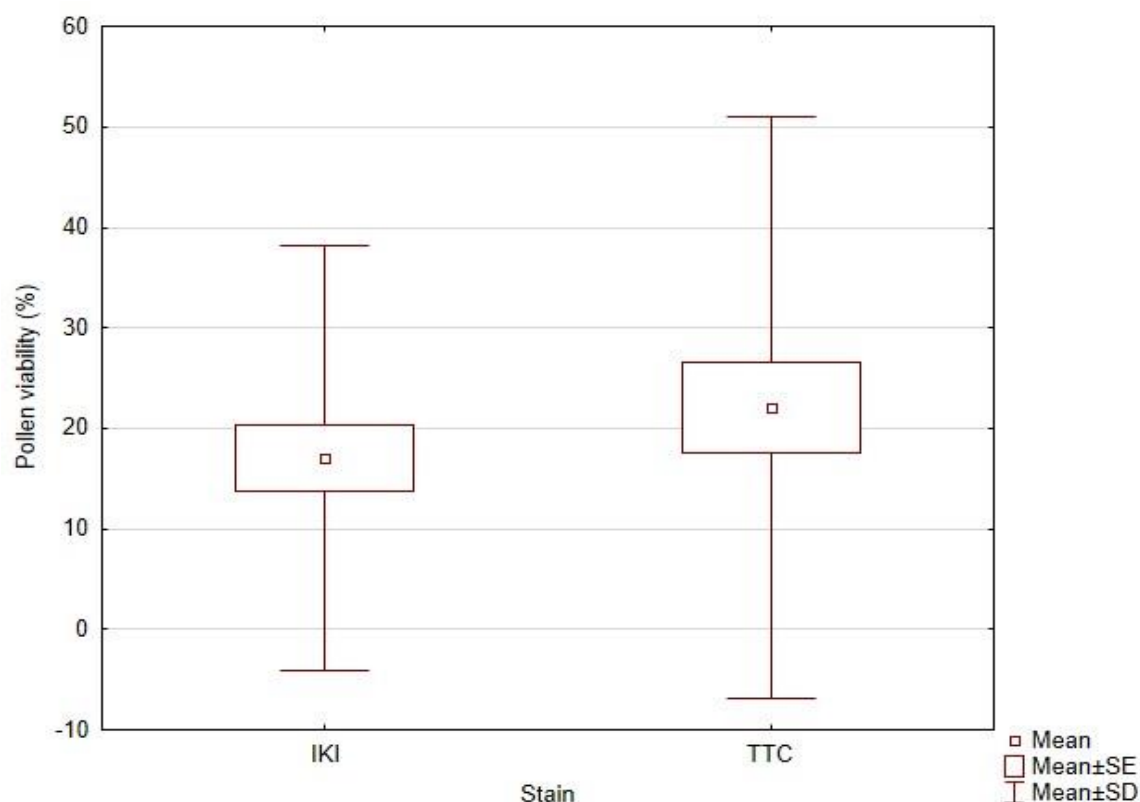


Figure 5.6: Box and whiskers plot of the mean pollen viability of sugarcane tested for the IKI (Iodine potassium iodide) and TTC (2,3,5- triphenyl tetrazolium chloride) staining methods. $\pm$ Standard error (SE) and standard deviation (SD), $n=43$.

5.3.3 Flowering sugarcane varieties and geographic range of pollen viability

Pollen viability is classified as low, intermediate and high by considering the percentage of viable pollen observed for each sugarcane variety within study sites 1-9 in 2016 and 2017 (Figure 5.7). An average pollen viability value was used when a variety was tested more than once at the same study site (variety N14 at site 2 during 2016). Varieties with pollen viability above 70% were classified as high and the variety was considered a potential pollen donor (male), varieties with 30-70% pollen viability were classified as intermediate and can either be pollen donor or receptor, while varieties with viability less than 30% were classified as low and the variety would then be a pollen receptor (female) (Amaral *et al.*, 2013).

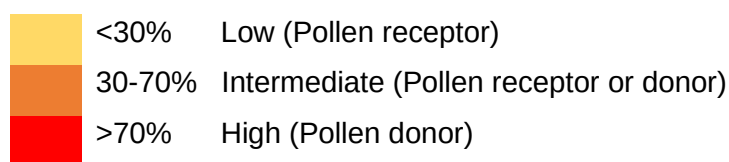


Figure 5.7: Colour codes assigned to each of the classes for pollen viability.

Site 4 (Jozini) had the most sugarcane varieties in flower (eight), followed by sites 9 and 5, each with six and five flowering varieties respectively (Table 5.1). Site 7 only had one variety flowering in 2017. Site 2 showed high pollen viability for N14 and N36 sugarcane varieties. Sites 1 and 5 each showed the highest variation in viable pollen from the different sugarcane varieties tested, i.e. low to high classes. Site 4 had pollen viability ranging from low to intermediate. Pollen viability was measured to be low for sites 3 and 6-9.

Pollen viability was found to sometimes exceed 70% for sugarcane varieties N14 and N36. Variation in pollen viability was found to range from low to high for these two varieties, indicating that pollen viability differs greatly from one site to another. Intermediate pollen viability was observed for N14 (two out of seven tests), N36 (five out of eight tests), as well as for N19 (one out of two tests). Pollen viability of variety N19 was found to only range from intermediate to low between two sites across the years 2016 and 2017. Varieties N23, N25, N27, N28, N39, N41, N42, N43, N58 and NCo 376 had low pollen viability at all the sites they were tested at.

On average, sites 1, 2, 4 and 5 of the irrigated regions showed that N14 and N36 are potential pollen donors. The highest pollen viability (>70%) was measured at sites 2 and 5 for N14 while for N36 it was measured at sites 1 and 2. Intermediate pollen viability percentages were measured at site 4 for N14, and at sites 4 and 5 for N36. Within these sites, and out of the seven pollen viability tests for N14, four showed high pollen viability while two showed intermediate pollen viability. Likewise with N36, and the seven tests conducted under the irrigated regions, four showed high pollen viability while three showed intermediate pollen viability.

Table 5.1: Pollen viability percentage classes (low, intermediate or high) for 2016 and 2017 for sugarcane varieties and sites. Sites 1: Malelane and 2: Komatipoort (Mpumalanga), sites 3: Pongola, 4: Jozini, 5: Mtubatuba, 6: Empangeni, 7: Umhlali, 8: Mount Edgecombe and 9: Port Shepstone (KwaZulu-Natal). The sum of tests conducted for each variety within a site is shown in each block.

Variety	No. of tests	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
N14	7	1	3	1	1	1				
N19	3				2		1			
N23	6	2		2	1	1				
N25	1				1					
N27	2						1			1
N28	1				1					
N36	8	2	2		2	1				1
N39	1									1
N41	1				1					
N42	6					1	2	1	1	1
N43	2				2					
N58	1									1
NCo376	4					1			2	1
Total varieties		3	2	2	8	5	3	1	2	6

5.3.4 Environmental variability across sites

5.3.4.1 Minimum temperatures

Minimum temperatures during the flowering season ranged from 11-15.5°C across the study area (Figure 5.8). Site 5 had the highest minimum daily temperature, which was significantly higher ($p < 0.05$) than the lowest minimum daily temperature at site 9 (Appendix D-1). Furthermore, sites 6 and 7 also had significantly higher daily minimum temperatures than that of site 9, indicated by p -values of 0.035 and 0.003 respectively. Generally, daily mean minimum temperature increased from site 1 to 7, which is from the northern inland to south eastern region of the study area. In contrast to this trend, mean minimum temperature decreased between sites 8 and 9.

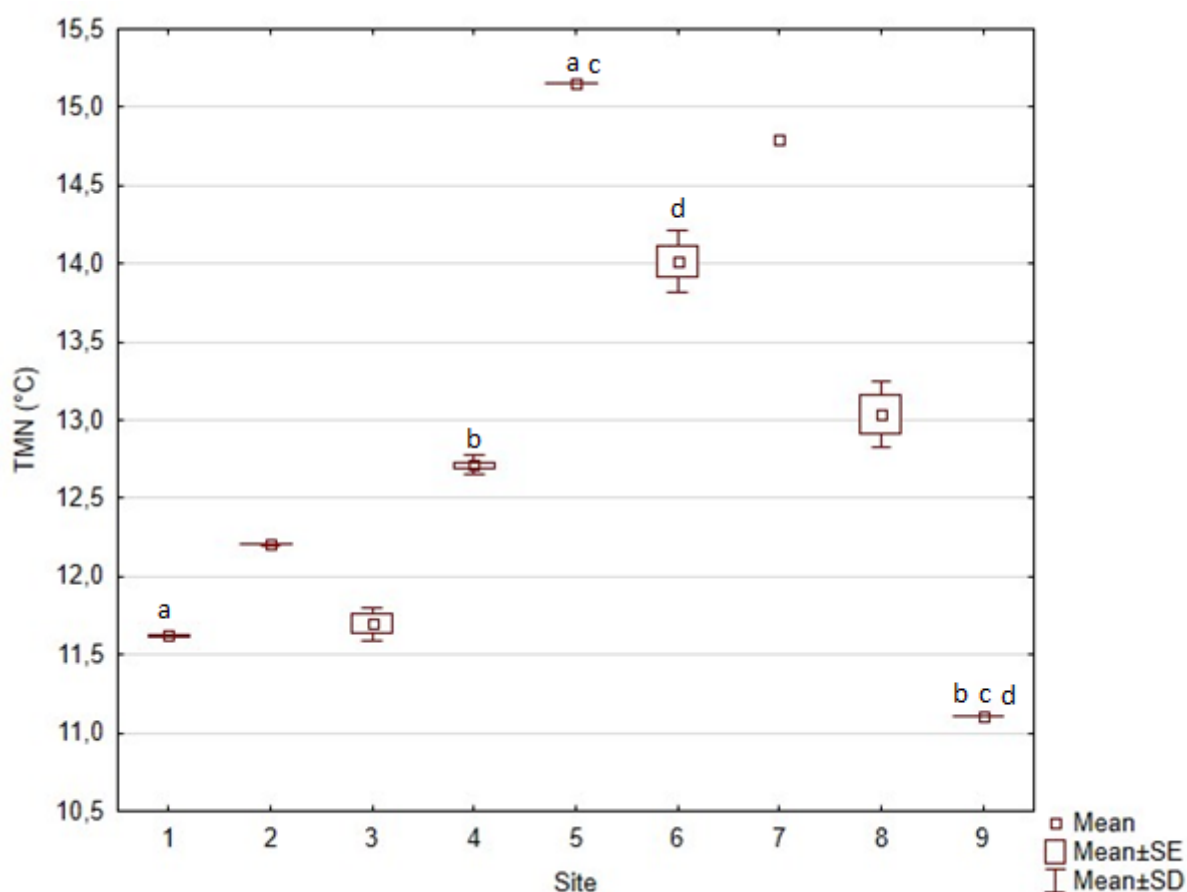


Figure 5.8: Box and whiskers plot showing the mean minimum temperature (TMN) among different sites, $\pm$ standard error (SE) and standard deviation (SD) $n=43$. A $p<0.05$ indicated that there was a significant difference in daily minimum temperatures between the sites highlighted in Appendix D-1. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

5.3.4.2 Maximum temperature

Mean maximum daily temperature during the flowering season (May-July) of commercial sugarcane ranged from 23-28.5°C (Figure 5.9). Site 1 had the highest mean maximum daily temperature while site 9 had the lowest. Site 9 had significantly ($p<0.05$) lower maximum temperatures than sites 1, 2 and 4, supported by p -values of <0.001 , 0.004 and 0.018 respectively (Appendix D-2). The mean maximum temperature generally decreased from the northern to the south eastern extent of the sugarcane cultivation region (sites 1-9). Site 8 was outlying from this trend. Sites 5, 7 and 9 showed little variation from the mean maximum temperature, while the

other sites showed a deviation of less than 1°C both above and below the mean value. This means that maximum daily temperatures were constant throughout the flowering season.

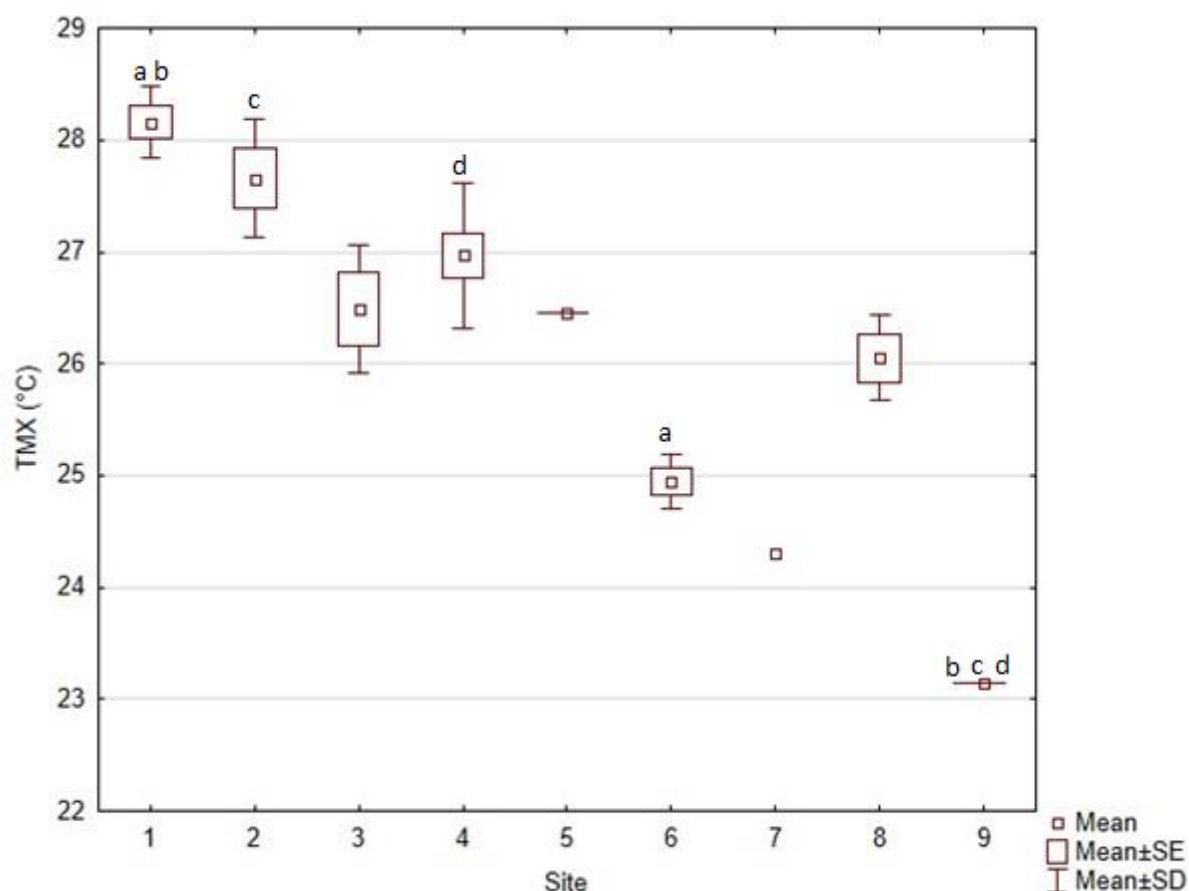


Figure 5.9: Box and whiskers plot showing the mean maximum temperature (TMX) across the study sites, $\pm$ standard error (SE) and standard deviation (SD) $n=43$. A p -value <0.05 indicated that there was a significant difference in maximum temperatures between the sites highlighted in Appendix D-2. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

5.3.4.3 Day length

Mean day length decreased from 10.75 to 10.35 hours from sites 1-9, indicating that none of the sites have a mean day length greater or equal to 12.50 hours that is favourable to pollen viability. Site 2 had low standard deviation of less than 0.07 hours from the mean, while all the other sites showed only a slight deviation from the mean value (Figure 5.10). Site 1 and 2 had significantly

longer DL than sites 8 and 9 (Appendix D-3). The shortest DL was observed at site 9 and is probably responsible for the lower mean TMX.

