University of São Paulo "Luiz de Queiroz" College of Agriculture

Climatic niches and geographical ranges of sugarcane wild relatives: a comparative and evolutionary approach

Igor Araújo Santos de Carvalho

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba 2020 Igor Araújo Santos de Carvalho Bachelor in Biological Sciences

Climatic niches and geographical ranges of sugarcane wild relatives: a comparative and evolutionary approach

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:

Prof. Dr. GIANCARLO CONDE XAVIER OLI-VEIRA

Dissertation presented to obtain the degree of Master in Science. Area: Genetics and Plant Breeding

Piracicaba 2020

Dados Internacionais de Catalogação na Publicação DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Carvalho, Igor Araújo Santos de

Climatic niches and geographical ranges of sugarcane wild relatives: a comparative and evolutionary approach / Igor Araújo Santos de Carvalho. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2020 .

52 p.

Dissertação (Mestrado) $\,$ – – USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Parentes selvagens da cana-de-açúcar 2. Nichos climáticos 3. Distribuição geográfica I. Título

DEDICATORY

To my parents Isidoria and Francisco and to my sister Camila, who have always supported and encouraged me.

ACKNOWLEDGEMENTS

To the University of São Paulo, the "Luiz de Queiroz" college of Agriculture, the Genetics Department and the Genetics and Plant Breeding graduate program for the opportunity

To the Coordenadoria de Aperfeiçoamento do Pessoal de Nível Superior (Capes) for the scholarship.

To the Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp) for the financial support for this research.

To Dr. Giancarlo Conde Xavier Oliveira for advising me during this degree, supervising this research, for the invaluable lectures on population genetics and evolution and for being an inspirational teacher.

To Dr. Eduardo de Andrade Bressan for the help during several moments of this research, mainly the collection expeditions, and for the precious advices.

To the technician Carlos Alberto Verissimo for the support in the laboratory and the technician Ronaldo Jose Rabello of the Evolution Lab for the help during the collection expeditions.

To my friends I made in the Department of Genetics, you made me feel welcomed and your friendship was essential for me to complete this journey.

To my friends Thassia, Alisson, Brenda and Liz, who have always been there for me despite the distance.

To my parents Francisco and Isidoria and my sister Camila, who have always supported me in my career and have been my biggest inspiration in life.

SUMMARY

Re	sumo			6
Ab	stract	t		7
1	Intro	duction	1	9
	Refe	rences		10
2	Clim	atic nic	hes and geographical ranges of sugarcane wild relatives: a comparative and evolutio-	
	nary	approa	ch	13
	Abst	ract .		13
	2.1	Introd	uction	13
	2.2	Materi	ials and Methods	16
		2.2.1	Study species and occurrence data	16
		2.2.2	Background area delimitation	16
		2.2.3	Environmental data	16
		2.2.4	Ecological Niche Models (ENMs) and range metrics	17
		2.2.5	Niche quantification and measurement of niche overlap	17
		2.2.6	Phylogenetic inference and divergence time estimation	18
		2.2.7	Range and niche evolution	19
	2.3	Result	S	19
		2.3.1	Biotic data	19
		2.3.2	ENMs, range overlap and asymmetry	20
		2.3.3	Climatic niche quantification	20
		2.3.4	Niche comparison	31
		2.3.5	Phylogeny and Divergence dating	35
		2.3.6	Range and Niche Evolution	36
	2.4	Discus	sion	38
	2.5	Conclu	isions	46
	Refe	rences		47

RESUMO

Nichos climáticos e amplitudes geográficas de parentes selvagens de cana-de-açúcar: um enfoque comparativo e evolutivo

A distribuição das espécies é resultado da interação entre seus atributos ecológicos e forças evolutivas, aspectos que são estudados pelas biogeografia ecológica e histórica, respectivamente. Apesar de haver uma separação histórica entre essas duas áreas, a aplicação do conceito de nicho ecológico na modelagem de nichos ecológicos e distribuição de espécies tem permitido uma uma maior integração entre ambas. Os resultados dessa integração têm permitido o preenchimento de diversas lacunas no conhecimento científico sobre a biodiversidade. Essas lacunas são especialmente relevantes para um grupo de plantas conhecido como parentes selvagens de plantas cultivadas (PSPCs), que são um recurso genético valioso com um papel crucial na garantia da segurança alimentar. Nesta dissertação, nós usamos dados de ocorrência e variáveis bioclimáticas para modelar a distribuição geográfica potencial de 15 espécies de parentes selvagens da cana-de-açúcar, assim como para caracterizar seus nichos climáticos em um espaço ambiental bidimensional. Uma filogenia, recentemente publicada, foi reconstruída e utilizada para investigar os mecanismos evolutivos que moldaram a geografia da especiação e a evolução dos nichos climáticos nesse grupo. A modelagem da distribuição potencial das espécies mostrou que os parentes selvagens da cana-de-açúcar são potencialmente distribuídos pelo globo. Adicionalmente, os modelos produzidos podem ser utilizados para informar futuras expedições de coleta, já que os parentes da cana são um dos menos representados em bancos de germoplasma. Espécies mais próximas filogeneticamente exibiram um alto grau de simpatria e assimetria de distribuições, enquanto espécies mais distantes ocorrem em total alopatria. Além disso, foi encontrada uma correlação negativa, e estatisticamente significante, entre o tempo de divergência dos clados e o nível de simpatria, sugerindo que a atual distribuição geográfica dessas espécies pode ter resultado de dispersão de longa distância seguido por eventos de especiação simpátrica recentes. A caracterização dos nichos mostrou que variáveis relacionadas a temperatura mínima e aridez são aquelas que mais contribuem para a variação entre od nichos climáticos das espécies. A maioria das espécies possuem nichos climáticos restritos, enquanto duas apresentam nichos climáticos amplos. Em geral, os nichos climáticos de parentes selvagens da cana-de-açúcar não são idênticos e o nível de similaridade entre eles é o mesmo que o esperado ao acaso, apesar de espécies mais próximas apresentarem os maiores níveis de sobreposição de nicho. Adicionalmente, os modelos evolutivos sugerem que diferentes processos atuaram para moldar as tolerâncias climáticas dessas espécies, a dimensão do nicho climático relacionado à temperatura foi melhor modelada pelo modelo Ornstein-Uhlenbeck com dois ótimos e a dimensão relacionada à aridez foi melhor modelada por um modelo de movimento Browniano. Portanto, nosso estudo mostra a importância de utilizar dados já disponíveis para analisar os atributos ecológicos e aspectos da evolução de PSPCs, que podem ser utilizados para fomentar iniciativas de conservação e para o uso mais eficiente destas espécies no melhoramento de plantas.

Palavras-chave: Parentes Selvagens da Cana-de-açúcar, Nicho Climático, Distribuição Geográfica

ABSTRACT

Climatic niches and geographical ranges of sugarcane wild relatives: a comparative and evolutionary approach

Species distributions are the result of the interplay between ecological requirements and evolutionary mechanisms, which are studied by ecological and historical biogeography, respectively. Although these fields have been historically separate, the concept of ecological niche has allowed their integration. The modelling of ecological niches and species distributions in light of the evolutionary processes has allowed the filling of considerable gaps of scientific knowledge about biodiversity. These gaps are especially relevant for a group of plants known as crop wild relatives (CWR), which are a valuable genetic resource with a crucial role in guaranteeing food security. In this dissertation, we used available occurrence data and bioclimatic variables to model the potential geographic distribution of 15 species of the sugarcane relatives and to characterize their niches in a bidimensional environmental space. We also reconstructed a recently published phylogeny and used it to investigate the predominant mechanism of speciation and the mechanisms shaping niche evolution in the group. The results showed that sugarcane relatives are potentially distributed worldwide and that range models can be used to inform future collection expeditions, considering that the relatives of this crop is one of the most underrepresented in genebanks. Closely related species show a high degree of sympatry with asymmetrical ranges, whilst species from different clades occur in total allopatry. We found a significant negative correlation between age of divergence between clades and pairwise level of sympatry, which suggests that the current geographical distribution may be a result of long-distance dispersion followed by recent events of sympatric speciation. Characterization of niches showed that climatic variables related to minimum temperatures and aridity were the ones that most contributed to variation among climatic niches of theses species. Whilst most species show narrow climatic niches, two show a very broad niche. In general, sugarcane relatives' climatic niches are not identical and are just similar as expected by chance, even though closely related species showed the highest levels of niche overlap. Additionally, models of evolution suggests that different processes acted to shape the climatic tolerances of sugarcane wild relatives, the first dimension of climatic niches related to minimum temperature was best modelled by a Ornstein-Uhlenbeck model with two optima and the dimension related to aridity is best modelled by a Brownian Motion model. Therefore, our study shows the importance of using already available data to analyse the ecological attributes and evolutionary aspects of CWR, which can be used to foment conservation initiatives and a more efficient use of these species in breeding.

Keywords: Sugarcane Wild Relatives, Climatic Niches, Geographical Ranges

1 INTRODUCTION

The geographical distribution of species has been a subject of interest for a long time. For example, the non random distribution of species that leads to clear patterns of biodiversity distribution in the planet has been investigated for animals by Wallace (1860) and for plants by Von Humboldt e Bonpland (2009). The field of biology that is concerned with the spatial distribution of biodiversity is biogeography. Historically, this field can be subdivided in two realms: historical biogeography, concerned with the evolutionary processes that have shaped species distributions; and ecological biogeography, which investigates, amongst others, the current distribution of species and their relationships with the physical and biotic variables of the environment (Lomolino et al., 2006). However, these two areas of biogeography have been put together by recent advances in the concept of ecological niches (Peterson et al., 2011).

The ecological niche is a concept with a long history in ecology and it is a term that has been attributed several meanings (Chase e Leibold, 2003). In the Grinellian definition, the ecological niche is defined by the group of environmental conditions that are needed in order for a species to survive (Grinnell, 1924). Additionally, in his work, Grinnell (1924) points out that species distributions are very marked and rarely a species has a global distribution and that occurs because of the limitations imposed by environmental conditions.