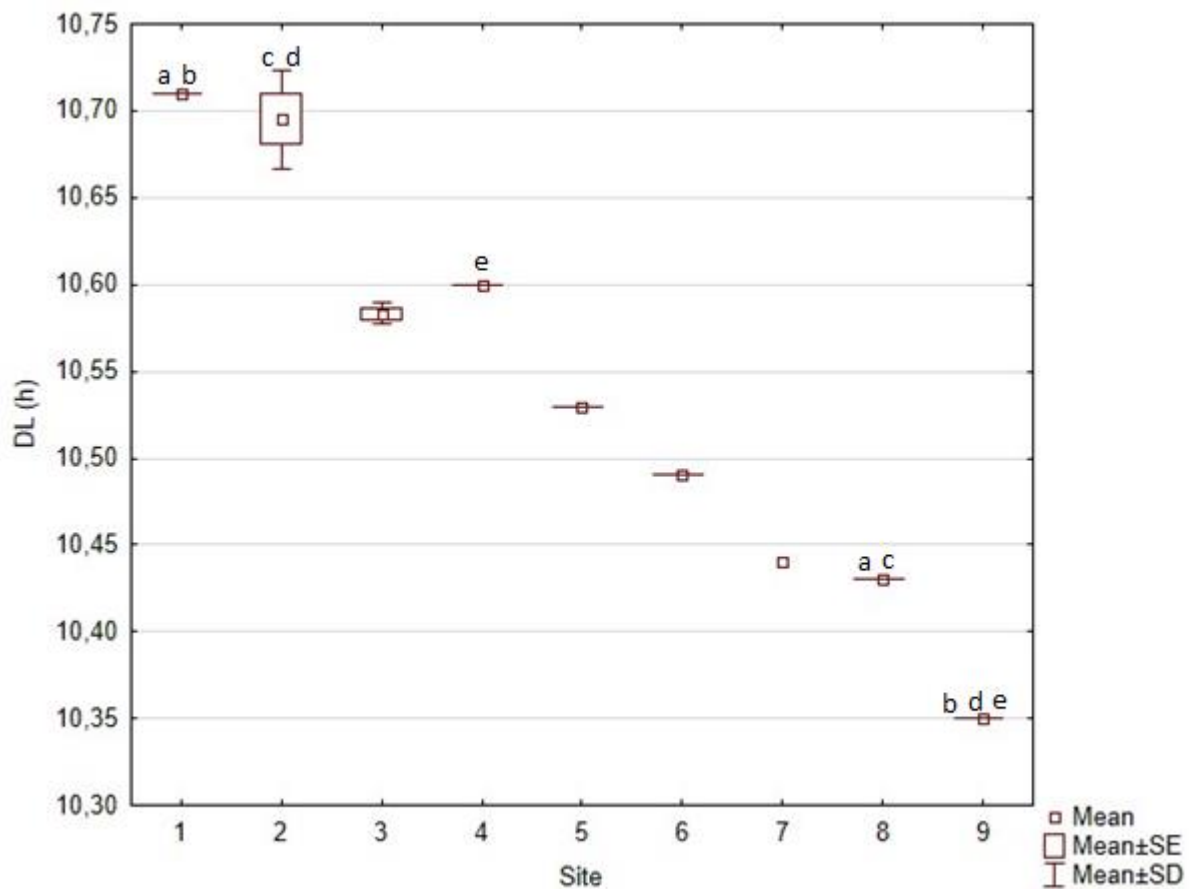


Figure 5.10: Box and whiskers plot showing the mean day length (DL) across the study sites, $\pm$ standard error (SE) and standard deviation (SD) $n=43$. A $p < 0.05$ indicated that there was a significant difference in day lengths between the sites highlighted in Appendix D-3. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

5.3.4.4 Relative humidity

Site 1 and 2 showed $\pm 2\%$ deviation from mean relative humidity as shown in Figure 5.11. The relative humidity measured from sites 3-9 showed slight variation from the mean. Relative humidity ranged from 62-77%, with sites 2, 4-6, 8 and 9 having the optimum humidity above 67%. Sites 1 and 3 had a significantly lower relative humidity than sites 5 and 6 (Appendix D-4).

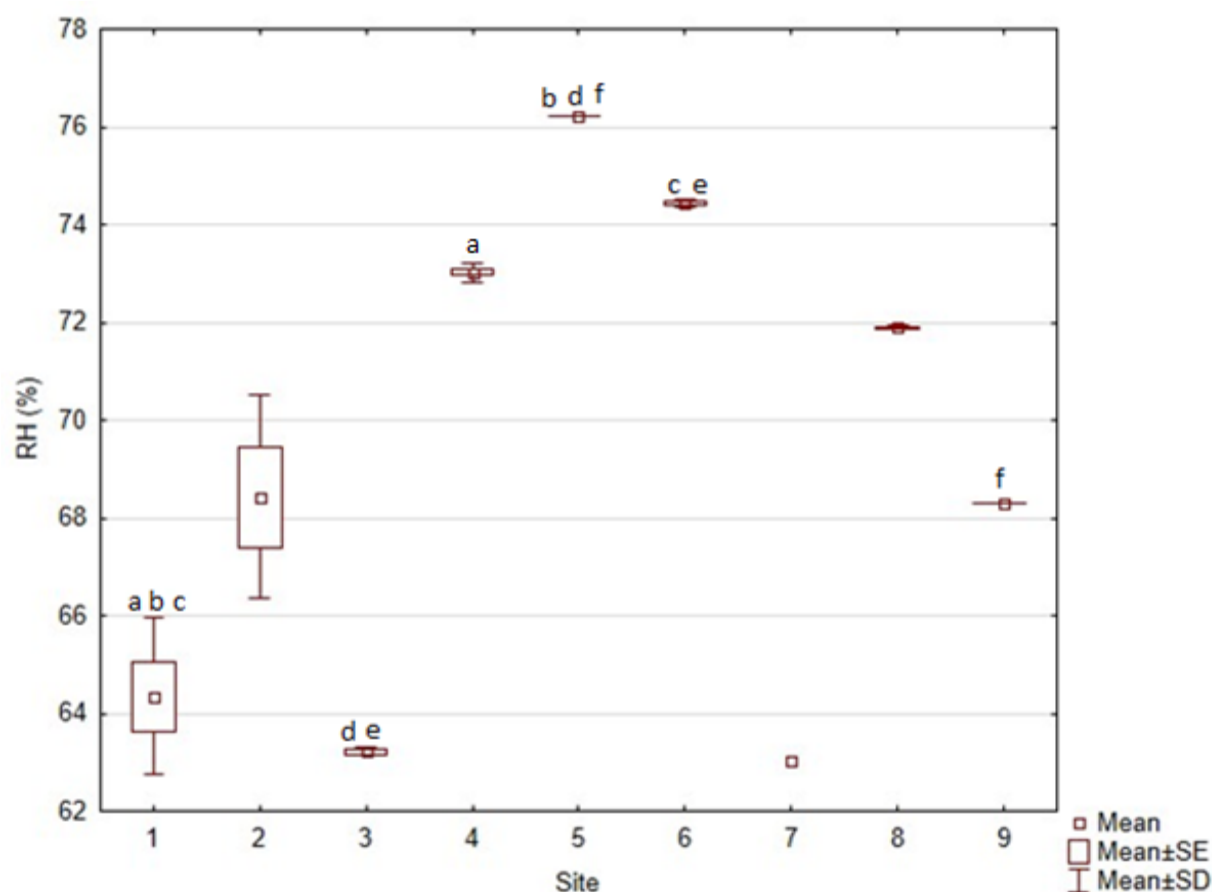


Figure 5.11: Box and whiskers plot showing the mean relative humidity (RH) across the sites, $\pm$ standard error (SE) and standard deviation (SD) $n=43$. A $p<0.05$ indicated that there was a significant difference in relative humidity between the sites highlighted in Appendix D-4. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

5.3.4.5 Soil water content at 100mm

Sites 1-5 fall under the irrigated sugarcane crop regions of the study area, while the other sites are rain fed. The mean daily SWC was found to be less than 15mm for site 1-4, while site 5-9 had greater than 30mm (Figure 5.12). Site 8 had the highest soil moisture content ($>60\text{mm}$) whilst site 1 had the lowest SWC ($<10\text{mm}$). The SWC at site 1 was significantly lower than that of sites 5, 8 and 9 (Appendix D-5). Generally, SWC increased from northern inland regions (Site 1) to the southern coastal region (site 9). Site 5 in contrast was the only irrigated site to have SWC above 20mm. This was attributed to unlimited water quota at this site, compared to the other irrigated sites limited to 995-1300 mm irrigation water per annum (Personal communication- Vusi Mtungwa- Biosecurity Technician (SASRI)).

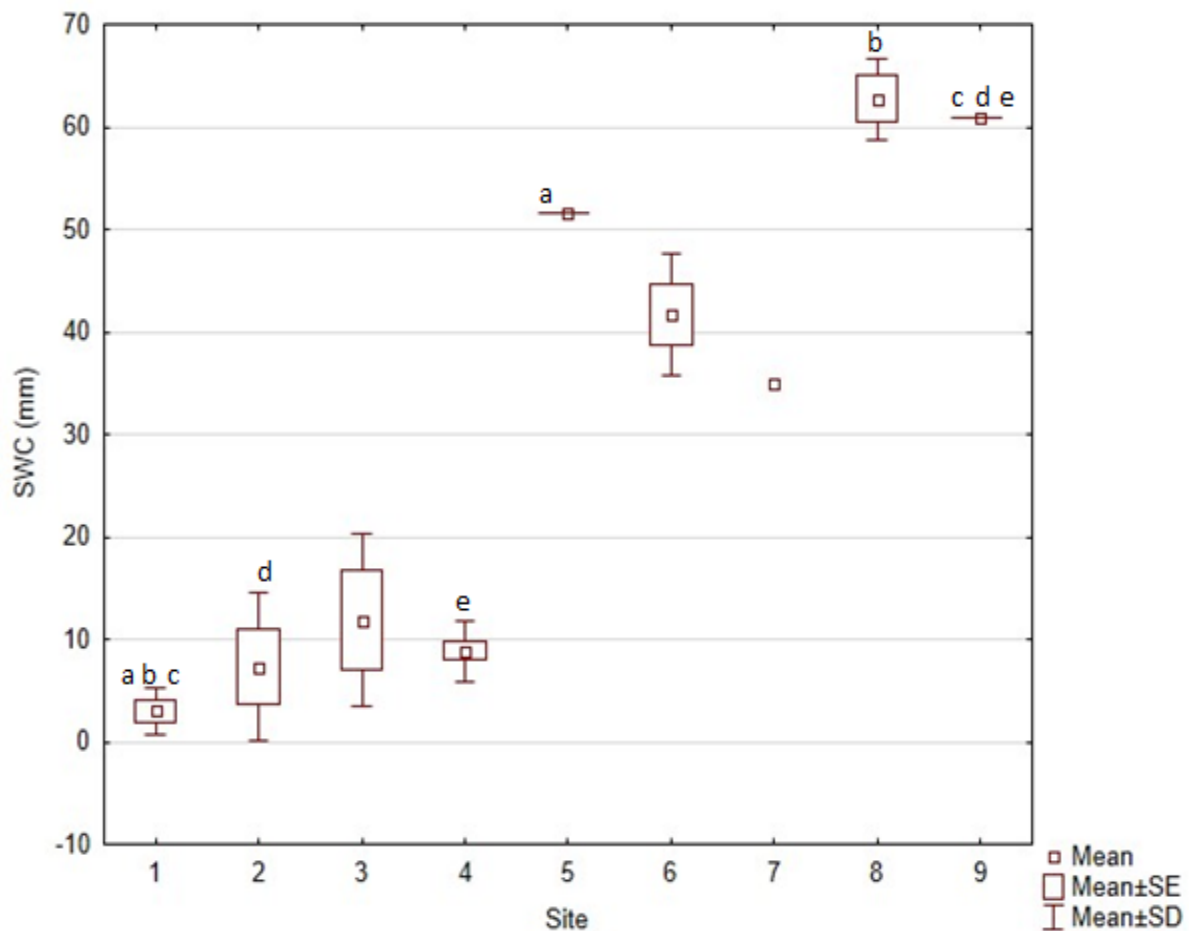


Figure 5.12: Box and whiskers plot showing the mean soil water content (SWC) at 100mm across the study sites, $\pm$ standard error (SE) and standard deviation (SD) $n=43$. A $p<0.05$ indicated a significant statistical difference between the observed SWC at 100mm between the sites highlighted in Appendix D-5. Box plots with different letters indicate significant differences. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site 5: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

While the source of crop water (irrigated against rain fed) affects the SWC, a Kruskal–Wallis ANOVA test for statistical difference revealed that there was a significant difference between pollen viability of irrigated regions to that of rain fed regions (Figure 5.13). Pollen viability was higher at irrigated regions, having a mean value of 30% while at rain fed regions the mean pollen viability was approximately 5%. Furthermore, the maximum pollen viability at irrigated regions was greater than 60% while it was lower than 10% at rain fed regions.

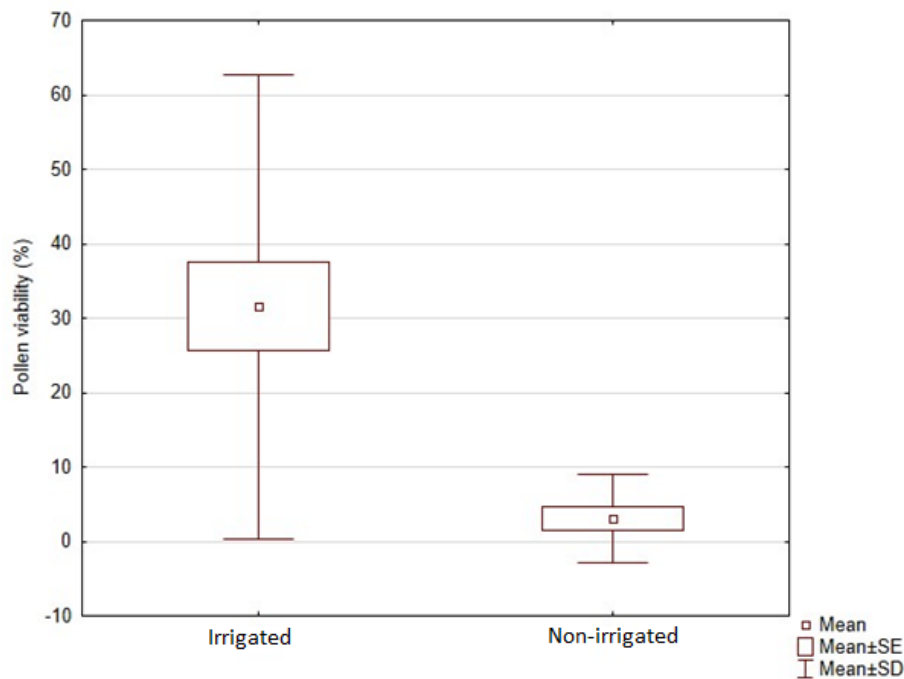


Figure 5.13: Box and whiskers plot showing the difference in pollen viability for all varieties between irrigated and non-irrigated regions, $\pm$ standard error (SE) and standard deviation (SD) $n=43$.

5.3.5 Optimal environmental variables for pollen viability

Since the desired range of mean minimum and maximum temperature for pollen viability is between 18 and 32°C, none of the sites met this requirement (Table 5.2). All sites had mean maximum temperatures not exceeding the threshold for highest temperature, which is favourable to pollen viability. Furthermore, all sites had day lengths shorter than the minimum threshold of 12.5 hours. Sites 2, 4-6, 8 and 9 were the only sites with the optimal mean relative humidity greater than 67%.

Site 2, with the highest mean pollen viability of 60%, had the same set of optimum environmental conditions (maximum temperature <32°C and relative humidity $\geq$ 67%) as sites 6 and 8, where the mean pollen viability was the lowest (0%). Similarly, site 1 with the second highest mean pollen viability (48%) had the same set of optimal environmental conditions (maximum temperature <32°C) as site 7, with the second lowest pollen viability (2%). The difference in pollen viability observed between two sites with more or less the same environmental parameters indicates that there is no correlation between pollen viability and the environmental gradients.

Table 5.2: The optimal range of environmental conditions required for pollen viability was compared to the actual mean pollen viability observed in each site in order to assess their relationship. Rows highlighted with the same colour indicate that the sites had the same sets of optimal environmental parameters. Site 1: Malelane and site 2: Komatipoort (Mpumalanga), site 3: Pongola, site 4: Jozini, site: Mtubatuba, site 6: Empangeni, site 7: Umhlali, site 8: Mount Edgecombe and site 9: Port Shepstone (KwaZulu-Natal).