In line with the Grinellian concept of niche, Hutchinson (1957) defined the niche as a hyperdimensional volume with n-dimensions, in which each axis that define this volume is a relevant environmental condition for the survival of the species of interest. In his seminal paper, Hutchinson defines the n-dimensional niche in the context of two competing species. In this context, he subdivides the concept of niche in the fundamental niche and the realized niche. The first is the volume defined by the environmental conditions that influence the survival of populations of a species, whilst the later is a subset of the fundamental niche that encompass the conditions in which a species can survive in the presence of a competitor.

Currently, there are two main approaches to measure the ecological niche of a species: a mechanistic approach and a correlative approach. The first one involves the measurement of the survival ability of individuals in a different set of conditions of a variable of interest by experimentation under controlled conditions, which provides an estimate of the fundamental niche (Kearney e Porter, 2009). Alternatively, correlative approaches measure the niches based on correlations between occurrence data and sets of environmental variables. The digitization of biodiversity data, making occurrences available at data banks easily available, and the availability of worldwide environmental data in fine scale resolution (especially climate data) made the correlative approaches a burgeoning field in ecological research (Elith e Leathwick, 2009).

It is important to highlight that correlative approaches are based on the occurrence records collected within the actual geographical distribution of the species, which is delimited by three requirements:: i) favourable abiotic conditions; ii) biotic interactions, positive and negative, are appropriate; iii) the geographical region has been accessible to the species via dispersion. Therefore, correlative approaches often measure the realized niche rather than the fundamental niche (Soberon e Peterson, 2005).

Correlative approaches for niche estimation can be divided in two main categories: the ones based on direct observations and the ones based in ecological niche models (ENMs). Measurements of ecological niche models based on direct observations are not based on the predictions of a particular model, instead, they often rely on ordination methods (e.g. principal component analysis) to measure the ecological niches in a multivariate environmental space. In order to avoid bias due to sampling effort, they can be improved by the use of smoothed densities of species occurrences and available environmental conditions (Guisan et al., 2014). On the other hand, measurement of niches that rely on model predictions use the observational data as an input for modelling algorithms, which will estimate the ecological niche as a function of a set of environmental variables, the output of the algorithm is projected back on the geographical space in order to predict the areas that are suitable for the species occurrence (the potential distribution) (Peterson et al., 2011).

The theoretical and methodological development in niche modelling by correlative approaches has allowed the emergence of methods to compare niches between species and to test hypotheses about niche evolution, such as the tendency that closely related species tend to be more ecologically similar than expected, phylogenetic niche conservation or PNC (Wiens e Donoghue, 2004), or to be more divergent than expected by chance, known as phylogenetic niche divergence or PND (Pyron et al., 2015). Furthermore, these processes are deeply related to the geographical aspect of speciation, since PNC is expected to arise under allopatric speciation and PND in sympatric and parapatric speciation (Peterson et al., 1999).

In this context, Warren et al. (2008) proposed a method that relied on the comparison of the geographical projection of ENMs to quantify the niche overlap between two species. Additionally, Warren et al. (2008) also introduced the use of null model tests to assess the statistical significance of the similarities and differences among niches. This statistical framework was then adapted to a niche comparison method based on a direct observations approach proposed by Broennimann et al. (2012), which compares ecological niches in a gridded bi-dimensional environmental space defined by a PCA. Simulations with virtual species, show that this approach measures more accurately the overlap between niches than the one based on ENMs (Broennimann et al., 2012). Other methods based on direct observations have been proposed to quantify niche differences between taxa, such as the n-hypervolume (Blonder et al., 2014; Blonder, 2018) and the random translocation and rotation method (RTR) (Nunes e Pearson, 2017). However, the n-hypervolume method does not provide a statistical framework to test the significance of its results and the RTR can only be used to compare species that share a common accessible area.

Hence, methods based on ENMs and direct observations are useful for generating about abiotic tolerances and potential geographical ranges of species. Additionally, methods that compare niches differences and similarities are able to improve our understanding about the processes and mechanisms that generate biodiversity. Knowledge about ecological niches and the underlying evolutionary mechanisms that shape them is especially important for groups that contain economically relevant species, such as the crop wild relatives (CWR). CWRs are a valuable genetic resource to guarantee food security in the future since they are potential donors of alleles associated with important traits for crops, mainly abiotic stress tolerance and disease resistance (McCouch et al., 2013). However, there is an important lack of information about CWRs geographical distribution and potential uses leading to underrepresentation of this species in genebanks and hindering their use in crop breeding programs (Castañeda-Álvarez et al., 2016; Dempewolf et al., 2017).

In this study, we gathered data of occurrence of sugarcane wild relatives available on major biodiversity databases and by field expeditions. This occurrence data were then used to estimate and compare the climatic niche of 15 species related to the sugarcane. Additionally, taking into account a newly published phylogeny of the group, we investigate the geography of speciation and the mechanisms that underlie the evolution of climatic tolerances in this group.

References

Blonder, B. (2018). Hypervolume concepts in niche-and trait-based ecology. *Ecography*, 41(9):1441–1455.

Blonder, B., Lamanna, C., Violle, C., e Enquist, B. J. (2014). The n-dimensional hypervolume. Global Ecology and Biogeography, 23(5):595–609.

- Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N. E., et al. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global ecology and biogeography*, 21(4):481– 497.
- Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., Guarino, L., Harker, R. H., Jarvis, A., Maxted, N., et al. (2016). Global conservation priorities for crop wild relatives. *Nature plants*, 2(4):16022.
- Chase, J. M. e Leibold, M. A. (2003). *Ecological niches: linking classical and contemporary approaches*. University of Chicago Press.
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., e Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Science*, 57(3):1070–1082.
- Elith, J. e Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, 40:677–697.
- Grinnell, J. (1924). Geography and evolution. Ecology, 5(3):225–229.
- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., e Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. *Trends in ecology & evolution*, 29(5):260–269.
- Hutchinson, G. E. (1957). Concluding remarks. In Cold Spring Harbor symposia on quantitative biology, volume 22, pages 415–427. Cold Spring Harbor Laboratory Press.
- Kearney, M. e Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters*, 12(4):334–350.
- Lomolino, M. V., Riddle, B. R., Brown, J. H., e Brown, J. H. (2006). *Biogeography.* Number QH84 L65 2006. Sinauer Associates Sunderland, MA.
- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., Burke, J. M., Charest, D., Cloutier, S., Cole, G., et al. (2013). Agriculture: feeding the future. *Nature*, 499(7456):23.
- Nunes, L. A. e Pearson, R. G. (2017). A null biogeographical test for assessing ecological niche evolution. Journal of Biogeography, 44(6):1331–1343.
- Peterson, A., Soberón, J., e Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. Science, 285(5431):1265–1267.
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., e Araújo, M. B. (2011). *Ecological niches and geographic distributions (MPB-49)*, volume 56. Princeton University Press.
- Pyron, R. A., Costa, G. C., Patten, M. A., e Burbrink, F. T. (2015). Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biological Reviews*, 90(4):1248–1262.
- Soberon, J. e Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas.
- Von Humboldt, A. e Bonpland, A. (2009). Essay on the geography of plants—with a physical tableau of the equinoctial regions (1807). Essay on the Geography of Plants, pages 61–155.
- Wallace, A. R. (1860). On the zoological geography of the malay archipelago. Journal of the Proceedings of the Linnean Society of London. Zoology, 4(16):172–184.

- Warren, D. L., Glor, R. E., e Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution: International Journal of Organic Evolution*, 62(11):2868–2883.
- Wiens, J. J. e Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. Trends in Ecology & Evolution, 19(12):639–644.

2 CLIMATIC NICHES AND GEOGRAPHICAL RANGES OF SUGARCANE WILD RELATIVES: A COMPARATIVE AND EVOLUTIONARY APPROACH

Abstract

Crop wild relatives (CWR) are a potential source of alleles related to abiotic and biotic stress, thus, valuable for plant breeding. Despite their importance, CWR use in plant breeding research programs is hindered by the lack of knowledge of their distribution, abiotic tolerances and the mechanisms underlying their diversification. Sugarcane is one of the most important sources of carbohydrates for human consumption and one of the most relevant bioenergy crops; however, knowledge about their wild relatives is still scarce. In this context, we used occurrence data of 15 sugarcane wild relatives along with 19 bioclimatic variables to model their potential geographic distribution and characterize their climatic niches, in a bidimensional environmental space. We also used a recently published phylogeny to investigate the geography of speciation and the macroevolutionary forces shaping the diversification of climatic niches in this group. Our ecological niche models showed that sugarcane relatives are distributed worldwide and that closely related species tend to occur in sympatry and present asymmetrical ranges and that more distantly related species occur in allopatry. We found that the level of sympatry in this group is negatively correlated with divergence events, suggesting that speciation probably occurred by sympatric mechanisms. Characterization of niches showed that minimum temperatures and aridity are the main climatic variables influencing the abiotic tolerances of these species. Several species close to sugarcane show niche optima in cold climates and are prospective candidates to future research for breeding cold tolerant cultivars. Additionally, we found that most species of sugarcane wild relatives are able to endure a narrow set of climatic conditions with two notable exceptions, S. spontaneum and M. floridulus, which present a very broad niche. Comparison of niches in environmental space showed that although niches are not identical, it is not possible to support neither hypotheses of niche divergence or conservatism. Finally, we found evidence that selection and neutral processes acted on the evolution of the two dimensions of these species' climatic niche. Thus, using an ecological and evolutionary comparative approach, we were able to decrease the extent of knowledge gaps about the geographical distribution, abiotic tolerances and the evolution of species related to sugarcane, contributing to future conservation of genetic resources and breeding efforts.

Keywords: Sugarcane Crop Wild Relatives; Climatic Niches; Geographical Ranges; Niche Evolution.

2.1 Introduction

Yield projections show that, given current increase rates, food production will not attend the demands of the growing human population (Ray et al., 2013). Also, arable land, already under pressure by degradation and urbanization, will diminish in tropical areas due to climate change (Zabel et al., 2014). Also, an increase in the occurrence of extreme climatic events, such as droughts, heat waves and floods is expected (Cai et al., 2014; Cook et al., 2014). Therefore, in order to promote sustainable increase in food production, crop breeding will have to produce not only more productive varieties but also varieties more resilient to biotic and abiotic stress and with increased input efficiency (Godfray et al., 2010).

Although crop breeding has been extremely successful in increasing productivity of agriculture in the last century, the challenges laid out above demand the advent of new approaches to this scientific field (Wallace et al., 2018). In this regard, it is of paramount importance to explore the diversity stored in crop wild relatives (CWR), as well as landraces and undomesticated wild species, since they can be potential donors of alleles associated with important traits for crops, such as abiotic stress tolerance and disease resistance (McCouch et al., 2013). The domestication process of a crop is normally characterized by strong selection and genetic bottlenecks - although the frequency and intensity of those demographic processes varies among different species - which leads to reduced genetic diversity within crops as compared to their wild relatives (Meyer e Purugganan, 2013).

Despite their importance, several biological and technical difficulties hinder the use of CWR in plant breeding (e.g., interspecific crossing barriers, linkage drag and difficulties in gene transfer), however, the most important barrier to the use of CWR is the lack of data on these species (taxonomic, genotypic, potential uses) (Dempewolf et al., 2017). In fact, a recent comprehensive study used bioclimatic variables and occurrence records to model the geographical distribution of the 1,076 wild relatives of 81 crops, showed that 72% of the crop gene pools were poorly represented in genebanks and were classified as high priority for collecting. Additionally, 95% of CWR are underrepresented when considered their full geographical range and their ecological variation (Castañeda-Álvarez et al., 2016).

Therefore, information on CWR climatic niches, geographical distribution and the processes and mechanisms underlying their evolution are crucial to their conservation and effective use in crop breeding. Studies of this nature have been done for CWR of some crops, such as: tomato (Nakazato et al., 2010), sunflower (Kantar et al., 2015), strawberry (Johnson et al., 2014), brinjal eggplant (Syfert et al., 2016) and pigeon pea (Khoury et al., 2015). However, there is a significant lack of information about crops of major economic and social importance such as cassava, maize and sugarcane (Castañeda-Álvarez et al., 2016).

Sugarcane is one of the most economically important crops worldwide in terms of tons harvested and planted area (FAOSTAT, 2017). It is also the third plant source of calories in human consumption, only surpassed by rice and wheat (Moore e Botha, 2013). Modern sugarcane cultivars are hybrids from species of the genus *Saccharum* L., that belongs to the tribe Andropogoneae and subtribe Saccharinae, which comprises other genera (Soreng et al., 2017). Besides *Saccharum*, this subtribe also includes the genus *Miscanthus* Andersson, that contains species used as a source of biofuels and cellulose (Brosse et al., 2012) and as ornamental species (Meyer, 2004).

The delimitation of species and genera related to sugarcane has been historically difficult due to the high frequency of natural hybridization events between sympatric species, different modes of chromosomal transmission in these events and high intraspecific chromosome number variation (Amalraj e Balasundaram, 2006). This high frequency of hybridization led to the creation of the concept of the "Saccharum Complex", a group of interbreeding species that were supposed to be involved in the sugarcane origin and included the genera Saccharum, Narenga Bor, Sclerostachya (Andersson ex Hack.) A. Camus and Erianthus Michx. (Mukherjee, 1957).

In the literature it is possible to identify two main currents of thought about the systematics of the sugarcane relatives: one understands *latu sensu* (*l. s.*) the genera *Saccharum* and *Miscanthus*, in which the first includes all the species of *Erianthus* and the second includes *Narenga*, *Sclerostachya* and the genus *Miscanthidium* Stapf (Clayton et al., 1986; Kellogg, 2015), an hypothesis that has prevailed; and another, a *stricto sensu* (*s.s.*) view that supports the separation between *Saccharum* and *Erianthus*, restricting the former to the six species highly related to the sugarcane (*Saccharum officinarum* L., *Saccharum spontaneum* L., *Saccharum robustum* E.W.Brandes & Jeswiet ex Grass, *Saccharum barberi* Jeswiet, *Saccharum sinense* Roxb. and *Saccharum edule* Hassk.) and proposes the separation of *Miscanthus*, *Narenga* and *Miscanthidium* (Watson et al., 1992).

Molecular systematics approaches have been recently used in order to address these questions. Hodkinson et al. (2002) showed that *Saccharum l. s.* and *Miscanthus l. s.* were not monophyletic, since species of *Erianthus* from the Old World (section *Ripidium*) should be allocated in a different genus *Tripidium* H. Scholz (under *Ripidium*), and that *Miscanthidium* species were placed outside *Miscanthus*. However, the authors did not find enough evidence to support the separation of *Saccharum l. s.* in *Saccharum s. s.,Erianthus* and *Narenga*. A more complete molecular phylogeny of the tribe Andropogoneae, based on the sequencing of five low copy nuclear genes and considering more species, also supported the polyphyly of *Erianthus* and confirmed the existence of *Tripidium* as a clade separated from *Saccharum* and *Miscanthus* (Welker et al., 2015). They found that the *Erianthus* from the New World form a clade with the *Miscanthidium* (represented by *Miscanthus ecklonii* (Nees) Mabb.), and they are related to the *Saccharum s. s.* (represented by *S. officinarum*). At last, they propose that *Narenga* (under *Saccharum narenga* (Nees ex Steud.) Hack.) is a hybrid between *Saccharum s. s.* and *Miscanthidium*.

Evans et al. (2019) published the most complete molecular phylogeny to date of the Andropogoneae, based on whole chloroplast genome and low copy nuclear genes. The authors confirmed previous evidence that the species from the former *Erianthus* sect. *Ripidum* belong to a different genus, *Tripidium*, distantly related to Saccharinae. They proposed the "core Saccharinae" clade, that comprises: *Miscanthus, Miscanthidium*, composed of species that were classified as belonging to *Miscanthus*; New World *Erianthus* and *Saccharum s. s.*.

Despite having made the systematics and taxonomy of sugarcane CWR complex, the high frequency of natural hybridization among these species has possibilitated the extensive use of wild relatives in sugarcane breeding. For instance, modern sugarcane cultivars are the result of crosses between S. officinarum and S. spontaneum, the latter of which was utilized in sugarcane breeding because of its high ratooning capacity and its known adaptation to several environmental stresses (Daniels et al., 1987). It is estimated that approximately 20% of the genome of modern sugarcane cultivars is derived from S. spontaneum (Piperidis et al., 2010).

Further attempts to use wild relatives in sugarcane breeding have been recorded, especially using species from the former genus *Erianthus* from the Old World, in detriment of other groups, such as: *Saccharum arundinaceum* Retz. (Piperidis et al., 2010; Ram et al., 2001), *Saccharum longisetum* (currently *Melinis longiseta* (A.Rich.) Zizka) (Cai et al., 2005), and *Saccharum rufipilium* Steud. (Wang et al., 2009). However, studies have shown that crosses between sugarcane hybrids and *S. arundinaceum* results in a small number of true hybrids with aberrant chromosome behaviour (aneuploidy, chromosome loss and duplication, interspecific recombination) (Wu et al., 2014; Huang et al., 2015). The clarification of the phylogenetic relationship may help us understand the difference of effectiveness in using genetic resources from *Saccharum s.s.* and *Tripidium* in sugarcane breeding, since the latter one is evolutionarily distant.

Therefore, the effective use of these genetic resources effectively demands a thorough understanding of the phylogenetic relationships and the evolution of the sugarcane wild relatives to be able to access their potential as trait donors. In this context, although the results from Castañeda-Álvarez et al. (2016) showcase the important gap in our knowledge about sugarcane CWRs, it is necessary to highlight that, from the eleven species analyzed, four belong to the 'core Saccharinae' (three subspecies of *S. spontaneum* and *S. robustum*) and the remaining seven species belong to *Tripidium*, which are not considered sugarcane wild relatives in light of the most recent phylogeny of the group.

Hence, given the importance of sugarcane and *Miscanthus* as economically relevant crops and the knowledge gap about the ecology and evolution of their wild relatives, this study aimed to: i)to use Ecological Niche Models (ENMs) to model the potential geographical distribution of wild "core Saccharinae" species; ii) to use ordination techniques to characterize their climatic ecological niches and identify species that are potential trait donors to sugarcane and commercial accessions of *Miscanthus*; iii) to compare the climatic niches of the species under study in search of ecological patterns in the group; iv) to use phylogenetic information in conjunction with the ecological data produced in order to understand the geographical mode of speciation of the group and the evolutionary forces driving the diversification of sugarcane wild relatives climatic niches.

16

2.2 Materials and Methods

2.2.1 Study species and occurrence data

We limited our analysis to the group of species closely related to sugarcane, referred to by Evans et al. (2019) as the "core Saccharineae", which includes species from four groups: Miscanthus, Miscanthidium, Saccharum sensu stricto and Erianthus (excluding the Old World species, that should be allocated in Tripidium, outside of the Saccharineae). We searched The Plant List database (www.theplantlist.org; accessed in May of 2019) in order to identify the currently accepted species of *Miscanthus* and *Sac*charum, which currently include species from the four groups mentioned above. Occurrence data for these species were downloaded from GBIF (Global Biodiversity Information Facility, www.gbif.org) and BIEN (Botanical Information and Ecology Network - www.biendata.org). Records from BIEN that were already present in the GBIF dataset were removed. Occurrence data of species native from Brazil (S.angustifolium, S. asperum and S. villosum) were complemented by data collected by the authors. Occurrence records with coordinates were submitted to an automated cleaning proceeding in the R package Coordinate Cleaner (Zizka et al., 2019) in order to remove records that were duplicated, outliers, located in scientific institutions, inside capital and countries centroids, in seas and urban areas. When available, metadata was used to remove records with low precision and records previous to the 1970s, in order to include records compatible with climate measurements, or with no year of collection. Additionally, records outside the known native area of the species were removed, in order to avoid occurrence records of specimens under cultivation or anthropogenic introgressions. In order to have a reliable estimate of their climate niches, we only considered species that had at least 30 occurrence records for subsequent analyses.

2.2.2 Background area delimitation

The estimation of the geographic area that has been available to a species during a relevant period through dispersal is crucial when estimating ecological niches with correlative methods, since it may strongly influence the final results. One of the approaches utilized to do this estimation is to utilize biotic regions occupied by the target species, since their delimitation may be a reflection of dispersal limitations shared by multiple species (Barve et al., 2011). For that reason, we defined the species-specific background areas using the ecorregions delimited by Word Wide Fund For Nature (WWF) (Olson et al., 2001) that were enclosed by the minimum convex polygon that contained all the occurrence records considered for the climatic niche estimation.