Site	Mean % pollen viability	Tmin >18°C	Tmax <32°C	Daylength 12h30min-12h55min	Humidity ≥67%
1	48	X	✓	X	X
2	60	X	✓	X	✓
3	5	X	✓	X	X
4	23	X	✓	X	✓
5	28	X	✓	X	✓
6	0	X	✓	X	✓
7	2	X	✓	X	X
8	0	X	✓	X	✓
9	7	X	✓	X	✓

5.3.6 Correlation between pollen viability and environmental variables

Day length, SWC and TMX all with $p < 0.05$ showed a significant correlation with mean pollen viability (Figures 5.14-5.16 & Appendix E). Day length and TMX showed a positive correlation, thus increasing pollen viability as each of these variables increased. SWC in contrast showed a negative correlation with pollen viability. As opposed to a perfect correlation that would be indicated by a value of +1 or -1, day length and TMX had the highest correlation coefficient of 0.5 and 0.6 respectively. SWC showed a correlation coefficient of -0.4. The insignificant correlation ($p > 0.05$) between pollen viability and relative humidity and TMN (Figure 5.17-5.18) was shown by correlation coefficients close to zero (0.09 for both RH and TMN), indicating that there was no association between each of these two variables and pollen viability.

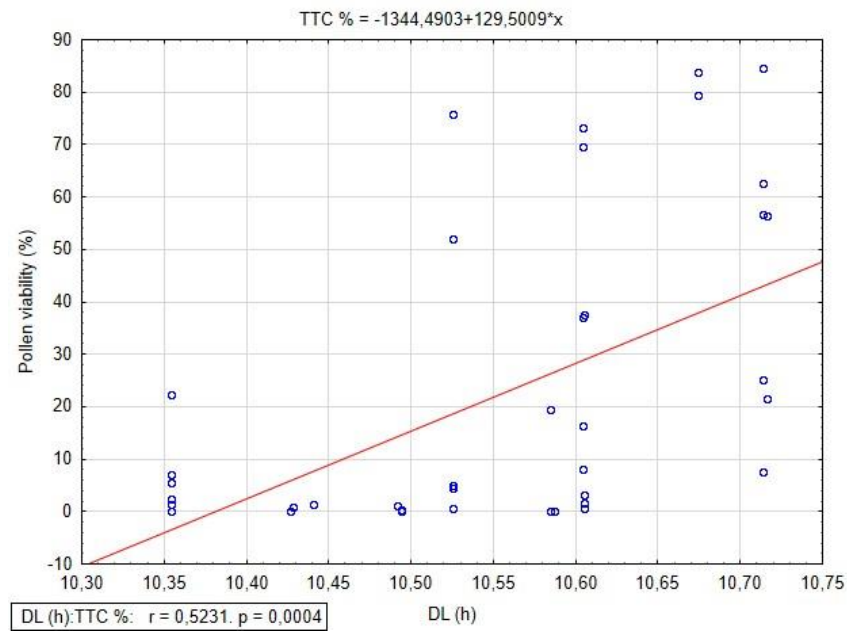


Figure 5.14: Correlation between mean pollen viability and day length for viability assessed using TTC stain ($p=0.0004$)

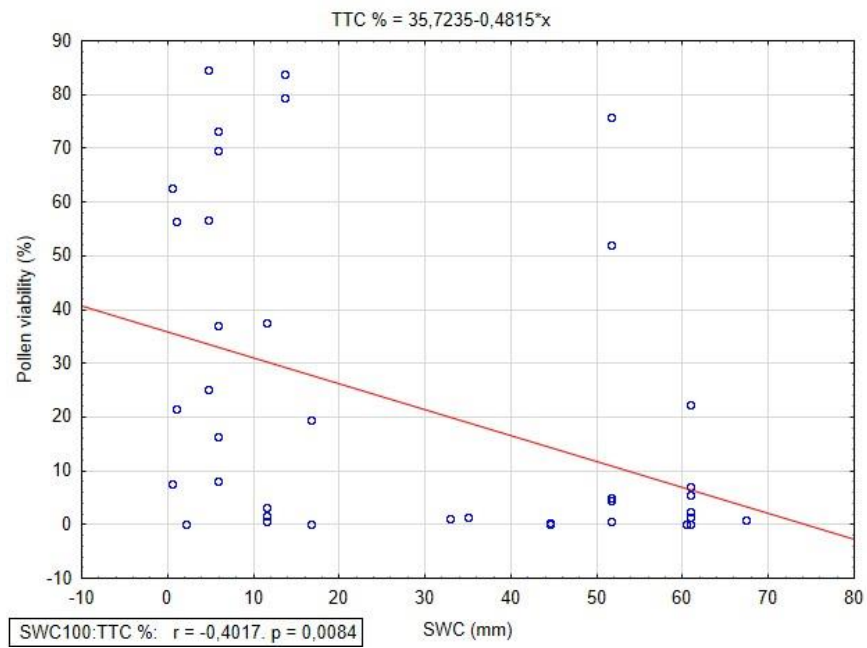
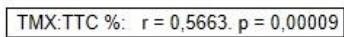


Figure 5.15: Correlation between mean pollen viability and soil water content for viability assessed using TTC stain ($p=0.0084$)



TTC % = 48,4636-2,0751*x

Pollen viability (%)

TMN (°C)

TMN:TTC %: $r = -0,0917$. $p = 0,5636$

93

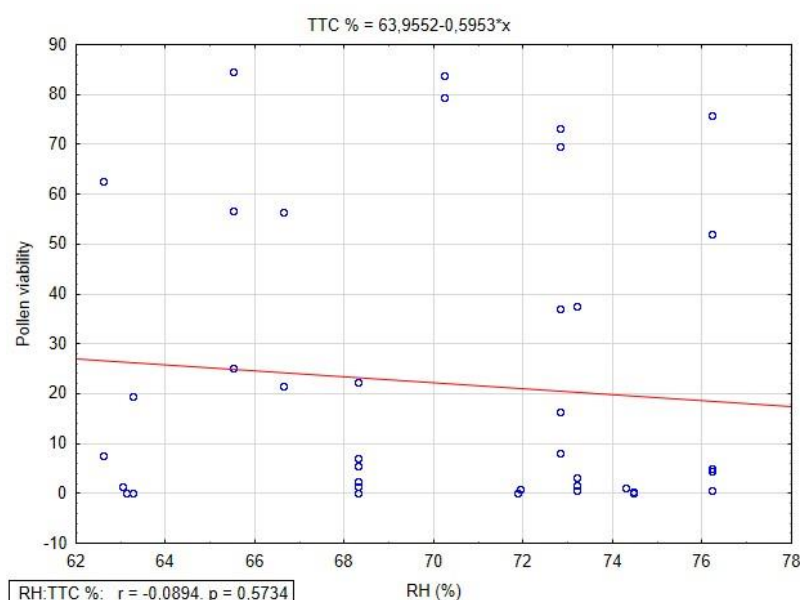


Figure 5.18: Correlation between mean pollen viability and relative humidity for viability assessed using TTC stain ($p=0.5734$)

5.3.7 Multivariate analysis of principal components

The first axis (PC 1) of the PCA explained 71.7% of the variance in the data (Table 5.3 and Figure 5.19). SWC and pollen viability of sugarcane varieties at sites 5-7 and 9 associated closely with the first axis. Their vectors however indicated a strong, negative correlation. This was confirmed by a statistically negative Spearman rank correlation (Figure 5.15 & Appendix E). The second axis (PC 2) explained 21.5% of the variance in the data and was closely associated with temperature, day length and relative humidity. Although there was a negative correlation between pollen viability and soil moisture content there was a significant positive correlation between pollen viability and day length as well as mean maximum temperature. Samples collected from sites 1-4 also strongly associate with these variables. Samples collected from sites 5, 6, 7 and 9 strongly associated with lower soil moisture content.

Table 5.3: Eigenvalues and cumulative percentage variance contributed by the four axes on the PCA ordination.

Axes	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative % variance of species data
1	0.717	71.7	0.717	71.7
2	0.215	21.5	0.932	93.2
3	0.056	05.6	0.988	98.8
4	0.011	01.1	0.999	99.9

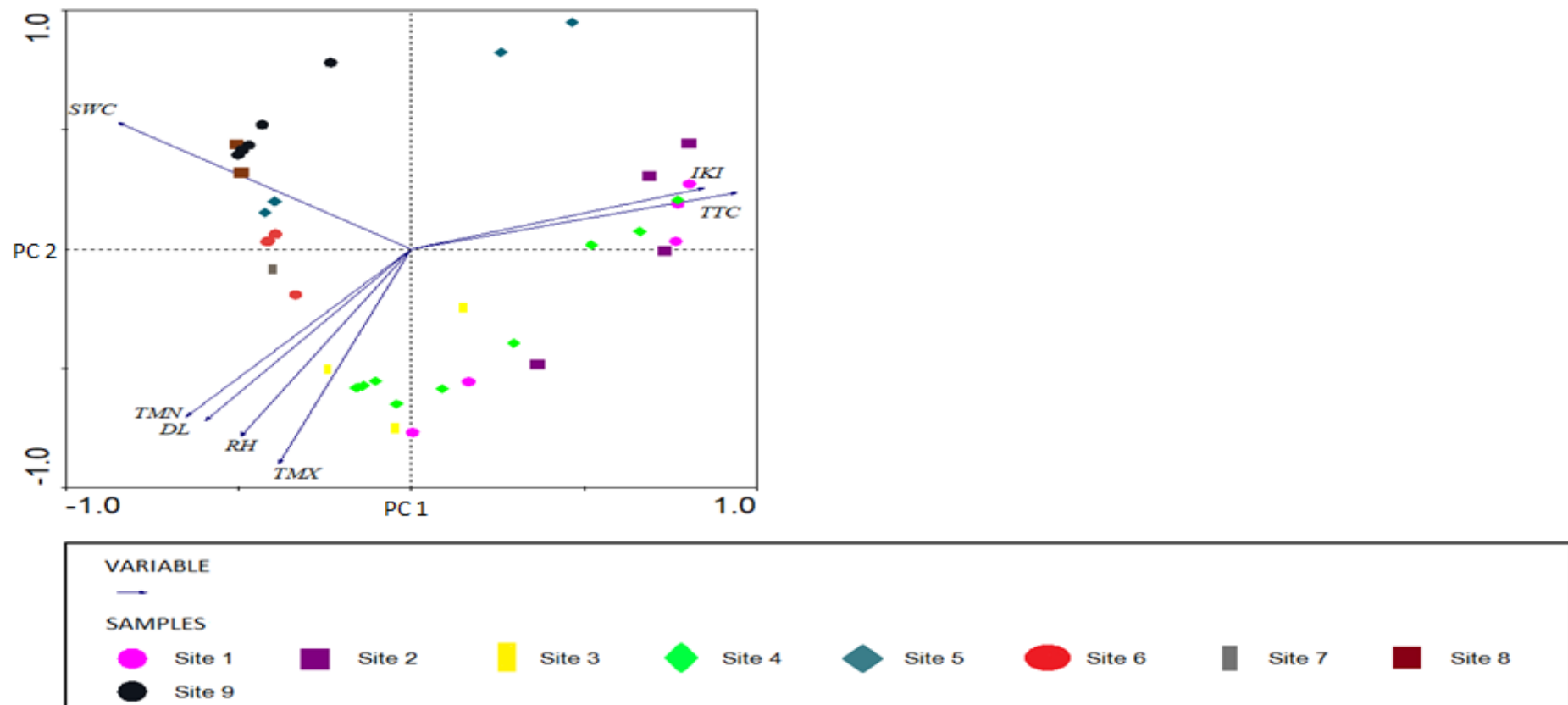


Figure 5.19: PCA of variables contributing to the variance within the pollen viability data, their relation to the first and second axes, as well as the individual sugarcane varieties at different sites. Variables represent soil water content (SWC), minimum temperature (TMN), maximum temperature (TMX), relative humidity (RH) and day length (DL) as environmental variables and 2,3,5 Triphenyl tetrazolium chloride (TTC) and Iodine potassium iodide (IKI) as staining methods

5.4 Discussion

5.4.1 Correlation of staining methods

While the TTC staining method was found to have recorded higher pollen viability than the IKI stain, the difference was found to be statistically insignificant. Several studies found that IKI and TTC were good estimates of pollen viability (Beck-Pay, 2012; Gaaliche *et al.*, 2013; Huang *et al.*, 2004) while others contradicted this (Soares *et al.*, 2013a; Soares *et al.*, 2013b; Sulusoglu & Cavusoglu, 2014). The reliability of staining methods may depend on the type of species investigated, as these previous studies worked on different species. The use of vital stains however provides a rough estimate of pollen viability; germination tests or possibly seed setting should be carried out for more accurate estimates. The presence of active enzymes may not necessarily imply that pollen would fertilise the ovule upon reaching the stigma (Beck-Pay, 2012). Vital stains may indicate the ability of a pollen grain to produce a germination tube, however because of the pollen tube length that may be less than the grain diameter, it would not be considered viable in germination tests. Furthermore, Beck-Pay (2012) states that low seed set is usually associated with incompatibility within the style or ovary, which is not related to the inability to produce a germination tube or pollen viability.

5.4.2 Pollen viability and the real determinant factors

“Sugarcane does not produce viable pollen in South Africa because flowering occurs in winter when temperatures fluctuate below 20°C, the minimum threshold for pollen survival” (Horsley & Zhou, 2013). With sugarcane commercial varieties N14 and N36 showing pollen viability greater than 80% in some regions of South Africa, the findings of the current study are in contrast with this. Pollen viability, however, showed a gradual decrease from the northern inland regions to the south coastal regions of the study. Horsley & Zhou (2013) however confined their study to Mount Edgecombe (site 8 of this study). Even though they did not identify the varieties they worked on, this study confirmed their results and reported pollen viability less than 1% at this particular site for varieties N42 and NCo376.

Within certain study sites (e.g. site 5), some varieties showed pollen viability of 70%, while others had less than 10% of viable pollen. A similar study in Brazil, by Melloni *et al.* (2015), also discovered that some varieties were 100% viable while others showed pollen viability less than 9%, regardless of the similar levels of optimal humidity, temperature and photoperiod. Likewise, Ethirajan (1987 cited by Horsley & Zhou, 2013) stated that even under ideal conditions, some sugarcane varieties would still show poor pollen viability. So, while environmental conditions may have been acknowledged to affect pollen viability, this indicates that it might be worthwhile to investigate the genotype as having an influence on pollen viability. In agreement with this, Melloni

et al. (2015) also associated the constant pollen viability obtained across different treatments to an individual's genotype.

Spearman rank correlation showed that the variation in pollen viability across the study sites was significantly associated with day length (Figure 5.14). However, all the sites had day lengths less than the optimum photoperiod of 12h30min, and regardless of that, some varieties still produced viable pollen while others did not. Therefore, this correlation between day length and pollen viability should no longer be considered valid. Srivastava *et al.* (2006) showed that day length did not have an influence on pollen viability and floral organs, as they recorded a non-significant variation in amount of viable pollen between sugarcane clones grown under natural sunlight and dark treatment.

Spearman rank correlation also showed a significant negative association between soil water content and pollen viability. This is in contrast to what was documented in the literature. For instance, adequate soil moisture has been shown to aid panicle development (Clements, 1975; Hale *et al.*, 2017; Pereira *et al.*, 1983; Shanmugavadivu & Rao, 2009). In addition, the SASRI weather web measures the extent of flowering based on soil water content (and temperature), in such that varieties cultivated in soil with a water content above 50% of capacity of total available moisture are scored high for the propensity to flower. Accordingly, in an event of correlation of pollen viability with SWC, a positive rather than a negative correlation could at least be accepted to be meaningful. Furthermore, if this correlation was to be considered true, then variety N14 would not be equally viable (>70%) between site 1 with less than 10mm SWC and site 5 with approximately 50mm SWC. Therefore, the significant difference (p -value>0.05) in pollen viability obtained between irrigated (high) and rain-fed regions of this study could rather be attributed to the different types of sugarcane varieties cultivated in these regions rather than the varying SWC.