2.2.3 Environmental data

We used climatic variables as predictors of the species' abiotic niche. We utilized 19 standard bioclimatic variables available at the CHELSA database (Climatologies at high resolution for the Earth's land surface areas - chelsa-climate.org): bio1-Annual Mean Temperature; bio2 - Mean Diurnal Range; bio3 - Isothermality; bio4- Temperature Seasonality; bio5- Max Temperature of Warmest Month; bio6-Min Temperature of Coldest Month; bio7 - Temperature Annual Range; bio8 - Mean Temperature of Warmest Quarter; bio11 -Mean Temperature of Driest Quarter; bio10 - Mean Temperature of Warmest Quarter; bio11 -Mean Temperature of Coldest Quarter; bio12 - Annual Precipitation; bio13 - Precipitation of Wettest Month; bio14 - Precipitation of Driest Month; bio15 - Precipitation Seasonality; bio16 - Precipitation of Wettest Quarter; bio19 - Precipitation of Driest Quarter. These variables are analogous to the ones available at the WorldClim database (Hijmans et al., 2005), widely used in ecological studies. However, CHELSA data is based on the monthly temperatures and precipitation from 1976 - 2013 obtained

from statistical downscaling of the ERA interim global circulation model, rather than by interpolation of data from climatic stations (Karger et al., 2017). Additionally, we used the aridity index (AI) and the potential evapo-transpiration (PET) from the Global Aridity and PET Database (Trabucco e Zomer, 2009). All climatic variables utilized in this study are georeferenced layers at 30 arc-seconds resolution (approximately 1 km² along the Equator). The values of these variables were extracted for all occurrence points and for random points plotted in the background area of each species.

2.2.4 Ecological Niche Models (ENMs) and range metrics

ENMs of all analyzed species were generated with the software Maxent 3.4.1 (Phillips et al., 2006). The utized parameters were: convergence treshold of $1e^{-5}$, 500 iterations e 10.000 background points. Every model consisted of the average over 10 replicates and it were validated by the bootstrap method. The occurrence dataset was partitioned into a training set (70% of the points) and a testing set (30% of the points). Models were evaluated using the AUC (Area Under the Curve) index. The contribution of climatic variables in the final models were tested by the jackknife method. All the cells with suitability value equal or above the maximum test sensitivity plus specificity cloglog threshold were considered within the range of the species. The species' ranges derived from the ENMs were used to calculate the pairwise range overlap and range asymmetry. The range overlap ranged from 0 to 1 and it was calculated as the area occupied by both species divided by the area of the species divided by the area occupied by the smaller range species (Barraclough e Vogler, 2000).

2.2.5 Niche quantification and measurement of niche overlap

We quantified the climatic niche of the species and measured the niche overlap of every possible pair of species using the framework proposed by Broennimann et al. (2012), which allows the quantification of niche and niche overlap between any two biological "entities" (e.g. sister species, different lineages of the same species) in a bi-dimensional environmental space. Both steps were performed using the R package *ecospat* (Di Cola et al., 2017).

For niche quantification, a Principal Component Analysis (PCA-env) was performed utilizing the values of the 21 environmental variables extracted for each data point (random points and species occurrences). This analysis transforms the original variables in new ones determined by orthogonal synthetic axes (one for each variable) that maximize the variation existent in the data. The first two axes, that explain the most variation in relation to the others, are retained and used to define a bi-dimensional environmental space in which the niches are calibrated. This environmental space is bounded by the minimal and maximal values in both axes and the space is divided into a grid of 100 x 100 cells, in which each cell (v_{ij}) represents a different set of unique environmental conditions.

As occurrence data may be biased, the set of occurrences may not be a good representation of the range of environmental conditions that can be occupied by the species. In order to solve this problem, a kernel density function is applied to determine a 'smoothed' density of occurrences for each cell of the gridded environmental space (o_{ij}) . Analogously, some environmental conditions may be more frequent in the background of a species in comparison to others, especially when comparing different ranges. In order to correct for this "environmental bias", a smoothed density of available environments (e_{ij}) for each species is also calculated.

Finally, the occupancy (z_{ij}) of each cell in the environmental grid is calculated dividing the values of o_{ij} by e_{ij} , and normalizing all values by the maximum value of o/e, in a way that an occupancy value ranging from 0 to 1 is obtained. In order to characterize the species' climatic niche, we obtained measures of niche optima for both axis of the environmental space and a measure of niche breadth. To

obtain these values, 100 samples were obtained from each species niche, the sampling probability was proportional to the occupancy of each species in the environmental space. From these 100 samples we calculated the niche optimum in the two PCs and a measure of niche breadth was obtained by multiplying the variance of the PC1 and variance of PC2. This procedure was performed 1000 times for each species in order to obtain a distribution of variance and niche optima. In order to calculate the niche overlap, the values of z_{ij} are compared between each pair of species to calculate the metric D, as presented by Broennimann et al. (2012) and adapted from Schoener (1970). The statistical significance of the value of D for each species pair was verified by two different tests: the niche equivalency test and the niche similarity test, both adapted by Broennimann et al. (2012) from the tests developed by Warren et al. (2008).

In the niche equivalency test, the occurrence points of both species are pooled together and then randomly allocated in the environmental range of each of the species, maintaining the original number of occurrences for each species. This process was repeated 1000 times and the value of D is calculated in each iteration in order to create a null distribution. The equivalency tests verify if the actual value of D is greater or lower than expected by random (p-value < 0.025 or p-value > 0.975). On the other hand, in the niche similarity test, the measured niche of one species is compared to a niche obtained by randomly sampling the background area of the other species. This test is performed in a reciprocal manner. As in the equivalency test, this process was also repeated a 1000 times, each time the value of D is calculated in order to create null distribution to which the statistical significance of the actual D is tested, considering the same confidence levels described above.

Although D is a very informative metric, the overlap between the niche of two species can be further characterized with the stability, unfilling and expansion indexes, which measure, respectively, the portion of the niche of one species that is shared with the other species of the pair, the portion of the species' niche that is not occupied by the other species and the amount of the environmental space available to the species that is not occupied (Petitpierre et al., 2012; Guisan et al., 2014). Considering that environmental heterogeneity may bias metric of comparison between niches of allopatrically distributed species (Godsoe, 2010; Warren et al., 2014), we only took into account the cells in the environmental space that were present in the background areas of both species.

2.2.6 Phylogenetic inference and divergence time estimation

We reconstructed the phylogeny from Evans et al. (2019) based on low copy nuclear genes. In order to do so, we downloaded the alignment used by the authors available at https://datadryad.org/resource /doi:10.5061/dryad.1k5s048. This alignment contains data for 63 taxa from the tribe Andropogoneae and a total length of 4408 nucleotides from 5 genes: *apo1*, *d8*, *ep2-exon7*, *ep2-exon8* and *rep1*.

We estimated the tree topology and divergence times between clades for both datasets with BEAST 2 (Bouckaert et al., 2014). Each gene was treated as a different partition. We used unlinked site models for each partition, and the bModelTest (Bayesian Model Test) was used as the site model. This functionality is based on a reversible jump MCMC (Monte Carlo Markov Chain) model that allows the chain to "walk" between several models of site substitution without assuming a fixed one (Bouckaert e Drummond, 2017). Additionally, we used the relaxed clock log model and the Yule Calibrated Model as clock and tree models. Both clock and tree models were linked for all dataset partitions. As in Evans et al. (2019), in the low copy nuclear gene phylogeny the root of the tree was calibrated at 19 ± 4 Mya and the divergence of Zea L. was calibrated at 13.8 ± 2 Mya.

The MCMC was composed of 20 million generation sampled at every 1000 generations. The output of the MCMC was analysed in Tracer 1.7 (Rambaut et al., 2018) to check for convergence and to verify if the ESS (Effective Sample Size) of all estimated parameters were superior to 200. The burn-

in was of 25%. Trees on the posterior distribution were summarized with TreeAnotator to generate a maximum clade credibility tree (MCCT) with mean node heights. Each node was dated based on the median of the nodes' ages of all trees in the posterior distribution and the dating precision was based on the 95% highest posterior distribution (HPD) of each node. The final trees were pruned with the R package *phytools* to exclude all the Andropogoneae species that do not belong to the core Saccharinae group.

2.2.7 Range and niche evolution

In order to investigate the main geographical mode of speciation within the Saccharinae, we used the *phyloclim* package (Heibl et al., 2018) to fit an age-range correlation (Fitzpatrick e Turelli, 2006) of the pairwise sympatry and asymmetry indexes on the divergence dates estimated by our molecular phylogeny. The significance of the association between the two variables was accessed by a Mantel test with a 1000 simulations and significance threshold of 5%.

Continuous trait models of evolution were fitted to the niche optima values of the nine species of Saccharineae, for which genetic data was available in the dataset mentioned above, and the trees obtained in the previous step. Initially, we used the *geiger* package v.2.0 (Pennell et al., 2014) in the R platform to fit three different macroevolutionary models to the data considering the existence of a single evolutionary regime in the phylogeny: 1) a simple Brownian Motion model - BM1; 2) An Ornstein-Uhlenbeck model with one optimum - OU; and 3) An Early Burst model - EB.

Subsequently, we used the *OUwie* package (Beaulieu e O'Meara, 2014), in the R platform, to fit alternative and more complex evolutionary models to the data. *OUwie* allows to fit multi-regime and multi-rate models in different parts of the phylogenetic tree. Hence, we used the values of species' niche optima in PC1 and PC2 to separate them in ecological groups (see Results - Climatic niche quantification) and test if they evolved at different regimes. In order to do so, we used six different models of continuous character evolution : 1) A simple Brownian Motion (BM), with a unique sigma rate, equivalent to the first model fitted with geiger, BM1; 2) A BM model that allowed for variable rates between groups, BMS; 3) An Ornstein-Uhlenbeck (OU) model with different optima value for groups, OUM; 4) An OU model with different optima and sigma rates but constant alpha among groups, OUMV; 5) An OU model with different optima and alpha values and unique sigma rate, OUMA; 6) An OU model with different optima, alpha and sigma rates, OUMVA.

For the purpose of accounting for uncertainty in niche optima and phylogenetic estimation, the process of model fitting with both packages was repeated 1000 times using a randomly sampled mean value of each species' niche and randomly sampled trees from the posterior distribution. The choice of the best model was based on the value of the mean Akaike's Information Criteria corrected for small samples (AICc) and the mean AICc weight for all repetitions.