The significant correlation between pollen viability and temperature should at least be accepted to affect pollen viability to some extent. With a general decrease in pollen viability from the northern towards the southern regions, it was also evident that daily maximum temperatures decreased towards the south coastal regions of the study. Hence, if the significant correlation between pollen viability with both day length and soil water content is ruled out, and temperature is only acceptable to some extent, then the reason for varying pollen viability of sugarcane varieties remains unresolved. This study proposes that mean daily temperature together with genotype plays a vital role in pollen viability.

In agreement with the influence of temperature, Brett (1950) concluded that temperature was the only factor that limited the development of viable pollen of sugarcane at SASRI, KwaZulu-Natal. From these experiments, he ruled that increased day length and humidity was unnecessary for

the development of viable pollen. Brett (1950) further proposed there was a possibility that temperature is responsible for the limited amount of sugarcane flowering in this country.

According to Ethirajan (1987 cited by Cheavegatti-Gianotto *et al.*, 2011; Horsley & Zhou, 2013), the poor pollen viability observed from sugarcane varieties even when under ideal conditions is due to the cytogenetic abnormalities that occur during meiosis that are associated with sugarcane's high polyploid multispecies genetic complex. Normal meiotic division ensures gamete viability as gametogenesis is controlled by a great number of genes that act from premeiotic to postmeiotic mitosis (Pagliarini, 2000). This was confirmed by Nair (1975), where crosses between *Sa. officinarum* and *Sa. spontaneum* showed that 80% of hybrids had a $2n+n$ constitution of chromosomes, with *Sa. officinarum* as a female parent having transmitted its somatic chromosomes. As the number of cells with bridges, laggards and fragments increased in an individual hybrid, a decrease in pollen viability was also observed. For instance, an individual with 100% cells containing laggards had 0% pollen viability (Nair, 1975). Pagliarini (2000) also stated that abnormality in meiotic division was a dominant factor reducing pollen viability in sugarcane. Likewise, this was found to be the cause of poor fertility in hybrids of other crops like sunflower (Quillet *et al.*, 1995), wheat (Rezaei *et al.*, 2010), sage (Haque & Ghoshal, 1980), and many other crops (Bajpai & Signh., 2006; Cavalcantea *et al.*, 2000; Pagliarini, 2000; Palma-Silva *et al.*, 2004; Silva *et al.*, 2011).

According to Price (1959), adventitious embryos resulting from a series of somatic cell divisions would exactly re-produce the maternal genotype, in such that sugarcane seedlings derived from this process are known to resemble their maternal parents as one stool of a vegetatively cultivated cane variety would resemble another. Since the literature (Amaral *et al.*, 2013; Melloni *et al.*, 2015; Melloni *et al.*, 2012) documented that female parents in crosses are chosen according to the amount of viable pollen produced (<30%), this could be the reason some varieties do not produce viable pollen regardless of the environment, as was seen in varieties N19, N23, N25, N42, N43 and NCo376. Varieties N14 and N36, in contrast, maintained their high pollen viability throughout the sites. However, N36 was found to have only 10% viable pollen at site 9. Thus, since this site was recorded with significantly lower minimum and maximum temperature than the other sites, it would be expected that temperature played a role in suppressing pollen viability in this case.

Brett (1950) in his observations on the flowering patterns of sugarcane varieties noted that flowering was rather mainly variety based, as some varieties would always flower while others did not. This study also suggests that just like flowering, pollen viability could be mainly variety dependant. Possibly supporting this, Melloni *et al.* (2015) observed that once an inflorescence of a certain variety has been classified as either pollen donor (male) or receptor (female), this classification would rarely change throughout the different breeding programmes.

5.4.3 Current sugarcane varieties in relation to the unreleased GM varieties

Variety N14 from female and male parents (N7 x unknown parent) is designated for irrigated regions in South Africa, and is also one (of two, the other being NCo376) of the widely cultivated varieties in Zimbabwe (Zhou, 2004). N36 on the other hand (82F1225 x 78Z1635) is widely adapted and has dual purpose (irrigated and rain-fed), and is one of the best performing varieties (Personal communication: Sifiso Thwala Biosecurity Technician (SASRI)). This variety was however recorded only once in the rain fed regions where it exhibited low pollen viability. Current development on GM sugarcane has not been recorded to have utilized these varieties or their direct parental lines.

For instance, completed breeding projects utilised commercial varieties N12 (NCo376 x Co331), N19 (NCo376 x CB40/35) and NCo310 (Co421 x Co312) for herbicide resistance; N27 (NiN2 x N52/219) and unreleased genotypes 88H0019 and 93F0234 for insect resistance; NCo310 and unreleased genotype 75E0247 for virus resistance (Meyer & Snyman, 2013; Snyman & Meyer, 2012; Sooknandan *et al.*, 2003; Watt *et al.*, 2010). As of 2012, GM projects under development were for sucrose metabolism on NCo310, drought tolerance on N12 and NCo376 (Co421 x Co312) and nitrogen use efficiency on NCo376 and N41 (77F790 x 82W1542) (Snyman & Meyer, 2012).

Of these varieties, only N12 and NCo310 were not tested for pollen viability in this study. However, NCo376 and NCo310 share exactly the same sets of parents, the only difference being that they were released 10 years apart (Zhou, 2013), thus it could be assumed that their pollen viability would be the same. All these varieties (NCo376, N19, N27, and N41) showed low or no pollen viability in this study. It would also be likely that N12 might show low pollen viability as the maternal line was NCo376. Noted from these varieties was that N12, N27 and NCo376 are designated for irrigated regions while N41 and N19 are for both irrigated and rain fed regions (Redshaw *et al.*, 1999; Redshaw & Nuss, 2001). The parental lines and pollen viability of the unreleased genotypes that were also included for GM modification developments however remain unknown and will need to be investigated.

5.5 Conclusion

Varieties N14 and N36 were found to have the highest percentage of viable pollen, showing a possibility of acting as pollen donors. Pollen viability decreased from the northern irrigated regions towards the south coastal regions of the study area. This gradient is believed to be largely determined by temperature. The effect of temperature would however need to be investigated further because while site 9 had a significantly lower temperature, all the sites generally had temperatures below the optimum for flowering recommended by literature. While the study

provided an estimate of pollen viability in field conditions, more accurate tests will need to be performed to assess the probability of sugarcane pollen fertilising another species. It can therefore be suggested that pollen viability is most likely dependent on genotype, which in turn is influenced by temperature. If the GM variety shares parental lines with N14 and N36 it should only be cultivated at site 9 and further south to decrease the chances of gene flow.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

A total of 36 incidents of hybridization between sugarcane cultivars and other grass species were reported in the literature, of which none were spontaneous. These crosses were mainly conducted with *Saccharum* (Syn. *Erianthus*) and *Sorghum*. This confirms that sugarcane can exchange genetic material with members of the Sorghinae and Saccharinae. However, efforts to generate strong and fertile hybrids from crosses between sugarcane and several related species have been largely unsuccessful and indicated that when natural barriers to gene-flow are being crossed, the likelihood of a transgene persisting in a recipient population is extremely low. Crosses between *Saccharum* species have displayed that no true hybrids could be obtained with the use of *Saccharum* sp. hybrid (sugarcane), as opposed to *Sa. officinarum* as a parental line. This was attributed to the high meiotic balance of sugarcane. In addition, whether sugarcane was used as a female or male parent, attempts to generate true hybrids were still unsuccessful. This shows that sugarcane could rarely accept genes from related species, thus also lowering the likelihood of seed establishment from GM sugarcane fertilised by a related species (first route of transgene escape).

As hybridisation remains a possibility, and occurs most successfully with members of the Saccharinae and Sorghinae genera, it was necessary to determine the relatedness of wild relatives of *Saccharum* in the study area. Phylogenetic studies based on DNA barcodes identified *Sorghum arundinaceum* (Sorghinae), *Miscanthus ecklonii* and *Imperata cylindrica* (Saccharinae) as the closest relatives to sugarcane. These species occur frequently in the sugarcane production regions and flower from May to July in synchrony with sugarcane. When *I. cylindrica* was artificially crossed with sugarcane, efforts did not produce any true hybrids. Even though there were no direct artificial crosses between sugarcane and *M. ecklonii* or *So. arundinaceum*, 3% (highest reported true hybrid percentage from artificial hybridization with sugarcane) and 0.16% of crosses between these genera produced genuine hybrids from thousands of seedlings respectively.

Even though sugarcane has the potential to hybridize with the closely related members of the Sorghinae and Saccharinae which are also present in the sugar production regions, with synchronous flowering, gene flow would not occur without the production of viable pollen. Pollen viability studies revealed that sugarcane varieties N36 and N14 could potentially act as pollen donors in the northern irrigated, sugar production regions of South Africa (Malelane (site 1) to Mtubatuba (site 5)). Out of 13 sugarcane varieties tested for pollen viability in 2016 and 2017, these were the only two varieties to exhibit up to 85% viable pollen, while other varieties were potential pollen receptors. The factors most likely to be responsible for pollen viability according to the results presented in this study were suspected to be linked to the genetic make-up of a

variety and daily mean maximum temperature. Tests of pollen viability of these commercial varieties have provided insight into potential gene flow of the GM lines of sugarcane under development as these lines have not been developed from N14 and N36 and neither share direct parental lines with these two varieties. More importantly, the commercial varieties (NCo376, N19, N27, and N41) for which transgenes have been inserted were found to exhibit low or no pollen viability. Because of this, pollen of GM lines is not expected to be viable. Therefore, this lowers the probability that pollen from GM sugarcane may cross-pollinate a related species outside of cultivation, which is the most common route for transgene dispersion (second route).

Since the unreleased varieties for which GM lines developed have not been evaluated for pollen viability, pollen movement needs to be taken into consideration. While the production of viable pollen and presence of closely related species within the cultivation regions with synchronous flowering may present a probability of pollen-mediated gene flow, the dispersal distance of pollen remains unknown. Due to its high sensitivity to desiccation and a half-life of 12 minutes, it is suspected that pollen would not disperse very far. True hybrids with sugarcane and wild relatives have rarely been bred (after crossing all gene-flow barriers and assisting the survival of seedlings), it would therefore remain highly unlikely for a chance event like wind pollination to produce hybrids from GM sugarcane.

As a precautionary measure, further research is recommended:

- Pollen viability of the unreleased varieties 88H0019, 93F0234 and 75E0247 from which GM cane was developed remains unknown, thus future studies may need to investigate this especially when cultivated under irrigation.
- The true lower threshold of maximum temperature which is detrimental to pollen viability was not identified and needs to be determined before boundaries can be drawn for high gene flow potential areas.
- The dispersal distance of pollen from sugarcane remains unknown. Information on this will determine the required isolation distance of sugarcane (either by mowing and herbicide control) from the closely related species within and around the cultivation areas.
- The ability of closely related species to donate pollen could lead to the formation of field hybrids that can cross back with wild relatives and create super-weeds. Thus, pollen viability of wild species may need to be assessed.

REFERENCE LIST

- Aitken, K., Li, J., Wang, L., Qing, C., Fan, Y.H. & Jackson, P. 2006. Characterization of intergeneric hybrids of *Erianthus rockii* and *Saccharum* using molecular markers. *Genetic Resources and Crop Evolution*, 54(7):1395-1405.
- Amaral, A.L., Santos, J.M., Câmara, T.M.M., Filho, L.S.C. & Barbosa, G.V.S. 2013. Pollen Storage Methodology in Sugarcane. *Comunicado Técnico*, 129:1-11.
- Anderson, E. 1949. Introgressive hybridization. *Wiley, New York*. DOI:10.1111/j.1469-185X.1953.tb01379.x
- Andow, D.A. & Zwahlen, C. 2006. Assessing environmental risks of transgenic plants. *Ecology Letters*, 9(2):196-214.
- Arencibia, A., Vázquez, R.I., Prieto, D., Tèllez, P., Carmona, E.R., Coego, A., Hernández, De la Riva G. A & Selman-Housein, G. 1997. Transgenic sugarcane plants resistant to stem borer attack. *Molecular Breeding*, 3:247–255.
- Arro, J.A., Veremis, J.C., Kimbeng, C.A. & Botanga, C. 2006. Genetic diversity and relationships revealed by AFLP markers among *Saccharum spontaneum* and related species and genera. *Journal American Society Sugar Cane Technologists*, 26:101-115.
- Asante, D.K.A. 2008. Genetically modified food -The dilemma of Africa. *African Journal of Biotechnology*, 7(9):1204-1211.
- Australian Government. 2004. The biology and ecology of sugarcane (*Saccharum* spp. hybrids) in Australia.
- Babu, C., Koodalingam, K., Natarajan, U.S., Shanthi, R.M. & Govindaraj, P. 2009. Genetic enhancement of sugarcane (*Saccharum* sp. Hybrids) for resistance to red rot disease and economic traits. *The Journal of Agricultural Sciences*, 4(3):97-107.
- Bacci Jr., M., Miranda, V.F.O., Martins, M.S., Figueira, A.V.O., Lemos, M.V., Pereira, J.O. & Marino, C.L. 2001. A search for markers of sugarcane evolution. *Genetics and Molecular Biology*, 24(1):169-174.
- Bajpai, A. & Singh, A. K. 2006. Meiotic behavior of *Carica papaya* L.: Spontaneous chromosome instability and elimination in important cvs. in North Indian conditions. *Cytologia*, 71(2): 131–136.
- Barnard, S., Diedericks, V. & Conradie, K.R. 2013. Genetic diversity of vetiver isolates (*Chrysopogon zizanioides/nigritanus*) available in South Africa based on *ITS*, *ndhF* and *rbcL* sequencing analyses. *South African Journal of Botany*, 86:63-67.
- Beck-Pay, S.L. 2012. Optimisation of pollen viability tests for *Acacia podalyriifolia* and two ploidys of *Acacia mearnsii*. *South African Journal of Botany*, 78:285-289.
- Bischoff, K.P. & Gravois, K.A. 2004. The development of new sugarcane varieties at the *Isu* agcenter. *Journal American Society Sugar Cane Technologists*, 24:142-164.
- Bonnett, G.D., Berding, N., Morgan, T. & Fitzgerald, P. 2007. Implementation of genetically modified sugarcane—The need for a better understanding of sexual reproduction. *Proceedings of the Australian Society of Sugar Cane Technologists*, 29:1-9.