2.3 Results

2.3.1 Biotic data

Our search in The Plant List resulted in 51 accepted species' names for the core Saccharinae, from these, 15 belong to *Miscanthus*, including 3 species formerly placed in *Miscanthidium*, and 36 to *Saccharum*. From this total, only 15 species (29.4% of the total) presented sufficient information for downstream analysis: *M. ecklonii* and *Miscanthus junceus* (Stapf) Pilg. (*Miscanthidium* group); *Saccharum asperum* (Nees) Steud., *Saccharum angustifolium* (Nees) Trin., *Saccharum villosum* Steud., *Saccharum alopecuroides* (L.) Nutt., *Saccharum brevibarbe* (Michx.) Pers., *Saccharum contortum* (Elliott) Nutt. and *Saccharum giganteum* (Walter) Pers. (*Erianthus* group); *S. spontaneum* (*Saccharum* group); and Miscanthus floridulus (Labill.) Warb. ex K.Schum. & Lauterb., Miscanthus sacchariflorus (Maxim.) Hack., Miscanthus sinensis Andersson, Miscanthus oligostachyus Stapf and Miscanthus tinctorius (Steud.) Hack. (Miscanthus group).

Miscanthus changii Y.N.Lee and Miscanthus paniculatus (B.S. Sun) Renvoize & S.L. Chen had no georeferenced records in any of the databases. Miscanthus erectus Gibbs.-Russ. had only one occurrence and Miscanthus depauperatus Merr. had only 3 occurrences. After the cleaning of the occurrence data, Miscanthus fuscus (Roxb.) Benth. had no occurrences left (the newest record was from 1954), the records for the natural hybrid Miscanthus x giganteus J.M.Greef, Deuter ex Hodk., Renvoize were all located in Europe, outside its native range and presumably under cultivation. Miscanthus nepalensis (Trin.) Hack. and Miscanthus violaceus (K.Schum.) Pilg. had only 12 and 10 occurrence records, respectively, and therefore were not considered in subsequent analysis either.

Within Saccharum, S. officinarum and S. edule are domesticated species under cultivation; therefore the environmental conditions in which they are found may not be good indicators of the niche of these species. We did not consider 20 species formerly placed in Erianthus from the Old World, which do not belong to Saccharinae. S. robustum (4 occurrences), S. sinensis (2 occurrences) and Saccharum coarctatum (Fernald) R.D.Webster (11 occurrences) had insufficient number of records for niche estimation.

2.3.2 ENMs, range overlap and asymmetry

The final ENMs for the 15 species considered in this study and the occurrence points of each species can be visualized in Figure 2.1. With the exception of *S. alopecuroides*, all ENMs presented high AUC values (above 0.75, considering the standard deviation) and low omission rates. The values of AUC, the maximum test sensitivity plus specificity cloglog threshold along with their respective omission rates, the three variables that most contributed to the models and their respective percent contribution for all species are summarized in Table 2.1.

The comparison of ranges delimited by the ENMs showed that most species in the Saccharinae clade are allopatrically distributed (range overlap equals to zero), this is especially true for pairs of species from different taxonomic groups, and presumably, less evolutionary related. On the other hand, species pairs from the same taxonomic group present high levels of sympatry (Table 2.2). Furthermore, species that are sympatric at some level tend to present high levels of range asymmetry (Table 2.3).

2.3.3 Climatic niche quantification

The climatic niches of all Saccharinae species projected in the environmental space can be visualized in Figure 2.2. It is possible to notice that the climatic niches breadth varies greatly and that there is a high level of overlap among niches within the clade. Even though there is a noticeable difference in the niche optima of groups of species in both axis of the environmental space, a graphical inspection of niche occupancy shows that niches present a nested arrangement.

In the niche quantification procedure, the first two PCs of the PCA-env, that bound the environmental space, explained 38.8% and 28.86% of the total variance, respectively. The correlation of the original variables to the PCs are presented in Table 2.4. The first principal component has a strong correlation with temperature variables. PC1 has a strong positive correlation (r = 0.955) with the variables bio 06 (minimum temperature of the coldest month) and bio11 (mean temperature of the coldest quarter; r = 0.934) and a strong negative correlation with bio07 (temperature annual range, r = -0.924). Conversely, the second principal component is more influenced by variables related to water availability. PC2 has a strong positive correlation (r = 0.840) with yearly potential evapotranspiration, a strong negative



Figure 2.1: Potential geographical distribution of the "core Saccharinae" based on Ecologial Niche Models (ENMs); A) Worldwide "core Saccharinae" species richness based on the ENMs of 15 species. Maps B to P show the potential distribution for each species studied, black dots represent the occurrence points used to build the ENM, yellow contours represent each species ' background area, heat colors represent the suitability level of the area. B) M. ecklonii; C)M. junceus; D)S. asperum; E) S. angustifolium.



Figure 2.1: Continuation. F) S. villosum; G)S. alopecuroides; H)S. brevibarbe; I) S. contortum;



Figure 2.1: Continuation. L) *M. floridulus*; M) *M. sinensis*; N) *M. sacchariflorus*; O) *M. tinctorius*; P) *M. oligostachyus*



Figure 2.1: Continuation. J) S. giganteum; K) S. spontaneum; L) M. floridulus; M) M. sinensis



Figure 2.1: Continuation. N) M. sacchariflorus; O) M. tinctorius; P) M. oligostachyus

Species	Mean Test AUC	Mean Test AUC standard devia- tion	Maximum test sen- sitivity plus spe- cificity Cloglog th- reshold	Maximum test sen- sitivity plus spe- cificity test omission	Maximum test sen- sitivity plus spe- cificity binomial probabi- lity	Variables that most contri- buted for the model	Variable contri- bution (%)
M. ecklonii	0.8658	0.0326	0.3637	0.1	0	ai bio14 bio17	25.2 20.9 14.4
M. junceus	0.9218	0.0284	0.27	0.0952	0	bio4 bio15 bio11	30.1 9.5 9.2
S. asperum	0.9412	0.0193	0.1962	0.1	0	et bio05 bio09	45.1 15.7 5.4
S. angustifolium	0.9173	0.0164	0.2152	0.0536	0	bio14 et bio17	55.1 10.7 7.4
S. villosum	0.9046	0.0158	0.2763	0.1068	0	bio4 ai bio06	27.3 22.2 13.5
S. alopecuroides	0.7555	0.0534	0.504	0.4167	0.0055	et bio2 ai	$24.7 \\ 14.7 \\ 11.6$
S. brevibarbe	0.921	0.0364	0.1533	0.1286	0	bio2 bio18 bio13	27.4 19 10.4
S. contortum	0.8533	0.041	0.396	0.05	0.0018	bio3 bio6 bio7	22.2 20.4 17.5
S. giganteum	0.8454	0.0229	0.3211	0.1769	0	bio7 bio4 bio8	$ 19.4 \\ 16.1 \\ 12.1 $
S. spontaneum	0.9111	0.0214	0.2602	0.1692	0	bio2 ai bio18	$ 18.2 \\ 15.8 \\ 12.2 $
M. floridulus	0.9039	0.0266	0.2531	0.1733	0	bio2 bio5 bio11	29.1 14.9 13.3
M. sinensis	0.9478	0.0058	0.2209	0.0494	0	bio19 bio10 ai	49.8 13.2 12.6
M. sacchariflorus	0.9442	0.0247	0.0931	0.0437	0	bio17 bio15 bio6	50.7 9.1 8.2
M. tinctorius	0.9142	0.0247	0.3204	0.1348	0	bio19 bio14 bio15	36.9 18.2 10.8
M. oligostachyus	0.872	0.0354	0.2344	0.1625	0	bio3 bio10 bio8	31.8 16.4 9

Table 2.1: Evaluation statistics for the Ecological Niche Models (ENMs) generated for each species of the "core Saccharinae" and top three bioclimatic variables contribution.

· (minda or modern la															
	M. ec-	М.	S. as-	S. an-	S. vil-	S. alo-	S. bre-	S.	S.	S.	M. flo-	M. si-	М.	М.	M. oli-
	klonii	jun-	perum	gusti-	losum	pecu-	vibarbe	con-	gigan-	-uods	ridulus	nensis	sac-	tincto-	gosta-
		ceus		folium		roides		tortum	teum	ta-			chari-	rius	chyus
										neum			florus		
M. ecklonii															
M.~junceus	0.273	1													
$S. \ asperum$	0	0	1												
$S. \ angustifolium$	0	0	0.506	1											
$S. \ villosum$	0	0	0.979	0.967	1										
S. a lope curvides	0	0	0	0	0	1									
$S. \ brevibarbe$	0	0	0	0	0	0.600	1								
$S. \ contortum$	0	0	0	0	0	0.435	0.479	1							
$S. \ giganteum$	0	0	0	0	0	0.519	0.905	0.453	1						
$S. \ spontaneum$	0	0	0	0	0	0	0	0	0	1					
M. floridulus	0	0	0	0	0	0	0	0	0	0.603	1				
$M. \ sinensis$	0	0	0	0	0	0	0	0	0	0.063	0.254	1			
M. sacchariflorus	0	0	0	0	0	0	0	0	0	0	0.164	0.708	1		
$M. \ tinctorius$	0	0	0	0	0	0	0	0	0	0	0	0.950	0.873	, - 1	
M. oligostachyus	0	0	0	0	0	0	0	0	0	0	0.007	0.982	0.627	0.412	1

secies) to 1 (completely	
ompletely allopatric sp	
Values of sympatry varies from 0 (c	
le "core Saccharinae".	
/mpatry index for species of th	
Table 2.2: Pairwise sy	ympatric species).

|--|

abo	Ta
ove	ble :
1 in	2.3:
dice	\mathbf{Pai}
utes	wise
ran	rai
ge a	ıge
syn	asyı
Ime	nme
try.	etry
	for
	spe
	cies
	of S
	accl
	hari
	nae
	that
	000
	ur i
	n sy
	/mp
	atry
	As
	ymr
	netr
	3A A.
	alue
	of 1
	ind
	licat
	es t
	hat
	boti
	h ra
	nge
	3 ha
	ve si
	imil
	ar si
	zes,
	valı
	ues



Figure 2.2: Climatic niches of Saccharinae species projected on a bidimensional environmental space. The axes of the graphs represent the first and second axis of the principal component analysis with the value of all available 19 bioclimatic variables. A) Superimposition of the climatic niche of all 15 species considered in this study, cell colors represent the number of species whose niches includes this set of environmental space for each species, the intensity of the colors in each map represent the value of occupancy in each cell of the grid.