- Bonnett, G.D., Nowak, E., Olivares-Villegas, J.J., Berding, N., Morgan, T. & Aitken, K.S. 2008. Identifying the risks of transgene escape from sugarcane crops to related species, with particular reference to *Saccharum spontaneum* in Australia. *Tropical Plant Biology*, 1(1):58-71.
- Bonnett, G.D., Olivares-Villegas, J.J., Berding, N. & Morgan, T. 2010. Sugarcane sexual reproduction in a commercial environment: Research to underpin regulatory decisions for genetically modified sugarcane. *Proceedings of the Australian Society of Sugar Cane Technology*, 32:1-9.
- Bothma, G., Mashaba, C., Mkhonza, N., Chakauya, E. & Chikwamba, R. 2010. GMOs in Africa: opportunities and challenges in South Africa. *GM Crops*, 1(4):175-180.
- Bourne, B.A. 1935. A comparative study of certain morphological characters of sugarcane x sorgho hybrids. *Journal of Agricultural research*, 5(6):539-552.
- Brett, P.G.C. 1950. Investigations on sugarcane breeding in Natal during 1949. *Proceedings of the South African Sugar Technologists Association*, 24:99-103.
- Brett, P.G.C. 1954. *Saccharum-Miscanthidium* hybrids. *Journal of Genetics*, 52:542-546.
- Burke, S.V., Wysocki, W.P., Zuloaga, F.O., Craine, J.M., Pires, J.C., Edger, P.P., Mayfield-Jones, D., Clark, L.G., Kelchner, S.A. & Duvall, M.R. 2016. Evolutionary relationships in Panicoid grasses based on plastome phylogenomics (Panicoideae; Poaceae). *BioMed Central Plant Biology*, 16(1):140-150
- Burner, D.M. & Webster, R.D. 1994. Cytological studies on North American species of *Saccharum* (Poaceae: Andropogoneae). *Sida, Contributions to Botany*, 16:233-244.
- Cai, Q., Aitken, K., Deng, H.H., Chen, X.W., Fu, C., Jackson, P.A. & McIntyre, C.L. 2005a. Verification of the introgression of *Erianthus arundinaceus* germplasm into sugarcane using molecular markers. *Plant Breeding*, 124:322—328.
- Cai, Q., Aitken, K.S., Fan, Y.H., Piperidis, G., Jackson, P. & McIntyre, C.L. 2005b. A preliminary assessment of the genetic relationship between *Erianthus rockii* and the “*Saccharum* complex” using microsatellite (SSR) and AFLP markers. *Plant Science*, 169(5):976-984.
- Campbell, L.G., Snow, A.A., Sweeney, P.M. & Ketner, J.M. 2009. Rapid evolution in crop-weed hybrids under artificial selection for divergent life histories. *Evolutionary Applications*, 2(2):172-186.
- Cavalcante, H.C., Schifino-Wittmann, M.T. & Dornelles, A.L.C. 2000. Meiotic behaviour and pollen fertility in an open-pollinated population of ‘Lee’ mandarin [*Citrus clementina* x (*C. paradise* x *C. tangerina*)]. *Scientia Horticulturae*, 86:103-114.
- CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):12794-12797.
- Chandler, S. & Dunwell, J.M. 2008. Gene flow, risk assessment and the environmental release of transgenic plants. *Critical Reviews in Plant Sciences*, 27(1):25-49.
- Cheavegatti-Gianotto, A., de Abreu, H.M., Arruda, P., Bessalho Filho, J.C., Burnquist, W.L., Creste, S., di Ciero, L., Ferro, J.A., de Oliveira Figueira, A.V., de Sousa Filgueiras, T., Grossi-de-Sa, M.D., Guzzo, E.C., Hoffmann, H.P., de Andrade Landell, M.G., Macedo, N., Matsuoka, S., de Castro Reinach, F., Romano, E., da Silva, W.J., de Castro Silva Filho, M. & Cesar Ulian, E. 2011. Sugarcane (*Saccharum X officinarum*): A reference study for the regulation of genetically modified cultivars in Brazil. *Tropical Plant Biology*, 4(1):62-89.

- Clements, H.F. 1975. Flowering of sugarcane: Mechanics and control. *Bulletin of Hawaii Agricultural Experiment Station*, 92:1-57.
- Cornille, A., Gladioux, P. & Giraud, T. 2013. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications*, 6(5):737-748.
- Dai, Q.Y., Gao, Q., Wu, C.S., Chesters, D., Zhu, C.D. & Zhang, A.B. 2012. Phylogenetic reconstruction and DNA barcoding for closely related pine moth species (*Dendrolimus*) in China with multiple gene markers. *PLoS One*, 7(4):e32544.
- Dale, P.J., Clarke, B. & Fontes, E.M.G. 2002. Potential for the environmental impact of transgenic crops. *Nature Biotechnology*, 20:567-574.
- de Vienne, D.M., Giraud, T. & Martin, O.C. 2007. A congruence index for testing topological similarity between trees. *Bioinformatics*, 23(23):3119-3124.
- Deng, Z., Zhang, M., Lin, W., Cheng, F., Zhang, C., Li, Y., Lai, L., Lin, Y. & Chen, R. 2010. Analysis of disequilibrium hybridization in hybrid and backcross progenies of *Saccharum officinarum* × *Erianthus arundinaceus*. *Agricultural Sciences in China*, 9(9):1271-1277.
- Dhanda, S.S. & Munjal, R. 2006. Inheritance of cellular thermotolerance in bread wheat. *Plant Breeding*, 125:557-564.
- D'Hont, A., Rao, P.S., Feldmann, P., Grivet, L., Islam-Faridi, N., Taylor, P. & Glaszmann, J.C. 1995. Identification and characterisation of sugarcane intergeneric hybrids, *Saccharum officinarum* × *Erianthus arundinaceus*, with molecular markers and DNA *in situ* hybridisation. *Theoretical and Applied Genetics* 91:320-326.
- Dillon, S.L., Lawrence, P.K. & Henry, R.J. 2001. The use of ribosomal ITS to determine phylogenetic relationships within *Sorghum*. *Plant Systems and Evolution*, 230:97-110.
- Dillon, S.L., Shapter, F.M., Henry, R.J., Cordeiro, G., Izquierdo, L. & Lee, L.S. 2007. Domestication to crop improvement: genetic resources for *Sorghum* and *Saccharum* (Andropogoneae). *Annals of Botany*, 100(5):975-989.
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., Cheng, T., Guo, J. & Zhou, S. 2015. *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports*, 5:8348.
- Ellstrand, N.C. 2003. Current knowledge of gene flow in plants: Implications for transgene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1434):1163-1170.
- Ellstrand, N.C., Prentice, H.C. & Hancock, J.F. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics*, 30:539-563.
- Erickson, K. 2010. The Jukes-Cantor Model of Molecular Evolution. *Primus*, 20(5):438-445.
- Evans, D.L. & Joshi, S.V. 2016. Complete chloroplast genomes of *Saccharum spontaneum*, *Saccharum officinarum* and *Miscanthus floridulus* (Panicoideae: Andropogoneae) reveal the plastid view on sugarcane origins. *Systematics and Biodiversity*, 14(6):548-571.
- Fish, L., Mashau, A.C., Moeaha, M.J. & Nembudani, M.T. 2015. Identification guide to southern African grasses. An identification manual keys, descriptions and distributions. *Strelitzia*, 36: South African National Biodiversity Institute, Pretoria. ISBN: 978-928224-00-6.
- Fukuhara, S., Terajima, Y., Irei, S., Sakaigaichi, T., Ujihara, K., Sugimoto, A. & Matsuoka, M. 2012. Identification and characterization of intergeneric hybrid of commercial sugarcane (*Saccharum* spp. hybrid) and *Erianthus arundinaceus* (Retz.) Jeswiet. *Euphytica*, 189(3):321-327.

- Gaaliche, B., Majdoub, A., Trad, M. & Mars, M. 2013. Assessment of Pollen Viability, Germination, and tube growth in eight Tunisian caprifig (*Ficus carica* L.) Cultivars. *ISRN Agronomy*, 2013:1-4.
- Gao, S., Yang, Y., Wang, C., Guo, J., Zhou, D., Wu, Q., Su, Y., Xu, L. & Que, Y. 2016. Transgenic sugarcane with a cry1Ac gene exhibited better phenotypic traits and enhanced resistance against sugarcane borer. *PLoS One*, 11(4):1-16.
- Gao, Y., Liu, X., Zhang, R., Zhou, H., Liao, J., Duan, W. & Zhang, G.-M. 2014. Verification of progeny from crosses between sugarcane (*Saccharum* spp.) and an intergeneric hybrid (*Erianthus arundinaceus* × *Saccharum spontaneum*) with molecular makers. *Sugar Technology*, 17(1):31-35.
- Ge, Y., Fu, C., Bhandari, H., Bouton, J., Brummer, E.C. & Wang, Z.-Y. 2011. Pollen viability and longevity of Switchgrass (L.). *Crop Science*, 51(6):1-8.
- Gepts, P. & Papa, R. 2003. Possible effects of (trans)gene flow from crops on the genetic diversity from landraces and wild relatives. *Environmental Biosafety Research*, 2(2):89-103.
- Goodman, R.M. & Gepts, P. 2004. Gene flow between crops and their wild progenitors. *Encyclopedia of Plant and Crop Science*, 3:488-492.
- Grivet, L., Daniels, C., Glaszmann, J.C. & D'Hont, A. 2004. A review of recent molecular genetics evidence for sugarcane evolution and domestication. *Ethnobotany Research & Applications*, 2: 9-17
- Guimaraes, C.T., Sills, G.R. & Sobral, B.W.S. 1997. Comparative mapping of Andropogoneae: *Saccharum* L. (sugarcane) and its relation to sorghum and maize. *Proceedings of the National Academy of Sciences*, 94:14261–14266.
- Guo, J., Gao, S., Lin, Q., Wang, H., Que, Y. & Xu, L. 2015. Transgenic sugarcane resistant to *Sorghum* mosaic virus based on coat protein gene silencing by RNA interference. *BioMed Research International*, 2015:1-9.
- Gupta, S.C., de Wet, J.M.J. & Harlan, J.R. 1978. Morphology of *Saccharum-Sorghum* hybrid derivatives. *American Journal of Botany*, 65 (9):936-942.
- Hale, A.L., White, P.M., Webber, C.L., & Todd, J.R. 2017. Effect of growing media and fertilization on sugarcane flowering under artificial photoperiod. *PLoS One*, 12(8):e0181639.
- Haque, M.S. & Ghoshal, K.K. 1980. Behaviour of meiotic chromosomes with special reference to their relationship in pollen sterility in *Salvia* L. *Cytologia*, 45:743-752.
- Hauke, J. & Kossowski, T. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2):87-93
- Haygood, R., Ives, A.R. & Andow, D.A. 2003. Consequences of recurrent gene flow from crops to wild relatives. *Proceedings of the Royal Society - Biological Sciences*, 270(1527):1879-1886.
- Hill, R.J. 1983. Johnsongrass, *Sorghum halepense* (L.) Pers. and Shattercane, *Sorghum bicolor* (L.) Moench ssp. *drummondii* (Steud.) de Wet. *Regulatory Horticulture*, 9(1):1-9.
- Hoang, N.V., Furtado, A., Botha, F.C., Simmons, B.A. & Henry, R.J. 2015. Potential for genetic improvement of sugarcane as a source of biomass for biofuels. *Frontiers in Bioengineering and Biotechnology*, 3(182):1-15.
- Hodkinson, T.R., Chase, M.W., Lledo, M.D., Salamin, N. & Renvoize, S.A. 2002. Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based

- on DNA sequences from *ITS* nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergenic spacers. *Journal of Plant Research*, 115(5):381-392.
- Hodnett, G.L., Hale, A.L., Packer, D.J., Stelly, D.M., da Silva, J. & Rooney, W.L. 2010. Elimination of a reproductive barrier facilitates intergeneric hybridization of *Sorghum bicolor* and *Saccharum* *Crop Science*, 50(4):1188-1195.
- Hollingsworth, P.M., Graham, S.W. & Little, D.P. 2011. Choosing and using a plant DNA barcode. *PLoS One*, 6(5):e19254.
- Horsley, T. & Zhou, M. 2013. Effect of photoperiod treatments on pollen viability and flowering at the South African Sugarcane Research Institute. *Proceedings of the South African Sugar Technologists Association*, 86:286-290.
- Huang, Z., Zhu, J., Mu, X. & Lin, J. 2004. Pollen dispersion, pollen viability and pistil receptivity in *Leymus chinensis*. *Annals of Botany*, 93:295-301.
- Ismail, R.M. 2013. Evaluation of genetically modified sugarcane lines carrying *Cry 1AC* gene using molecular marker techniques. *GM Crops Food*, 4(1):58-66.
- James, C. 2016. Global Status of Commercialized Biotech/GM Crops: 2016. ISAAA Brief 52. ISBN: 978-1-892456-66-4
- James, G. 2004. Sugarcane. Blackwell, Oxford.
<http://base.dnsgb.com.ua/files/book/Agriculture/Cultures/Sugarcane.pdf>. Date of access: 02 June 2017. ISBN 0-632-05476-X
- Jansen van Rijssen, F.W., Morris, E.J. & Eloff, J.N. 2015. The precautionary principle: Making managerial decisions on GMOs is difficult. *South African Journal of Science*, 111(3):1-9.
- Johnson, K.L., Raybould, A.F., Hudson, M.D. & Poppy, G.M. 2007. How does scientific risk assessment of GM crops fit within the wider risk analysis? *Trends in Plant Science*, 12(1):1-5.
- Joyce, P., Hermann, S., O'Connell, A., Dinh, Q., Shumbe, L. & Lakshmanan, P. 2014. Field performance of transgenic sugarcane produced using *Agrobacterium* and biolistics methods. *Plant Biotechnology Journal*, 12(4):411-424.
- Jukes, T.H. & Cantor, C.R. 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), *Mammalian protein metabolism*. New York: Academic Press, pp. 21-132.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647-1649.
- Kenkel, N.C. 2006. On selecting an appropriate multivariate analysis. *Canadian Journal of Plant Science*, 86:663–676.
- Khan, K.S., Kunz, R., Kleijnen, J. & Antes, G. 2003. Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 93:118-121.
- Komape, D.M. 2018. Spatial assessment of *Saccharum* species hybrids (sugarcane) and wild relatives in eastern South Africa. Potchefstroom: NWU. (Dissertation-MSc).
- Kress, W.J., Erickson, D.L., Jones, F.A., Swenson, N.G., Perez, R., Sanjur, O. & Bermingham, E. 2009. Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proceedings of the National Academy of Sciences of the United States of America*, 106(44):18621-18626.

- Kumar, S., Stecher, G. & Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7):1870-1874.
- Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y. & Chen, S. 2015. Plant DNA barcoding: from gene to genome. *Biological Reviews of the Cambridge Philosophical Society*, 90(1):157-166.
- Liu, Q., Liu, H. & Wen, J. 2017. Relationships between *Sorghum bicolor* (Poaceae) and its close relatives based on genomic *in situ* hybridization evidence. *Turkish Journal of Botany*, 41:11-24.
- Liu, J., Pang, C., Wei, H., Song, M., Meng, Y., Fan, S., & Yu, S. 2014. Proteomic analysis of anthers from wild-type and photosensitive genetic male sterile mutant cotton (*Gossypium hirsutum* L.). *BMC Plant Biology*. 14(390):1-16.
- Mashoko, L., Mbohwa, C. & Thomas, V.M. 2010. LCA of the South African sugar industry. *Journal of Environmental Planning and Management*, 53(6):793-807.
- McIntyre, C.L. & Jackson, P.A. 2001. Low level of selfing found in a sample of crosses in Australian sugarcane breeding programs. *Euphytica*, 117:245–249.
- Melloni, M.L.G., Melloni, M.N.G., Scarpari, M.S., Garcia, J.C., Landell, M.G.A. & Pinto, L.R. 2015. Flowering of sugarcane genotypes under different artificial photoperiod conditions. *American Journal of Plant Sciences*, 6(03):456-463.
- Melloni, M.L.G., Scarpari, M.S., de Mendonça, J.R., Perecin, D., de Andrade Landell, M.G. & Pinto, L.R. 2012. Comparison of two staining methods for pollen viability studies in sugarcane. *Sugar Technology*, 15(1):103-107.
- Meyer, G.M. & Snyman, S.J. 2013. Progress in research on genetically modified sugarcane in South Africa and associated regulatory requirements. *Acta Horticulture*, 974:43-50.
- Mitchell, H.J. 2011. Regulation of genetically modified (GM) sugarcane in Australia. *Proceedings of the Australian Society of Sugar Cane Technologists*, 33:1-8.
- Mnisi, M.S. & Dlamini, C.S. 2012. The concept of sustainable sugarcane production: Global, African and South African perceptions. *African Journal of Agricultural Research*, 7(31):4337-4343.
- Mohan, C. 2016. Genome editing in sugarcane: Challenges ahead. *Frontiers in Plant Science*, 7(1542):1-5.
- Mohlala, L.M., Bodunrin, M.O., Awosusi, A.A., Daramola, M.O., Cele, N.P. & Olubambi, P.A. 2016. Beneficiation of corncob and sugarcane bagasse for energy generation and materials development in Nigeria and South Africa: A short overview. *Alexandria Engineering Journal*, 55(3):3025-3036.
- Moore, P.H. & Botha, F.C. 2013. Sugarcane, Wiley, Somerset. Available from: ProQuest Ebook Central. [30 June 2017]. ISBN: 9781118771280
- Nair, M.K. 1975. Cytogenetics of *Saccharum officinarum* L. and *S. spontaneum* L. IV. chromosome number and meiosis in *S. officinarum* X *S. spontaneum* hybrids. *Caryologia*, 28(1):1-14.
- Nair, N.V., Nair, S., Sreenivasan, T.V. & Mohan, M. 1999. Analysis of genetic diversity and phylogeny in *Saccharum* and related genera using RAPD markers. *Genetic Resources and Crop Evolution*, 46:73-79.
- Nair, N.V., Selvi, A., Sreenivasan, T.V., Pushpalatha, K.N. & Mary, S. 2005. Molecular diversity among *Saccharum*, *Erianthus*, *Sorghum*, *Zea* and their hybrids. *Sugar Technology*, 7(1):55-59.

Nickson, T.E. 2008. Planning environmental risk assessment for genetically modified crops: problem formulation for stress-tolerant crops. *Plant Physiology*, 147(2):494-502.

OECD (Organisation for Economic Cooperation and Development) (2013). Consensus document on the biology of sugarcane (*Saccharum* spp.), Series on the harmonisation of regulatory oversight in biotechnology No. 56, OECD Environment Directorate, Paris.

Oliveira, J.M.S., Martins, M.S., Dorneles, M.P. & Freitas, C.C. 2015. Starch distribution in anthers, microspores and pollen grains in *Aechmea recurvata* (Klotzsch.) L.B.Sm., *Dyckia racinae* L.B.Sm. and *Tillandsia aeranthos* (Loisel.) L.B.Sm. (Bromeliaceae). *Acta Botanica Brasílica*, 29(1):103-112.

Ostertagová, E., Ostertag, O. & Kováč, J. 2014. Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials*, 611:115-120.

Pagliarini, M.S. 2000. Meiotic behavior of economically important plant species: the relationship between fertility and male sterility. *Genetics and Molecular Biology*, 23(4):997-1002.

Palma-Silva, C., Santos, D.G.D., Kaltchuk-Santos, E. & Bodanese-Zanettini, M.H. 2004. Chromosome numbers, meiotic behavior, and pollen viability of species of *Vriesea* and *Aechmea* genera (Bromeliaceae) native to Rio Grande do Sul, Brazil. *American Journal of Botany*, 91(6):804-807.

Parfitt, R.C. 2005. Release of sugarcane varieties in South Africa. *Proceedings of the South African Sugar Technologists Association*, 79:63-71.

Parthasarathy, N. 1948. Origin of Noble Sugar-canes (*Saccharum officinarum* L.). *Nature*, 161:608.

Pereira, A.R., Barbieri, V. & Nova, N.A.V. 1983. Climatic conditioning of flowering induction in sugarcane. *Agricultural Meteorology*, 29:103-110.

Piperidis, G., Christopher, M.J., Carroll, B.J., Berding, N. & D'Hont, A. 2000. Molecular contribution to selection of intergeneric hybrids between sugarcane and the wild species *Erianthus arundinaceus*. *Genome*, 43:1033-1037.

Price, B. & Cotter, J. 2014. The GM Contamination Register: a review of recorded contamination incidents associated with genetically modified organisms (GMOs), 1997-2013. *International Journal of Food Contamination*, 1(5):1-13

Price, S. 1959. Critique on apomixis in Sugarcane. *Economic Botany*, 13(1):76-74.

Quillet, M.C., Madjidian, N., Griveau, Y., Serieys, H., Tersac, M., Lorieux, M. & Bervilld, A. 1995. Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. *Theoretical and Applied Genetics*, 91:1195-1202.

Raghavan, T.S. 1952. Sugarcane x bamboo hybrids. *Nature*, 170(4321):329-330.

Redshaw, K.A. & Nuss, K.J. 2001. Yield and quality differences between irrigated sugarcane varieties. *Proceedings of the South African Sugar Technologists Association*, 75:160-164.

Redshaw, K.A., Donaldson, R.A. & McIntyre, R.K. 1999. Yield differences between five irrigated sugarcane varieties. *Proceedings of the South African Sugar Technologists Association*, 73:137-146.

Rezaei, M., Arzani, A. & Sayed-Tabatabaei, B.E. 2010. Meiotic behaviour of tetraploid wheats (*Triticum turgidum* L.) and their synthetic hexaploid wheat derivatives influenced by meiotic restitution and heat stress. *Journal of Genetics*, 89(4):401-407.

- Rieben, S., Kalinina, O., Schmid, B. & Zeller, S.L. 2011. Gene flow in genetically modified wheat. *PLoS One*, 6(12):e29730.
- Saitou, N. & Nei, M. 1987. The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4:406–425.
- Sanchez, M.A., Cid, P., Navarrete, H., Aguirre, C., Chacon, G., Salazar, E. & Prieto, H. 2016. Outcrossing potential between 11 important genetically modified crops and the Chilean vascular flora. *Plant Biotechnology Journal*, 14(2):625-637.
- Shanmugavadivu, R. & Rao, P.N.G. 2009. A comparison of flowering behaviour of sugarcane clones in two different locations. *Sugar Technology*, 11(4):401-404.
- Silva, N., Mendes-Bonato, A.B., Sales, J.G. & Pagliarini, M.S. 2011. Meiotic behavior and pollen viability in *Moringa oleifera* (Moringaceae) cultivated in southern Brazil. *Genetics and Molecular Research*, 10(3):1728-1732.
- Skendzic, E.M., Columbus, J.T. & Cerros-Tlatilpa, R. 2007. Phylogenetics of Andropogoneae (Poaceae: Panicoideae) based on nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. *Aliso: A Journal of Systematic and Evolutionary Botany*, 23(1):530–544.
- Smith, D.M., Inman-Bamber, N.G. & Thorburn, P.J. 2005. Growth and function of the sugarcane root system. *Field Crops Research*, 92(2-3):169-183.
- Smyth, S.J. 2017. Genetically modified crops, regulatory delays, and international trade. *Food and Energy Security*, 6(2):78-86.
- Snow, A.A. 2002. Transgenic crops- Why gene flow matters. *Nature Biotechnology*, 20:542.
- Snow, A.A., Moran-Palma, P., Rieseberg, L.H., Wszelaki, A. & Seiler, G.J. 1998. Fecundity, phenology, and seed dormancy of F1 wild-crop hybrids in sunflower (*Helianthus annuus*, Asteraceae). *American Journal of Botany*, 85(6):794-801.
- Snow, A.A., Uthus, K. L. & Culley, T.M. 2001. Fitness of hybrids between weedy and cultivated radish: Implications for weed evolution. *Ecological Applications*, 11(3):934-943.
- Snyman, S.J. & Meyer, G.M. 2012. Improvement of sugarcane in South Africa using genetic engineering: Requirements for potential commercialisation. *Proceedings of the Australian Society of Sugar Cane Technologists*, 85(96):96-101.
- Snyman, S.J., Baker, C., Rutherford, R.S., Watt, D.A., Hockett, B.I., McFarlane, S.A., van Antwerpen, T., Berry, S. & Omarjee, J. 2008. South African Sugarcane Research Institute: Embracing biotechnology for crop improvement research. *Sugar Technology*, 10(1):1-13.
- Snyman, S. J., Komape, D. M., Khanyi, H., van den Berg, J., Cilliers, D., Lloyd Evans, D., Barnard, S. & Siebert, S. J. 2018 Assessing the likelihood of gene flow from sugarcane (*Saccharum* hybrids) to wild relatives in South Africa. *Frontiers in Bioengineering and Biotechnology*, 6(72):1-23.
- Soares, T.L., De Jesus, O.N., Dos Santos-Serejo, J.A. & De Oliveira, E.J. 2013a. *In vitro* pollen germination and pollen viability in passion fruit (*Passiflora* spp.). *Revista Brasileira de Fruticultura*, 35(4):1116-1126.
- Soares, T.L., Jesus, O.N., Souza, E.H., Santos-Serejo, J.A. & Oliveira, E.J. 2013b. Morphology and viability of pollen grains from passion fruit species (*Passiflora* spp.). *Acta Botanica Brasilica*, 27(4):779-787.

- Sooknandan, S., Snyman, S.J., Potier, B.A.M. & Hockett, B.I. 2003. Progress in the development of mosaic resistant sugarcane via transgenesis. *Proceedings of the South African Sugar Technologists Association*, 77:624-627.
- Soreng, R.J., Peterson, P.M., Romaschenko, K., Davidse, G., Zuloaga, F.O., Judziewicz, E.J., Filgueiras, T.S., Davis, J.I. & Morrone, O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53(2):117-137.
- Srivastava, R.P., Singh, S.E., Singh, H. & Singh, S.B. 2006. Artificial induction of flowering in sugarcane under sub-tropical conditions - A successful approach. *Sugar Technology*, 8(2):184-186.
- Stewart, C.N., Jr., Halfhill, M.D. & Warwick, S.I. 2003. Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews Genetics*, 4(10):806-817.
- Su, X., Liu, Y.P., Chen, Z. & Chen, K.L. 2016. Evaluation of candidate barcoding markers in *Orinus* (Poaceae). *Genetics and Molecular Research*, 15(2):1-14.
- Sulusoglu, M. & Cavusoglu, A. 2014. In vitro pollen viability and pollen germination in cherry laurel (*Prunus laurocerasus* L.). *The Scientific World Journal*, 2014:1-8.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, 28(10):2731–2739.
- Tavares de Oliveira Melo, A. 2016. Exploring genomic databases for *in silico* discovery of *Pht1* genes in high syntenic close related grass species with focus in sugarcane (*Saccharum* spp.). *Current Plant Biology*, 6:11-18.
- Teshome, M. & Feyissa, T. 2013. Genetic diversity of Wild Sorghum (*Sorghum bicolor* ssp. *verticilliflorum* (L.) Moench) germplasm from Ethiopia as revealed by ISSR markers. *Asian Journal of Plant Sciences*, 12(3):137-144.
- The Plant List. 2013. Version 1.1. Published on the Internet; <http://www.theplantlist.org/> (accessed 1 October 2017).
- Tripathi, L. 2005. Techniques for detecting genetically modified crops and products. *African Journal of Biotechnology*, 4 (13):1472-1479.
- Van den Brink, P.J., Van den Brink, N.W. & Ter Braak, C.J. F. 2003. Multivariate analysis of ecotoxicological data using ordination: Demonstrations of utility on the basis of various examples. *Australasian Journal of Ecotoxicology*, 9:141-156.
- van der Vyver, C., Conradie, T., Kossmann, J. & Lloyd, J. 2013. *In vitro* selection of transgenic sugarcane callus utilizing a plant gene encoding a mutant form of acetolactate synthase. *In Vitro Cellular & Developmental Biology – Plant*, 49(2):198-206.
- Venkatraman, T.S. 1938. Agriculture. Hybridization in and with the genus *Saccharum*. Its scientific and economic aspects. *Current Science*, 6(8):425-427.
- Wafula, D., Waithaka, M., Komen, J. & Karembu, M. 2012. Biosafety legislation and biotechnology development gains momentum in Africa. *GM Crops Food*, 3(1):72-77.
- Warwick, S.I., Beckie, H.J. & Hall, L.M. 2009. Gene flow, invasiveness, and ecological impact of genetically modified crops. *Annals of the New York Academy of Sciences*, 1168:72-99.
- Watt, D.A., Sweby, D.L., Potier, B.A.M. & Snyman, S.J. 2010. Sugarcane genetic engineering research in South Africa: From gene discovery to transgene expression. *Sugar Technology*, 12(2):85-90.

- Welker, C.A., Souza-Chies, T.T., Longhi-Wagner, H.M., Peichoto, M.C., McKain, M.R. & Kellogg, E.A. 2015. Phylogenetic analysis of *Saccharum* s.l. (Poaceae; Andropogoneae), with emphasis on the circumscription of the South American species. *American Journal of Botany*, 102(2):248-263.
- Wu, J., Huang, Y., Lin, Y., Fu, C., Liu, S., Deng, Z., Li, Q., Huang, Z., Chen, R. & Zhang, M. 2014. Unexpected inheritance pattern of *Erianthus arundinaceus* chromosomes in the intergeneric progeny between *Saccharum* spp. and *Erianthus arundinaceus*. *PLoS One*, 9(10):e110390.
- Xue, B., Guo, J., Que, Y., Fu, Z., Wu, L. & Xu, L. 2014. Selection of suitable endogenous reference genes for relative copy number detection in sugarcane. *International Journal of Molecular Sciences*, 15(5):8846-8862.
- Yao, W., Ruan, M., Qin, L., Yang, C., Chen, R., Chen, B. & Zhang, M. 2017. Field performance of transgenic sugarcane lines resistant to sugarcane mosaic virus. *Frontiers in Plant Science*, 8:(104):1-9
- Yu, J., Xue, J.-H. & Zhou, S.-L. 2011. New universal *matK* primers for DNA barcoding angiosperms. *Journal of Systematics and Evolution*, 49(3):176-181.
- Zhang, H., Mittal, N., Leamy, L.J., Barazani, O. & Song, B.H. 2017. Back into the wild- Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, 10(1):5-24.
- Zhang, H.Y., He, L.L., Zhong, H.Q., Li, F.S., He, S.C. & Yang, Q.H. 2009. Identification of intergeneric hybrids between *Saccharum* spp. and *Erianthus fulvus* with ITSS. *African Journal of Biotechnology*, 8(9): 1841-1845.
- Zhou, D., Wang, C., Li, Z., Chen, Y., Gao, S., Guo, J., Lu, W., Su, Y., Xu, L. & Que, Y. 2016. Detection of *bar* transgenic sugarcane with a rapid and visual loop-mediated isothermal amplification assay. *Frontiers in Plant Science*, 7:(279):1-11.
- Zhou, M. 2004. Performance of varieties N14 and NCo376 in the south-east lowveld of Zimbabwe. *Proceedings of the South African Sugar Technologists Association*, 78:137-148.
- Zhou, M. 2013. Conventional sugarcane breeding in South Africa: Progress and future prospects. *American Journal of Plant Sciences*, 4:189-197.

APPENDICES

Appendix A-1: Andropogoneae members used in the phylogenetic analysis to confer relatedness between sugarcane varieties and related species

Species name	Origin	Occurrence within study area	Accession no	Place of collection
<i>Saccharum robustum</i>	New Guinea, Asia	No	14591	KZN, SASRI nursery
<i>Saccharum spontaneum</i>	New Guinea, Asia	No	15590	KZN, SASRI nursery
<i>Saccharum ravennae</i>	New Guinea, Asia	No	14593	KZN, SASRI nursery
<i>Saccharum arundinaceum</i>	New Guinea, Asia	No	14594	KZN, SASRI nursery
<i>Saccharum officinarum</i>	New Guinea, Asia	No	14586	KZN, SASRI nursery
<i>Saccharum sinense</i>	New Guinea, Asia	No	14597	KZN, SASRI nursery
<i>Saccharum</i> sp. Rowan Green	Sacharum hybrid	No	14598	KZN, SASRI nursery
<i>Saccharum</i> sp. Co745	Sacharum hybrid	No	14600	KZN, SASRI nursery
<i>Saccharum</i> sp. N14	Sacharum hybrid	Cultivated	14614	MP, Komatipoort sugarcane plantation
<i>Saccharum</i> sp. N36	Sacharum hybrid	Cultivated	14606	MP, Malelane sugarcane plantation
<i>Saccharum</i> sp. NCo376	Sacharum hybrid	Cultivated	14656	KZN, SASRI germplasm nursery
<i>Imperata cylindrica</i>	South east Asia	Yes	14604	KZN, Mount Edgecombe sugarcane plantation margin
<i>Miscanthus ecklonii</i>	Southern Africa	Yes	14652	KZN, New hanover roadside

<i>Sorghum arundinaceum</i>	Northeastern Africa	Yes	14603	KZN, SASRI weed garden
<i>Sorghum x drummondii</i>	Northeastern Africa	Yes	14646	KZN, Empangeni sugarcane plantation
<i>Trachypogon spicatus</i>	Southern Africa	yes	14655	NW, Ikageng roadside
<i>Miscanthus junceus</i>	Southern Africa	Yes	14654	KZN, Paulpietersburg, seasonal pond
<i>Sorghum versicolor</i>	Africa	Yes	10278	NW, Potchefstroom roadside
<i>Sorghum halepense</i>	Asia	Yes	GenBank	
<i>Chrysopogon zizanioides</i>	India	Yes	GenBank	
<i>Bothriochloa insculpta</i>	Africa, Asia, Europe	Yes	GenBank	
<i>Panicum maximum</i>	Middle America	Yes	GenBank	
<i>Sorghastrum nudipes</i>	Africa	Yes	GenBank	
<i>Microstegium nudum</i>	Tropical Asia	Yes	GenBank	
<i>Zea mays</i>	central America	Cultivated	GenBank	

KZN=KwaZulu-Natal, MP=Mpumalanga, NW=North-West

Appendix A-2: Herbarium specimen information of the Andropogoneae members used in the phylogenetic analysis