Figure 2.2: Continuation. Figures I to P represent the projection of the climatic niche in the environmental space for each species, the intensity of the colors in each map represent the value of occupancy in each cell of the grid.

Variable name	Variable code	PC1 correlation	PC2 correlation
Annual Mean Temperature	bio1	0.892	0.421
Mean Diurnal Range	bio2	-0.434	0.677
Isothermality	bio3	0.760	0.237
Temperature Seasonality	bio4	-0.893	-0.194
Max. Temperature of Warmest Month	bio5	0.391	0.656
Min. Temperature of Coldest Month	bio6	0.955	0.250
Temperature Annual Range	bio7	-0.924	-0.024
Mean Temperature of Wettest Quarter	bio8	0.420	0.300
Mean Temperature of Driest Quarter	bio9	0.843	0.319
Mean Temperature of Warmest Quarter	bio10	0.507	0.485
Mean Temperature of Coldest Quarter	bio11	0.934	0.323
Annual Precipitation	bio12	0.652	-0.718
Precipitation of Wettest Month	bio13	0.598	-0.486
Precipitation of Driest Month	bio14	0.290	-0.679
Precipitation Seasonality	bio15	-0.074	0.561
Precipitation of Wettest Quarter	bio16	0.610	-0.491
Precipitation of Driest Quarter	bio17	0.296	-0.681
Precipitation of Warmest Quarter	bio18	0.137	-0.704
Precipitation of Coldest Quarter	bio19	0.503	-0.465
Aridity Index	ai_ext	0.387	-0.840
Annual Evapotranspiration	et_yr_ext	0.319	0.810

Table 2.4: Correlation between original bioclimatic variables and the first two principal components defining the bidimensional environmental space in which the niches of Saccharinae species were quantified. The three highest correlation indexes for each principal component are in bold.

correlation (r = -0.810) with the aridity index, followed by a negative correlation (r = -0.718) with bio12 (annual precipitation).

The values of niche optima in PC1 seems to separate species that prevail in high latitudes from species that occur in intermediate and low latitudes. It is possible to notice that species from the *Miscanthidium* group, the North American species of the *Erianthus* group and four species of the *Miscanthus* group are located on the left half of the PC, which means that their niche optima on this dimension are represented by environments characterized by cold temperatures and high annual amplitude in temperature Figure 2.3.

On the other hand, the optima of the species seem to be distributed in the PC2 in a congruent manner to their phylogenetic groups. *Miscanthidium* species are located on the top of the graph, with high values on the PC, which means that their optima are places with low aridity and high values of potential evapotranspiration. *Saccharum spontaneum* and the species of the *Erianthus* group are located in an intermediary position, whilst all five *Miscanthus* species are placed on the bottom of the graph, with low values; thus, their optima are placed on environments subject to high hydric stress (Figure 2.3 B).

The measurements of niche breadth show that two species: *S. spontaneum* and *M. floridulus* have niches much wider than all others species considered in this study. *S. villosum* and *M. sinensis* have narrower niches than the species mentioned above, but bigger than the remaining species, which have narrow niches. Measures of niche optima and breadth could not be obtained for *S. brevibarbe* due to the low number of cells it occupied in the environmental space (Figure 2.3 C).

2.3.4 Niche comparison

The values of D obtained considering all pairwise comparison ranged from 0 to 0.482 with a mean of 0.115. These values are displayed in a heatmap in Figure 2.4 A . It is possible to notice that,



Figure 2.3: Niche optima of fourteen species of species of the subtribe Saccharinae for the first principal component (A), the second principal component (B) of a bidimensional environmental space and niche breadth (C). For each species, the value of niche optimum and breadth is represented by a boxplot that summarizes the distribution of the means of 100 sampling procedures, each consisting of 100 sampled points projected in each species' bidimensional niche; the probability of sampling a value was proportional to the occupancy of the species in the environmental space. The niche breadth measurement consisted of the variance of the 100 samples along the first principal component multiplied by the variance of points along the second principal component.

generally, the highest values of D occur in species pairs that are close phylogenetically and are distributed simpatrically. The only exception to this trend is *M. sacchariflorus* that presents relatively high values of niche overlap with S. villosum and S. angustifolium.

The equivalency test for D shows that niche non-equivalency is the general trend among the Saccharinae species. From the 105 possible comparisons, the equivalency tests showed that the niches of 10 species pairs (9.52%) were significantly more equivalent than expected (p-value < 0.025), whilst the niches of 77 species pairs were less equivalent than expected (p-value < 0.975) (Figure 2.4). Interestingly,

even though the highest values of overlap were more frequent among closely related species, the majority of these values were either non significant or were evidence that the niches were significantly non-equivalent. Among the 10 pairs of equivalent niches, five were made of species belonging to the same clade: M. ecklonii $x \ M.$ junceus (Miscanthidium clade); M. floridulus $x \ M.$ sinensis (Miscanthus clade); S. alopecuroides $x \ S.$ angustifolium, S. alopecuroides $x \ S.$ villosum, S. villosum $x \ S.$ giganteum (Erianthus clade). The remaining equivalent niche pairs were made by species from the Miscanthus clade and the Erianthus clade: M. sacchariflorus $x \ S.$ villosum, M. sacchariflorus $x \ S.$ alopecuroides, M. sacchariflorus $x \ S.$ giganteum, M. sinensis $x \ S.$ alopecuroides, M. sinensis $x \ S.$ contortum.

Furthermore, the 210 reciprocal similarity tests showed that of only 25 (11,9%) species pairs have niches more similar than expected (p-value > 0.975), 22 (10,47%) have niches less similar than expected and the 163 (77,6%) remaining comparison yielded non-significant results (Figure 2.4). It is possible to notice that the niche of *M. sacchariflorus* is similar to almost all species of the *Erianthus* clade, with the exception of *S. contortum*. The remaining species' niches show significant similarity to niches of closely related species. Interestingly, these pairs also show significant divergence, with the exception of the pair *M. sinensis* x *M. floridulus*.

Figures 2.4 C and D show that the results of reciprocal similarity tests are not always concordant (species A similar to species B, but the inverse is not true), except for the pairs S. angustifolium x S. villosum, S. brevibarbe x S. alopecuroides, M. sacchariflorus x S. brevibarbe and M. sinensis x M. floridulus. In addition, S. asperum, S. giganteum and S. brevibarbe have niches more similar than expected to several Miscanthus species. Finally, all the tests that resulted in significant non-similarity between niches involved the species of the Miscanthidium group, which seems to be ecologically divergent from other species in Saccharinae.

Differently from the metrics above, the stability, expansion and unfilling indexes compares the niches considering only the environmental conditions that are shared between the background areas of two species. The values of these metrics are calculated in relation to the niche of one of the species under consideration, so values are not symmetrical for the two species on one pair. Stability, expansion and unfilling values ranged from 0 to 1 with mean 0.512, 0.487 and 0.486, respectively (Figure 2.5).

Within the *Miscanthus* group, the values of stability are generally intermediate to high. Together, with intermediate to low values of expansion and unfilling, it is possible to say that niches are not differentiated in this group. Nonetheless, intermediate to low stability values are found when the niches of other *Miscanthus* species are compared to *M. floridulus*, especially the pair *M. floridulus* x *M. oligostachyus*. Within *Erianthus* high niche stability is found within the North-American species, and *S. giganteum* occupies almost all conditions that are occupied by *S. alopecuroides*, *S. brevivbarbe* and *S. contortum*. The same pattern is found amongst the South American species of this group, which show high to intermediate stability values and *S. villosum* encompasses almost all the conditions occupied by *S. angustifolium* and *S. asperum*. The latter show intermediate overlap of environmental conditions, in which *S. asperum* seems to occupy a bigger range of conditions than *S. angustifolium*. When we compare the Southern and Northern American clades, we can notice that *S. villosum* and *S. asperum* are more differentiated from northern species than *S. angustifolium*.

When compared, both *Miscanthidium* species show an intermediate stability value and relatively high values of expansion and unfilling, which means that although they share environmental conditions, both have unique niche features and do not occupy all available environment. These species also show great niche differentiation when compared to all other species in Saccharinae, which is reflected in low values of stability, and high values of expansion and unfilling.

When we compared species from different phylogenetic groups, it was possible to notice a very distinctive pattern, in which the niches of most species when compared to species with greater values of niche breadth (*S. spontaneum*, *M. floridulus* and *M. sinensis*) show a great value of overlap. However,

34



Figure 2.4: Pairwise climatic niche comparison of 15 species of Saccharinae. A) Pairwise values of overlap D index, the index varies from 0 to 1. Low values of D are represented by blue tones, intermediary values by green tones and high values by yellow tones; B) - D): results of the tests of equivalency and similarity of niches are represented in figures B to D. In all three figures red cells mean that the D value is significantly low, green cells mean that the value is significantly high and grey cells mean that the value is not lower or higher than the expected at random. Similarity tests were performed in a reciprocal manner. Figure C) shows the results of similarity of the species in the y-axis to species in the x-axis and figure D) shows the results of similarity of the species in the x-axis to species in the y-axis.

the reciprocal is not true, broad niches species show intermediate to high values of niche expansion and low values of unfilling when compared to narrow niche species. That means that in shared environmental conditions, almost all the environmental conditions occupied by narrowed niche species are also occupied by broad niche species, whilst the remaining conditions are almost totally filled by the broad niche species. *M. sacchariflorus*, which has a relatively narrow niche, also shows high values of stability when compared to other species, which is also reflected on the D values.

Stability, expansion and unfilling indexes could not be calculated for two species pairs (M.

 $ecklonii \ge M$. tinctorius and M. ecklonii and M. oligostachyus), because these species' background areas do not overlap in the environmental space. Therefore, niche comparison measurements between them should be considered with caution, as they may reflect differences in the environmental conditions of their geographical range rather than actual ecological patterns.