Species name	Collected by	Collector No.	Collection date	Identified by	GPS coordinates	Barcode
<i>Saccharum arundinaceum</i> Retz.	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0006	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0014594
<i>Saccharum ravennae</i> (L.) L.	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0005	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0014593
<i>Imperata cylindrica</i> (L.) Raeusch.	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0016	23/05/2016	Siebert, S.J.	29°42'14" S, 31°03'05" E	PUC0014604
<i>Miscanthus ecklonii</i> (Nees) Mabb.	Komape, D.M., Khanyi, H.	0047	05/03/2017	Komape, D.M.	29°19'20" S, 30°28'59" E	PUC0014652
<i>Miscanthus junceus</i> (Stapf) Pilg.	Komape, D.M., Khanyi, H.	0049	08/03/2017	Komape, D.M.	27°25'30" S, 30°48'21" E	PUC0014654
<i>Saccharum</i> sp. N14	Komape, D.M., Khanyi, H.	0009	04/07/2016	SASRI	25°33'01" S, 31°57'02" E	PUC0014614
<i>Saccharum</i> sp. N36	Komape, D.M., Khanyi, H.	0001	05/05/2016	SASRI	25°36'08" S, 31°33'30" E	PUC0014606
<i>Saccharum</i> sp. Co745	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0012	23/05/2016	SASRI	29°42'27" S, 31°03'00" E	PUC0014600
<i>Saccharum</i> sp. NCo376	Komape, D.M., Khanyi, H.	0051	09/07/2016	SASRI	29°42'12" S, 31°02'35" E	PUC0014656
<i>Saccharum</i> sp. Rowan Green	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0010	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0014598
<i>Saccharum sinense</i> Roxb.	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0009	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0014597
<i>Saccharum officinarum</i> L.	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0008	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0014586
<i>Saccharum robustum</i>	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0003	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0014591

Species name	Collected by	Collector No.	Collection date	Identified by	GPS coordinates	Barcode
E.W.Brandes & Jeswiet ex Grassl						
<i>Saccharum spontaneum</i> L.	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0002	23/05/2016	SASRI	29°42'18" S, 31°02'41" E	PUC0015590
<i>Sorghum arundinaceum</i> (Desv.) de Wet & J.R. Harlan	Komape, D.M., Siebert, S.J., Barnard, S., Khanyi, H	0015	23/05/2017	Mashau, A.C., Komape, D.M	29°42'14" S, 31°02'38" E	PUC0014603
<i>Sorghum x drummondii</i> (Nees ex Steud.) de Wet	Komape, D.M., Khanyi, H.	0041	08/03/2017	Mashau, A.C., Komape, D.M	28°43'55" S, 31°43'54" E	PUC0014646
<i>Sorghum versicolor</i> Andersson	Cilliers, S.S.	2504	12/03/1996	Cilliers SS	26°37'30" S, 27°07'30" E	PUC0010278
<i>Trachypogon spicatus</i> (L.f.) Kuntze	Komape, D.M., Khanyi, H.	0050	09/03/2017	Mashau, A.C., Komape, D.M	26°42'31" S, 27°02'25" E	PUC0014655

Appendix B 1-3: Best substitution model for each DNA barcode determined using the “find best DNA/Protein Models” within MEGA 7. Kimura 2-parameter (K2) was calculated to be the best DNA substitution models for all the barcodes. The analysis involved 24 nucleotide sequences and analysed for the ITS (B-1), *rbcL* (B-2) and *matK* (B-3) DNA barcodes

Note: Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) are also presented. Non-uniformity of evolutionary rates among sites may be modelled by using a discrete Gamma distribution (+G) with 5 rate categories and by assuming that a certain fraction of sites was evolutionarily invariable (+I). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown. Assumed or estimated values of transition/transversion bias (R) are shown for each model, as well. They are followed by nucleotide frequencies (freq.) for each nucleotide pair

Abbreviations: GTR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor

Appendix B-1: Best model for the ITS DNA barcode

Model	#Param	BIC	AICc	lnL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G
K2	48	1219,886951	968,9578516	-434,8295569	n/a	n/a	1,74	0,25	0,25	0,25	0,25
K2+G	49	1226,014927	969,929273	-434,2453383	n/a	5,00	1,81	0,25	0,25	0,25	0,25
K2+I	49	1226,540507	970,4548537	-434,5081286	0,048	n/a	1,77	0,25	0,25	0,25	0,25
T92	49	1227,567505	971,4818512	-435,0216273	n/a	n/a	1,74	0,20	0,20	0,30	0,30
T92+G	50	1232,622564	971,3833615	-433,9009504	n/a	3,46	1,86	0,20	0,20	0,30	0,30
K2+G+I	50	1233,212873	971,97367	-434,1961047	0,00	5,00	1,81	0,25	0,25	0,25	0,25
T92+I	50	1233,340073	972,1008705	-434,2597049	0,07	n/a	1,82	0,20	0,20	0,30	0,30

Model	#Param	BIC	AICc	lnL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G
JC	47	1237,587976	991,8184309	-447,3282764	n/a	n/a	0,5	0,25	0,25	0,25	0,25
T92+G+I	51	1239,821335	973,4315947	-433,852129	0	3,46	1,86	0,20	0,20	0,30	0,30
HKY	51	1240,646407	974,2566664	-434,2646649	n/a	n/a	1,76	0,23	0,18	0,32	0,28
JC+G	48	1243,632797	992,7036977	-446,70248	n/a	4,77	0,5	0,25	0,25	0,25	0,25
JC+I	48	1244,494982	993,5658825	-447,1335724	0,04	n/a	0,5	0,25	0,25	0,25	0,25
HKY+G	52	1245,778887	974,241628	-433,1826987	n/a	3,54	1,88	0,23	0,18	0,32	0,28
HKY+I	52	1246,994424	975,4571648	-433,7904671	0,06	n/a	1,82	0,23	0,18	0,32	0,28
TN93	52	1247,835898	976,2986388	-434,2112041	n/a	n/a	1,77	0,23	0,18	0,32	0,28
JC+G+I	49	1250,875759	994,790105	-446,6757542	0	4,77	0,5	0,25	0,25	0,25	0,25
HKY+G+I	53	1252,978284	976,2965296	-433,1341902	0,00	3,54	1,88	0,23	0,18	0,32	0,28
TN93+G	53	1253,040299	976,3585452	-433,165198	n/a	3,58	1,88	0,23	0,18	0,32	0,28
TN93+I	53	1254,234012	977,5522578	-433,7620543	0,06	n/a	1,82	0,23	0,18	0,32	0,28
TN93+G+I	54	1260,243531	978,4203126	-433,118607	0	3,58	1,88	0,23	0,18	0,32	0,28
GTR	55	1267,563856	980,6022117	-433,1305632	n/a	n/a	1,74	0,23	0,18	0,32	0,28
GTR+G	56	1272,623485	980,5264577	-432,012171	n/a	3,55	1,86	0,23	0,18	0,32	0,28
GTR+I	56	1274,29505	982,198023	-432,8479537	0,04	n/a	1,78	0,23	0,18	0,32	0,28
GTR+G+I	57	1279,920031	982,6906712	-432,0122375	0,00	3,55	1,86	0,23	0,18	0,32	0,28

Appendix B-2: Best model for the *rbcL* DNA barcode

Model	#Param	BIC	AICc	lnL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G
K2	48	2054,233832	1700,695764	-802,1473021	n/a	n/a	2,01	0,25	0,25	0,25	0,25
JC	47	2054,589597	1708,408855	-807,0120512	n/a	n/a	0,5	0,25	0,25	0,25	0,25
K2+G	49	2056,012039	1695,116988	-798,3495386	n/a	0,05	2,07	0,25	0,25	0,25	0,25
JC+G	48	2056,476169	1702,938101	-803,2684707	n/a	0,05	0,5	0,25	0,25	0,25	0,25
K2+I	49	2060,110348	1699,215296	-800,3986928	0,48	n/a	2,02	0,25	0,25	0,25	0,25
T92	49	2060,143918	1699,248866	-800,4154778	n/a	n/a	2,01	0,27	0,27	0,23	0,23
JC+I	48	2060,49049	1706,952422	-805,2756309	0,48	n/a	0,5	0,25	0,25	0,25	0,25
T92+G	50	2062,056586	1693,804895	-796,6849449	n/a	0,05	2,06	0,27	0,27	0,23	0,23
K2+G+I	50	2064,218205	1695,966514	-797,7657544	0,46	0,05	2,14	0,25	0,25	0,25	0,25
JC+G+I	49	2064,80462	1703,909568	-802,7458288	0,45	0,05	0,5	0,25	0,25	0,25	0,25
T92+I	50	2066,050683	1697,798993	-798,6819937	0,48	n/a	2,02	0,27	0,27	0,23	0,23
T92+G+I	51	2070,410825	1694,802839	-796,1751975	0,44	0,05	2,11	0,27	0,27	0,23	0,23
HKY	51	2077,868353	1702,260367	-799,9039614	n/a	n/a	2,01	0,26	0,29	0,22	0,23
HKY+G	52	2079,803865	1696,839929	-796,1848509	n/a	0,05	2,06	0,26	0,29	0,22	0,23
HKY+I	52	2083,780357	1700,81642	-798,1730966	0,48	n/a	2,02	0,26	0,29	0,22	0,23
TN93	52	2084,827787	1701,86385	-798,6968117	n/a	n/a	2,02	0,26	0,29	0,22	0,23
TN93+G	53	2087,096874	1696,777331	-795,1444884	n/a	0,05	2,06	0,26	0,29	0,22	0,23
HKY+G+I	53	2088,180969	1697,861426	-795,6865359	0,44	0,05	2,10	0,26	0,29	0,22	0,23
TN93+I	53	2090,810829	1700,491285	-797,0014654	0,48	n/a	2,03	0,26	0,29	0,22	0,23
TN93+G+I	54	2095,80575	1698,130944	-794,812059	0,37	0,05	2,09	0,26	0,29	0,22	0,23
GTR	55	2107,111459	1702,081735	-795,7780466	n/a	n/a	1,12	0,26	0,29	0,22	0,23
GTR+G	56	2109,102317	1696,718021	-792,0866089	n/a	0,05	1,19	0,26	0,29	0,22	0,23
GTR+I	56	2113,119595	1700,735299	-794,0952479	0,48	n/a	1,13	0,26	0,29	0,22	0,23
GTR+G+I	57	2117,479562	1697,741036	-791,5883641	0,31	0,05	1,26	0,26	0,29	0,22	0,23

Appendix B-3: Best model for the *matK* barcode

Model	#Param	BIC	AICc	lnL	Invariant	Gamma	R	Freq A	Freq T	Freq C	Freq G
K2	48	2646,831916	2332,569156	-1117,832357	n/a	n/a	2,10	0,25	0,25	0,25	0,25
K2+I	49	2654,277944	2333,487067	-1117,27238	0,01	n/a	2,10	0,25	0,25	0,25	0,25
K2+G	49	2655,463195	2334,672318	-1117,865005	n/a	200	2,10	0,25	0,25	0,25	0,25
T92	49	2660,838322	2340,047445	-1120,552569	n/a	n/a	2,17	0,32	0,32	0,18	0,18
K2+G+I	50	2662,852949	2335,53474	-1117,276891	0	200	2,10	0,25	0,25	0,25	0,25
T92+I	50	2668,282479	2340,964269	-1119,991656	0	n/a	2,17	0,32	0,32	0,18	0,18
T92+G	50	2669,458823	2342,140613	-1120,579828	n/a	200	2,18	0,32	0,32	0,18	0,18
HKY	51	2674,286141	2340,441382	-1118,710495	n/a	n/a	2,19	0,27	0,37	0,19	0,17
T92+G+I	51	2676,84769	2343,002931	-1119,991269	0	200	2,18	0,32	0,32	0,18	0,18
HKY+I	52	2681,721375	2341,350852	-1118,14512	0	n/a	2,19	0,27	0,37	0,19	0,17
TN93	52	2682,422122	2342,0516	-1118,495494	n/a	n/a	2,21	0,27	0,37	0,19	0,17
HKY+G	52	2682,899466	2342,528944	-1118,734166	n/a	65,52	2,20	0,27	0,37	0,19	0,17
TN93+I	53	2689,850358	2342,954857	-1117,92662	0,00	n/a	2,21	0,27	0,37	0,19	0,17
HKY+G+I	53	2690,278594	2343,383093	-1118,140738	0,01	66,44	2,20	0,27	0,37	0,19	0,17
TN93+G	53	2691,038123	2344,142621	-1118,520502	n/a	78,29	2,22	0,27	0,37	0,19	0,17
TN93+G+I	54	2698,409985	2344,990291	-1117,923442	0	79,55	2,22	0,27	0,37	0,19	0,17
GTR	55	2703,903957	2343,960856	-1116,387436	n/a	n/a	1,82	0,27	0,37	0,19	0,17
GTR+I	56	2706,575146	2340,109425	-1113,440039	0,01	n/a	2,20	0,27	0,37	0,19	0,17
GTR+G	56	2707,517368	2341,051647	-1113,91115	n/a	29,93	2,24	0,27	0,37	0,19	0,17
GTR+G+I	57	2715,006758	2342,019204	-1113,372853	0,99	30,28	2,24	0,27	0,37	0,19	0,17
JC	47	2718,422135	2410,688276	-1157,910459	n/a	n/a	0,5	0,25	0,25	0,25	0,25
JC+I	48	2725,932309	2411,66955	-1157,382554	0	n/a	0,5	0,25	0,25	0,25	0,25
JC+G	48	2734,303579	2420,04082	-1161,568189	n/a	8,75	0,5	0,25	0,25	0,25	0,25
JC+G+I	49	2734,55974	2413,768863	-1157,413278	0	200	0,5	0,25	0,25	0,25	0,25

Appendix C1-3: The number of base substitutions per site from between sequences are shown for the ITS (C-1), *matK* (C-2) and *rbcL* (C-3) barcodes.

Analyses were conducted using the Kimura 2-parameter model. The analysis involved 24 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 59 (ITS), 210 (*matK*) and 471 (*rbcL*) positions in the final dataset. Evolutionary analyses were conducted in MEGA7.