Figure 2.5: Stability index (A), Expansion index (B) and Unfilling index (C) obtained from the comparison of climatic niches of 15 species of Saccharinae. All three indexes vary from 0 to 1, low values represented by blue tones, intermediary values by green tones and high values by yellow tones. White cells outside the diagonal represent species pairs for which the indexes could not be calculated.

2.3.5 Phylogeny and Divergence dating

The topology of the phylogeny reconstructed by us match the one of Evans et al. (2019). However, in our MCCT, *S. asperum* and *S. giganteum* are grouped as a monophyletic clade. The distribution of posterior trees show that there is great disagreement among trees in the relationship of *S. giganteum* and *S. asperum* (Figure 2.6). Our molecular dating approach showed that: the *Miscanthus* group diverged from Saccharum s.l. around 6.213 Mya (95% HPD: 4.269 - 8.282 Mya); S. spontaneum diverged from the remaining groups of Saccharum s.l. 4.837 Mya (95% HPD: 3.226 - 6.588 Mya); Miscanthidium and Erianthus diverged 3.677 Mya (95% HPD: 2.343 - 5.2 Mya); and the Erianthus species from North America diverged from those in South America around 1.94 Mya (95% HPD: 1.124 - 2.828 Mya). Additionally, the divergence between: M. floridulus and M. sinensis occurred 2.811 Mya (95% HPD: 1.633 - 4.131 Mya), M. ecklonii and M. junceus at 1.42 Mya (95% HPD: 0.5425 - 2.485 Mya) and S. angustifolium and S. villosum around 1.522 Mya (95% HPD: 0.814 - 2.279 Mya). It is important to notice that the 95% HPD range is very large for older nodes, which indicates that these estimations are not precise. On the other hand, the more recent divergence events, the ones that originated the current species, show more precise estimates (Figure 2.7).



Figure 2.6: Phylogenetic tree of 9 species of Saccharinae. Each tree represents one of the fifteen thousand trees obtained in the posterior distribution of the Bayesian algorithm BEAST. Blue colors represent agreement among trees, whilst green and red colors represent disagreement on the topology of trees. The horizontal axis represents the age of the nodes in million years.

2.3.6 Range and Niche Evolution

The ARC resulted in a significant negative correlation between age of divergence events and sympatry levels (r = -0.1131; p-value: 0.004). This means that sympatric species tend to be more closely related; hence, it suggests that sympatric speciation may be the primary mode of geographical speciation in the Saccharinae (Figure 2.8 A). Although we found a negative correlation between age and asymmetry of ranges, this relationship was not statistically significant (r = -0.2358; p-value = 0.342) (Figure 2.8 B).

In relation to the models of climatic niches evolution, when niche optima from PC1 (more influenced by minimum temperature and temperature seasonality) were fitted to the posterior trees using



Figure 2.7: Dated maximum clade credibility tree (MCCT) of 9 species of the "core Saccharinae clade" along a geological scale. The distribution curves on the internal nodes of the trees represent the distribution of the median node ages of all trees on the posterior distribution of the BEAST analysis. Mi. = Miocene, Pl. = Ploiocene, P. = Pleistocene

the *geiger* package, the following median values of AICc were obtained considering all the 1000 model fitting steps: 33.83 for BM, 38.37 for OU and 38.62 for EB (Figure 2.9 A-D)

After preliminary fitting tests of the *OUwie* package, the more complex models (OUMA and OUMVA) resulted in the lowest AICc values. However, the estimated parameter values were often numerically nonsensical and therefore these models were discarded from subsequent analysis. We hypothesize that this result was due to the relatively low number of taxa, that may not provide enough data to estimate well the several parameters of these models. Therefore, we only analyzed the results of the first four models (BM2, BMS, OUM, OUMV). Among these, the OUM model presented the lowest median AICc value considering all the 1000 steps in model fitting (median AICc = 30.72), followed by the BM2 model (median AICc = 33.92) and the models BMS (median AICc = 41.4) and OUMV (median AICc = 41.98) (Figure 2.10 A-F)

It is interesting to notice that the median AICc of the BM1 model (OUwie) and the BM model (geiger) are very close, what is expected since these models are equivalent. The mean AICc weights



Figure 2.8: Age-range correlation (ARC) analysis in Saccharinae. (A) Correlation between common ancestor ages in million years (Mya) of Saccharinae species and mean sympatry level (r = -0.1131; p-value = 0.004). (B) Correlation between common ancestor ages in million years (Mya) of Saccharinae species and mean asymmetry level (r = -0.2358; p-value = 0.342).

of the four models were: 62.71% (OUM), 30.48% (BM1) , 5.64% (BMS) and 1.15% (OUMV). Hence, considering the totality of models, OUM presents the best fit to the data. However, the difference of the AICc values of the OUM model and the BM and BM2 models is only of 3.11 and 3.2, what does not allow us to discard the latter models. The median values for alpha and sigma were 3.06 and 1,33, respectively. The two different regimes showed different optima along the PC1 axis, the distribution median was -0.552 for the group adapted to colder climates and 1,786 for warmer climates.

Considering the fitting of the models to the niche optima in PC2 with geiger, the model BM (AICc = 39,52) had the lowest AICc value, followed by OUM (AICc = 44,3) and EB (AICc = 44,25) (Figure 2.9 E-H). The models fitted with OUwie had the following AICcs: BM2 - 39,55014; BMS - 79,81; OUM - 50.24; and OUMS - 145.56. Again, the AICc values of BM and BM2 are almost the same (Figure 2.10 F and G).

Differently from PC1, the models fitted for the data on PC2 showed that BM/BM1 is the best model among the ones considered. The median of sigma square estimate for this model was 0.684 in the *geiger* package and 0.691 in *OUwie*, whilst for both packages the median of optimum was approximately -1.45.

2.4 Discussion

Biodiversity is inherently complex; therefore, to characterize and to measure all of its aspects is a huge challenge that needs to be addressed by ecologists and evolutionary biologists. In order to measure the size of our ignorance about the biodiversity, seven main knowledge gaps have been identified by Hortal et al. (2015): i) the Linnean shortfall, the lack of knowledge of species description and cataloguing; ii) the Wallacean shortfall, in respect to the incomplete knowledge about species' geographical distribution; iii) Darwinian shortfall, deficient knowledge on the phylogenetic relationships, and on species and trait evolution; iv) the Hutchinsonian shortfall, the lack of knowledge on physiological tolerances and responses to abiotic conditions; v) the Prestonian shortfall, the insufficient data on abundance and the spatial and temporal dimensions of populations' dynamics; vi) the Raunkiæran shortfall, in respect to knowledge about species functions and morphological traits; vii) the Eltonian shortfall, lack of knowledge on species' interactions and its impacts on individual fitness.



Figure 2.9: Fitting of niche evolution models in two dimensions of the climatic niche of Saccharinae species with the *geiger* package in R. Fitting of the models were performed in a 1000 simulations with sampled phylogenetic trees from a Bayesian posterior distribution and niche optimum values sampled for the climatic niche of each species. Graphs A) to D) show the results of model fitting using niche optima in the first dimension of the climatic niche and graphs E) to H) for the second dimension of the niche. A) and F) are box plot charts showing the AIC values obtained for the Brownian Motion (BM), Ornstein-Uhlenbeck (OU) and Early Burst (EB) models. B) and F show the mean AIC weight over all simulation for each model. C) and G) show the distributions of the values of the sigma parameter in the BM model. D) and H) show the distribution of the values of the optimum parameter for the clade in the BM model. The dashed red line shows the median value for the estimated parameters.

In that context, even though the "core Saccharinae" contain cultivated species of high economic relevance, such as sugarcane and *Miscanthus* species, scientific knowledge for this group is still deficitary in almost all aspects mentioned above. In this study, we were able to gather data on species occurrence on the field and on databases. Based on them and available climatic data our results can partially bridge three of these shortfalls: the Wallacean, the Hutchinsonian and the Darwinian.

In regard to attempts to respond to the Wallacean shortfall for the Saccharinae, the oldest attempt to map the geographic distribution of the wild relatives of sugarcane was made by Mukherjee (1957) that described the distributional area of *Saccharum s.s.*, *Erianthus* (including *Tripidium* species), *Narenga* and *Sclerostachya*. The ranges were defined arbitrarily and determined by records in only



Figure 2.10: Fitting of niche evolution models in two dimensions of the climatic niche of Saccharinae species with the *OUwie* package in R. Fitting of the models were performed in a 1000 simulations with sampled phylogenetic trees from a Bayesian posterior distribution and niche optimum values sampled for the climatic niche of each species. Graphs A) to E) show the results of model fitting using niche optima in the first dimension of the climatic niche and graphs E) to I) for the second dimension of the niche. A) and F) are box plot charts showing the AIC values obtained for the Brownian Motion model with a single rate (BM1), Brownian Motion model with two rates (BM2), Ornstein-Uhlenbeck model with two optima (OUM) and Ornstein-Uhlenbeck model with two optima and two alpha parameters (OUMV). B) and G) show the mean AIC weight over all simulation for each model. C) and D) shows the distribution of the two optima parameters in the OUM model. H) and I) show the distribution of sigma values and optimum parameter for the BM1 model. The dashed red line shows the median value for the estimated parameters.

two herbaria. In this study, Mukherjee (1957) recognized some patterns, such as: the endemism of the American species of *Erianthus*, the widespread distribution of *S. spontaneum* and the Indo-Burma-Chinese Frontier as the place of origin and diversity center of the group. Recently, another attempt to model the geographic distribution of sugarcane wild relatives was present in Castañeda-Álvarez et al. (2016). However, both studies were based on an obsolete definition of the *Saccharum* relatives and, therefore, should be revisited.

Our systematic search in the GBIF and BIEN databases for occurrence data of the "core

Saccharinae" along with ENMs produced by the Maxent algorithm, allowed us to present models of potential distribution of fifteen species of sugarcane wild relatives. To our knowledge this is the most complete attempt to map the distribution of the sugarcane relatives. This information is essential to update the indexes of representativeness of sugarcane in gene banks and to guide collections of germplasm.

However, our results also showed that data of occurrence, with enough quality, is only available for less than one third of all recognized species (considering the current delimitation of *Saccharum* and *Miscanthus*) and approximately one half, when we exclude *Tripidium* species. This number shows that the Wallacean gap is still significant within the "core Saccharinae". The main groups affected are *Miscanthus*, *Miscanthidium* and species and cultivars of *Saccharum* s.s.