C-1: ITS DNA barcode

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Bothriochloa insculpta																								
2. Chrysopogon zizanioides	0.073																							
3. Saccharum arundinaceum	0.788	0.620																						
4. Saccharum ravennae	0.769	0.710	0.072																					
5. Imperata cylindrica	0.697	0.647	0.090	0.072																				
6. Microstegium nudum	0.090	0.090	0.710	0.815	0.686																			
7. Miscanthus ecklonii	0.726	0.670	0.017	0.053	0.071	0.710																		
8. Miscanthus junceus	0.788	0.620	-0.000	0.072	0.090	0.710	0.017																	
9. Saccharum sp. N14	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017																
10. Saccharum sp. N36	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000															
11. Saccharum sp. Co745	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000	-0.000														
12. Saccharum sp. NCo376	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000	-0.000	-0.000													
13. Saccharum sp. Rowan Green	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000	-0.000	-0.000	-0.000												
14. Saccharum sinense	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000	-0.000	-0.000	-0.000	-0.000											
15. Saccharum officinarum	0.974	0.897	0.214	0.261	0.285	0.952	0.192	0.214	0.192	0.192	0.192	0.192	0.192	0.192										
16. Saccharum robustum	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.192									
17. Saccharum spontaneum	0.697	0.647	0.090	0.072	-0.000	0.686	0.071	0.090	0.071	0.071	0.071	0.071	0.071	0.071	0.285	0.071								
18. Sorghastrum nudipes	0.864	0.942	0.247	0.225	0.197	1.001	0.221	0.247	0.221	0.221	0.221	0.221	0.221	0.221	0.494	0.221	0.197							
19. Sorghum arundinaceum	0.726	0.670	0.017	0.053	0.071	0.710	-0.000	0.017	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.192	-0.000	0.071	0.221						
20. Sorghum x drummondii	0.970	0.970	0.110	0.072	0.110	1.033	0.090	0.110	0.090	0.090	0.090	0.090	0.090	0.090	0.311	0.090	0.110	0.197	0.090					
21. Sorghum versicolor	0.943	0.860	0.073	0.072	0.110	0.913	0.054	0.073	0.054	0.054	0.054	0.054	0.054	0.054	0.262	0.054	0.110	0.197	0.054	0.110				
22. Sorghum halepense	0.110	0.110	0.860	0.913	0.815	0.053	0.860	0.860	0.860	0.860	0.860	0.860	0.860	0.860	1.177	0.860	0.815	1.288	0.860	1.365	1.169			
23. Trachypogon spicatus	0.970	0.970	0.110	0.072	0.110	1.033	0.090	0.110	0.090	0.090	0.090	0.090	0.090	0.090	0.311	0.090	0.110	0.197	0.090	-0.000	0.110	1.365		
24. Zea mays	0.978	0.978	0.769	0.753	0.864	0.788	0.710	0.769	0.710	0.710	0.710	0.710	0.710	0.710	0.952	0.710	0.864	0.783	0.710	0.739	0.769	0.810	0.739	

C-2: *matK* DNA barcode

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. <i>Bothriochloa insculpta</i>																								
2. <i>Chrysopogon zizanioides</i>	0.225																							
3. <i>Saccharum arundinaceum</i>	0.005	0.232																						
4. <i>Saccharum ravennae</i>	0.005	0.232	-0.000																					
5. <i>Imperata cylindrica</i>	0.005	0.232	-0.000	-0.000																				
6. <i>Miscanthus ecklonii</i>	0.005	0.232	-0.000	-0.000	-0.000																			
7. <i>Miscanthus junceus</i>	1.004	1.061	0.977	0.977	0.977	0.977																		
8. <i>Saccharum</i> sp. N14	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977																	
9. <i>Saccharum</i> sp. N36	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000																
10. <i>Saccharum</i> sp. Co745	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000															
11. <i>Saccharum</i> sp. NCo376	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000														
12. <i>Saccharum</i> sp. Rowan Green	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000													
13. <i>Saccharum sinense</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000												
14. <i>Saccharum officinarum</i>	0.014	0.246	0.010	0.010	0.010	0.010	1.023	0.010	0.010	0.010	0.010	0.010	0.010											
15. <i>Saccharum robustum</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.010										
16. <i>Saccharum spontaneum</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.010	-0.000									
17. <i>Sorghastrum nudipes</i>	0.010	0.239	0.005	0.005	0.005	0.005	0.994	0.005	0.005	0.005	0.005	0.005	0.005	0.010	0.005	0.005								
18. <i>Sorghum x drummondii</i>	0.682	0.738	0.696	0.696	0.696	0.696	n/c	0.696	0.696	0.696	0.696	0.696	0.696	0.671	0.696	0.696	0.700							
19. <i>Sorghum arundinaceum</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.010	-0.000	-0.000	0.005	0.696						
20. <i>Sorghum versicolor</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.010	-0.000	-0.000	0.005	0.696	-0.000					
21. <i>Trachypogon spicatus</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.010	-0.000	-0.000	0.005	0.696	-0.000	-0.000				
22. <i>Microstegium nudum</i>	0.024	0.254	0.019	0.019	0.019	0.019	0.984	0.019	0.019	0.019	0.019	0.019	0.019	0.029	0.019	0.019	0.024	0.720	0.019	0.019	0.019			
23. <i>Sorghum halepense</i>	0.005	0.232	-0.000	-0.000	-0.000	-0.000	0.977	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.010	-0.000	-0.000	0.005	0.696	-0.000	-0.000	-0.000	0.019		
24. <i>Zea mays</i>	0.238	0.263	0.245	0.245	0.245	0.245	1.204	0.245	0.245	0.245	0.245	0.245	0.245	0.259	0.245	0.245	0.251	0.841	0.245	0.245	0.245	0.258	0.245	



Pairwise evolutionary distance was uncalculatable

C-3 *rbcL* DNA barcode

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Bothriochloa insculpta																								
2. Chrysopogon zizanioides	0.013																							
3. Imperata cylindrica	0.004	0.013																						
4. Microstegium nudum	0.006	0.015	0.002																					
5. Miscanthus ecklonii	0.019	0.011	0.019	0.022																				
6. Miscanthus junceus	0.004	0.013	-0.000	0.002	0.019	-0.000																		
7. Saccharum sp. N14	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000																	
8. Saccharum sp. N36	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000																
9. Saccharum sp. Co745	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000															
10. Saccharum sp. NCo376	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000														
11. Saccharum sp. Rowan Green	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000													
12. Saccharum arundinaceum	0.002	0.011	0.002	0.004	0.017	0.002	0.002	0.002	0.002	0.002	0.002	0.002												
13. Saccharum officinarum	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.002												
14. Saccharum ravennae	0.002	0.011	0.002	0.004	0.017	0.002	0.002	0.002	0.002	0.002	0.002	-0.000	0.002											
15. Saccharum robustum	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	0.002										
16. Saccharum sinense	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	0.002	-0.000									
17. Saccharum spontaneum	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	0.002	-0.000	-0.000								
18. Sorghastrum nudipes	0.004	0.013	0.004	0.006	0.019	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.004	0.002	0.004	0.004	0.004							
19. Sorghum x drummondii	0.022	0.019	0.022	0.024	0.017	0.022	0.022	0.022	0.022	0.022	0.022	0.019	0.022	0.019	0.022	0.022	0.022	0.022						
20. Sorghum arundinaceum	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	0.002	-0.000	-0.000	-0.000	0.004	0.022					
21. Sorghum halepense	0.004	0.013	-0.000	0.002	0.019	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.002	-0.000	0.002	-0.000	-0.000	-0.000	0.004	0.022	-0.000				
22. Sorghum versicolor	0.006	0.015	0.002	0.004	0.022	0.002	0.002	0.002	0.002	0.002	0.002	0.004	0.002	0.004	0.002	0.002	0.002	0.006	0.024	0.002	0.002			
23. Trachypogon spicatus	0.004	0.013	0.004	0.006	0.019	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.004	0.002	0.004	0.004	0.004	-0.000	0.022	0.004	0.004	0.006		
24. Zea mays	0.004	0.013	0.004	0.006	0.015	0.004	0.004	0.004	0.004	0.004	0.004	0.002	0.004	0.002	0.004	0.004	0.004	0.004	0.022	0.004	0.004	0.006	0.004	

Appendix D 1-5: Analysis of variance by multiple comparisons of *p*-values (2-tailed) for independent variables, i.e TMN (minimum temperature: D-1), TMX (maximum temperature: D-2), DL (day length: D-3), RH (relative humidity: D-4) and SWC (soil water content: D-5) per site using the Kruskal-Wallis test.

D-1: Mean minimum temperature comparison across the study area

Depend.: TMN	Multiple Comparisons <i>p</i> -values (2-tailed); TMN (Spreadsheet final) Independent (grouping) variable: Site Kruskal-Wallis test: H (8, N= 43) =40,32886 p =,0000								
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
Site 1		1,000000	1,000000	0,842130	<u>0,002325</u>	0,069997	1,000000	0,506359	1,000000
Site 2	1,000000		1,000000	1,000000	0,154653	1,000000	1,000000	1,000000	1,000000
Site 3	1,000000	1,000000		1,000000	0,092918	0,783165	1,000000	1,000000	1,000000
Site 4	0,842130	1,000000	1,000000		0,561674	1,000000	1,000000	1,000000	<u>0,035741</u>
Site 5	<u>0,002325</u>	0,154653	0,092918	0,561674		1,000000	1,000000	1,000000	<u>0,000032</u>
Site 6	0,069997	1,000000	0,783165	1,000000	1,000000		1,000000	1,000000	<u>0,003259</u>
Site 7	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000		1,000000	0,412793
Site 8	0,506359	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000		0,054850
Site 9	1,000000	1,000000	1,000000	<u>0,035741</u>	<u>0,000032</u>	<u>0,003259</u>	0,412793	0,054850	

D-2: Mean maximum temperature comparison across the study area

Depend.: TMX	Multiple Comparisons p -values (2-tailed); TMX (Spreadsheet final) Independent (grouping) variable: Site Kruskal-Wallis test: $H(8, N=43) = 35.45609, p = .0000$								
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
Site 1		1,000000	1,000000	1,000000	1,000000	<u>0,011067</u>	0,596612	0,157405	<u>0,000055</u>
Site 2	1,000000		1,000000	1,000000	1,000000	0,170575	1,000000	1,000000	<u>0,004225</u>
Site 3	1,000000	1,000000		1,000000	1,000000	1,000000	1,000000	1,000000	0,977044
Site 4	1,000000	1,000000	1,000000		1,000000	1,000000	1,000000	1,000000	<u>0,018893</u>
Site 5	1,000000	1,000000	1,000000	1,000000		1,000000	1,000000	1,000000	0,208325
Site 6	<u>0,011067</u>	0,170575	1,000000	1,000000	1,000000		1,000000	1,000000	1,000000
Site 7	0,596612	1,000000	1,000000	1,000000	1,000000	1,000000		1,000000	1,000000
Site 8	0,157405	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000		1,000000
Site 9	<u>0,000055</u>	<u>0,004225</u>	0,977044	<u>0,018893</u>	0,208325	1,000000	1,000000	1,000000	

D-3: Mean day length comparison across the study area

Depend.: DL	Multiple Comparisons p -values (2-tailed); DL (h) (Spreadsheet final) Independent (grouping) variable: Site Kruskal-Wallis test: $H(8, N=43) = 40.65575, p = .0000$								
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
Site 1		1,000000	1,000000	1,000000	0,244728	0,069997	1,000000	<u>0,029247</u>	<u>0,000123</u>
Site 2	1,000000		1,000000	1,000000	0,385817	0,118312	1,000000	<u>0,049165</u>	<u>0,000475</u>
Site 3	1,000000	1,000000		1,000000	1,000000	1,000000	1,000000	1,000000	1,000000
Site 4	1,000000	1,000000	1,000000		1,000000	1,000000	1,000000	0,443368	<u>0,002995</u>
Site 5	0,244728	0,385817	1,000000	1,000000		1,000000	1,000000	1,000000	1,000000
Site 6	0,069997	0,118312	1,000000	1,000000	1,000000		1,000000	1,000000	1,000000
Site 7	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000		1,000000	1,000000
Site 8	<u>0,029247</u>	<u>0,049165</u>	1,000000	0,443368	1,000000	1,000000	1,000000		1,000000

Depend.: DL	Multiple Comparisons p -values (2-tailed); DL (h) (Spreadsheet final) Independent (grouping) variable: Site Kruskal-Wallis test: $H(8, N=43) = 40,65575$ $p = ,0000$								
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
Site 9	<u>0,000123</u>	<u>0,000475</u>	1,000000	<u>0,002995</u>	1,000000	1,000000	1,000000	1,000000	

D-4: Mean relative humidity comparison across the study area

Depend.: RH	Multiple Comparisons p values (2-tailed); RH (Spreadsheet final) Independent (grouping) variable: Site Kruskal-Wallis test: $H(8, N=43) = 39,63624$ $p = ,0000$								
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
Site 1		1,000000	1,000000	<u>0,022915</u>	<u>0,000296</u>	<u>0,009166</u>	1,000000	1,000000	1,000000
Site 2	1,000000		1,000000	1,000000	0,069997	0,557449	1,000000	1,000000	1,000000
Site 3	1,000000	1,000000		0,143887	<u>0,003370</u>	<u>0,040797</u>	1,000000	1,000000	1,000000
Site 4	<u>0,022915</u>	1,000000	0,143887		1,000000	1,000000	1,000000	1,000000	1,000000
Site 5	<u>0,000296</u>	0,069997	<u>0,003370</u>	1,000000		1,000000	0,212436	1,000000	<u>0,021513</u>
Site 6	<u>0,009166</u>	0,557449	<u>0,040797</u>	1,000000	1,000000		0,641205	1,000000	0,288147
Site 7	1,000000	1,000000	1,000000	1,000000	0,212436	0,641205		1,000000	1,000000
Site 8	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000		1,000000
Site 9	1,000000	1,000000	1,000000	1,000000	<u>0,021513</u>	0,288147	1,000000	1,000000	

D-5: Mean soil water content comparison across the study area

Depend.: SWC	Multiple Comparisons p values (2-tailed); SWC100 (Spreadsheet final) Independent (grouping) variable: Site Kruskal-Wallis test: $H(8, N=43) = 36,59586$ $p = ,0000$								
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
Site 1		1,000000	1,000000	1,000000	<u>0,026402</u>	0,329331	1,000000	<u>0,011719</u>	<u>0,000206</u>
Site 2	1,000000		1,000000	1,000000	0,754445	1,000000	1,000000	0,274557	<u>0,029468</u>
Site 3	1,000000	1,000000		1,000000	1,000000	1,000000	1,000000	1,000000	0,426207
Site 4	1,000000	1,000000	1,000000		0,366935	1,000000	1,000000	0,143887	<u>0,002995</u>
Site 5	<u>0,026402</u>	0,754445	1,000000	0,366935		1,000000	1,000000	1,000000	1,000000
Site 6	0,329331	1,000000	1,000000	1,000000	1,000000		1,000000	1,000000	1,000000
Site 7	1,000000	1,000000	1,000000	1,000000	1,000000	1,000000		1,000000	1,000000
Site 8	<u>0,011719</u>	0,274557	1,000000	0,143887	1,000000	1,000000	1,000000		1,000000
Site 9	<u>0,000206</u>	<u>0,029468</u>	0,426207	<u>0,002995</u>	1,000000	1,000000	1,000000	1,000000	

Appendix E: Spearman Rank Order Correlations

Variable	Spearman Rank Order Correlations Marked/underlined correlations are significant at $p < .05000$					
	IKI	TTC	Irrigation/ Rain	RH	SWC100	TMN
IKI	1.000000	<u>0.738142</u>	<u>-0.449538</u>	-0.015569	<u>-0.315177</u>	0.003800
TTC	<u>0.738142</u>	1.000000	<u>-0.509115</u>	-0.163806	<u>-0.408802</u>	-0.248462
Irrigation/ Rain	<u>-0.449538</u>	<u>-0.509115</u>	1.000000	-0.025127	<u>0.716131</u>	-0.050255
RH	-0.015569	-0.163806	-0.025127	1.000000	<u>0.318710</u>	<u>0.777869</u>
SWC100	<u>-0.315177</u>	<u>-0.408802</u>	<u>0.716131</u>	<u>0.318710</u>	1.000000	0.170241
TMN	0.003800	-0.248462	-0.050255	<u>0.777869</u>	0.170241	1.000000
TMX	<u>0.477322</u>	<u>0.661139</u>	<u>-0.812453</u>	-0.177116	<u>-0.738091</u>	-0.062040
DL	<u>0.432944</u>	<u>0.518985</u>	<u>-0.820829</u>	-0.193649	<u>-0.891799</u>	-0.098216

ADDENDUM

Addendum A: Flowering times of *Saccharum* wild relatives in sugarcane cultivation areas. Sugarcane is known to flower from May to August in South Africa (Sithole and Singels, 2013; Zhou, 2013). Information on flowering months of related species were sourced from Snyman *et al.* (2018). The overlap between the flowering time of *Saccharum* hybrids and each wild relative was calculated by dividing the number of overlapping months with the total number of months of sugarcane flowering (inside bold box)

Species name	Flowering period (Months)												Overlap (Months)
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
<u><i>Saccharum</i> Commercial hybrids</u>					x	x	x	x					
<i>Imperata cylindrica</i>	x	x	x	x	x	x	x	x	x	x	x	x	4/4
<i>Microstegium nudum</i>	x	x	x	x	x	x							2/4
<i>Miscanthus capensis</i>	x	x	x	x	x	x	x		x	x	x	x	3/4
<i>Miscanthus junceus</i>	x	x	x	x	x	x			x	x	x	x	2/4
<i>Trachypogon spicatus</i>	x	x	x	x	x				x	x	x	x	1/4
<i>Cleistachne sorghoides</i>		x	x	x									-
<i>Sorghastrum nudipes</i>	x	x	x	x									-
<i>Sorghastrum stipoides</i>	x	x	x	x	x			x	x	x	x	x	2/4
<i>Sorghum arundinaceum</i>	x	x	x	x	x	x	x	x	x	x	x	x	4/4
<i>Sorghum x drummondii</i>	x	x	x	x	x	x	x				x	x	3/4
<i>Sorghum halepense</i>	x	x	x	x	x	x	x	x	x		x	x	4/4
<i>Sorghum versicolor</i>	x	x	x	x	x							x	1/4