The Hutchinsonian shortfall is particularly important to CWR, since species that exist in extreme environments are good candidates to possess useful characteristics to crop breeding (Zhang et al., 2017). Filling this knowledge gap is especially important for sugarcane relatives, since sugarcane depends on rich soil, high water availability and warm temperatures to obtain optimal growth (Ebrahim et al., 1998). These requirements limit the production of sugarcane to tropical and subtropical regions; therefore, it is of great interest of breeders to produce varieties that are resilient to hydric stress and colder temperatures (Wu et al., 2014).

In that regard, our analysis shows that low temperatures and water availability are the most relevant climatic conditions determining the niches of the "core Saccharinae" species considered. Although sugarcane is mainly a tropical crop, most species analyzed have their niche optimum in climates characterized by cold temperatures with high annual amplitude. This is specially true to *Miscanthus* species, which have almost all of their niche optima in environments with cold winters (except *M. floridulus*) and all species have their optima in environments associated with hydric stress.

Miscanthus species have been cultivated as a bioenergy crop, especially the interspecific hybrid *M.* x giganteus, triploid hybrid between *M. sinensis* and *M. sacchariflorus* (Sacks et al., 2013). Therefore, the ability of surviving cold stress has been investigated in the wild germplasm of these two species. The range of *M. sacchariflorus* extends as far north as eastern Siberia, a region characterized by severe winters (Clark et al., 2016). Consistently with our results for niche optima, field experiments have shown that Siberian accessions of *M. sacchariflorus* have an outstanding tolerance to chilling (Pignon et al., 2019).

Furthermore, the capacity of surviving through winters of 564 accessions of M. sinensis was evaluated by field trials placed in three locations in East Asia and two locations in the United States. In this experiment, Dong et al. (2019) showed that there is a great difference in overwintering ability among genetic groups in M. sinensis and that the germplasm from East China, Southern and Northern Japan are more resistant to cold, in comparison to other groups from lower latitudes. This result is in line with our results for niche optima and the intermediate niche breadth value for M. sinensis. Dong et al. (2019) also found a significant relationship between bioclimatic variables (bio1, bio6, bio10 and bio11, from Worldclim) extracted from the collection sites and the capacity of resisting winter temperatures, whereas our results showed that bio06 and bio11 (from CHELSA) have a strong correlation with the first axis of the environmental space defining the climatic niches of the "core Saccharinae".

In addition, genome-wide analysis performed by Dong et al. (2019) showed 73 SNPs associated with overwintering ability in *M. sinensis*. This SNPs are candidate genes that can be behind the molecular mechanism of cold tolerance, among them: the MYB gene, COR47, CEX13, WRKY12 and CSPD1. They also identified several members of gene families involved with cold tolerance which were near or in genomic regions associated with overwintering, which may be evidence that genome duplications can be an important evolutionary mechanism behind this trait. This results provides insightful evidence about the molecular mechanisms that might be behind abiotic tolerance in *Miscanthus* analyzed in this paper, and that can be used in order to search for orthologs, possibilitating a better understanding of the ecology of the remaining Saccharinae.

Our results show that *M. oligostachyus* and *M. tinctorius* have niche optima equivalent to those of the other widely researched *Miscanthus* species, such as *M. sinensis* and *M. sacchariflorus*, or even colder. Therefore, collection and study of the germplasm of these species can be fruitful for *Miscanthus* breeding for cold tolerance.

In that context, chilling tolerance has been transferred to a certain extent to sugarcane commercial hybrids in *Miscanthus* x *Saccharum* hybrids, also known as "miscanes" (Kar et al., 2019). However, our results also show that species from the *Miscanthidium* group (*M. ecklonii* and *M. junceus*) and North American species of the *Erianthus* group (*S. contortum*, *S. alopecuroides* and *S. giganteum*) also have their niche optima in relatively cold climates, equivalent to the ones in *Miscanthus*. Because these species are phylogenetically closer to sugarcane, they could be used as an alternative source of alleles to transfer cold tolerance to commercial hybrids. This shows that the use of available biodiversity and environmental data combined with ecological theory can be a useful tool to open new research lines in plant breeding.

The recent application of molecular methods has helped the clarification of the phylogenetic relationships among sugarcane relatives (Hodkinson et al., 2002; Welker et al., 2015; Evans et al., 2019) and thus helping to minimize the Darwinian shortfall for these species. In this paper, we used the recently available data from Evans et al. (2019) to reconstruct the phylogenetic relationships of 9 species belonging to the "core Saccharinae" and to estimate the divergence date among clades and species. Despite general agreement between tree topologies, there are still some gaps that need to be addressed in order to obtain a better understanding of the evolution of sugarcane wild relatives.

In this context, the uncertainty in the trees of our analyses' posterior distribution about the relationship between *S. asperum* and *S. giganteum*, is an important open question. Although they belong to the same clade (referred to as the *Erianthus* group in this paper), it is still unclear how the *Erianthus* species from North America and South America diverged. In other phylogenies the relationship between these groups is represented by a polytomy (Welker et al., 2015; Evans et al., 2019). Therefore, it is important that genetic information about other North American species is included in new phylogenies of the group. Beyond the *Erianthus* group, there is still lack of representativeness throughout the clades of Saccharinae, only 9 species out of 31 species are represented in the most complete phylogeny of the group to date. This information is necessary to clarify the boundaries between the groups within Saccharinae, especially between *Miscanthus* and *Miscanthidium*.

Furthermore, our molecular dating approach resulted in more ancient estimates for the nodes in the phylogeny than the ones proposed by Evans et al. (2019). For instance, we estimated that the divergence of *Miscanthus* from *Saccharum s.l.* occurred around 6.213 Mya (95% HPD: 4.269 - 8.282 Mya) and the divergence between *Miscanthidium* and *Erianthus* occured around 3.677 Mya (95% HPD: 2.343 - 5.2 Mya), whilst the aforementioned work dated this divergence events at 3.6 Mya and 2.3 Mya, respectively. Our estimates are in line with the ones found by Estep et al. (2014), who dated the divergence of the parental genome of *Miscanthus* at 7.4 Mya (95% HPD: 3.7- 9.4 Mya) and the divergence of *Miscanthidium* (represented by *M. ecklonii*) at 3.8 Mya (95% HPD: 2.0 - 6.2 Mya). Nonetheless, the estimates of Evans et al. (2019) are in line with the ones found by Evans et al. (2017) and Kim et al. (2014). It is important to highlight that our age estimates of ancient nodes showed a very broad 95% HPD, which indicates a lack of precision. Differently, our age estimates for more recent speciation events are more precise and do not differ significantly from the ones of previous molecular dating attempts.

Despite the need for improvements, current phylogenetic data makes it possible to analyze biogeographical and climatic niche information about the Saccharinae in a comparative way, allowing us to dissect the geographical and ecological mechanisms underlying the diversification of this group. In that respect, the availability of dated molecular phylogenies has brought the possibility to investigate the geography of speciation in a macroevolutionary approach by analysing the patterns of range of overlap through a phylogeny. The premise for this kind of analysis is that the different mechanisms of speciation (allopatric, sympatric, parapatric and peripatric) leave imprints that determine the pattern of geographical distribution within a clade. For example, allopatric speciation is thought to result in closely related species with disjunct distribution (no range overlap), whilst sympatric processes are thought to result in closely related species which overlap each other (Barraclough e Vogler, 2000). The most utilized quantitative method to investigate this processes is the age-range correlation (ARC), in which the average range overlap of each node of a phylogenetic tree is regressed on the age of each node: allopatric processes are thought to result in a positive correlation between node age and average node sympatry, whilst sympatric processes are thought to result in a negative correlation (Fitzpatrick e Turelli, 2006).

In this respect, several studies that utilize ARC often are inconclusive about the predominant mode of geographical speciation in the clades analyzed. Criticism about this approach is based on the idea that species present dynamical ranges and that most demographic processes that occur after the speciation events can affect species distribution to the extent that current geographical distribution does not carry an imprint of the speciation process. For instance, current overlapping species may have arisen by allopatric speciation followed by secondary contact in the same manner that species that arose sympatrically could have dispersed to different regions resulting in current allopatric distribution (Pigot e Tobias, 2015). However, an approach based on approximate Bayesian simulation showed that summary statistics of current ranges, such as range overlap and asymmetry, are able to correctly predict the underlying process of geographical speciation in simulated species (Skeels e Cardillo, 2019).

The ARC using our phylogeny and the range overlaps obtained by the comparison of ENMs showed a significant negative correlation, which is evidence that sympatric speciation is the predominant mechanism within the core Saccharinae. Extrapolating from the phylogeny allow us to notice that the level of sympatry is zero (or almost zero) among almost all species pairs of different groups (*Miscanthidium*, *Erianthus*, *Saccharum s.s.* and *Miscanthus*), that can be a proxy for phylogenetic relationship. Additionally, there is no range overlap between the South American and North American *Erianthus*. On the other hand, the levels of sympatry are high for almost all species pairs within taxonomic groups, reaching almost complete sympatry for some pairs. Within sympatric species pairs, ranges are often asymmetrical and exceedingly high levels of asymmetry are found in almost all taxonomic groups, *Miscanthidium* being the exception.

Macroevolutionary evidence of the relevance of sympatric speciation in plants has been obtained by recent meta-analysis studies. Using an Approximate Bayesian Computation framework to compare the support of different models of geographic speciation in six plant genera, Skeels e Cardillo (2019) found evidence that the sympatric model is supported for three genera (*Protea* L., *Hakea* Schrad. & J.C.Wendl., *Bursera* Jacq. ex L.), whilst parapatric was supported for *Sidalcea* A. Gray ex Benth. and models composed of more than one mechanism of speciation were supported for *Banksia* L.f. and *Mimulus* L..

In a more traditional approach, Anacker e Strauss (2014) compared the range overlap and range asymmetry between 71 sister-species pairs of plants from the California Floristic Province (CFP), they found that 80% of sister-species pairs were sympatric and that ranges were highly asymmetrical, on average the range of the broad-range species was 10-fold bigger than the smaller-range species, which resembles the pattern we have found for the Saccharinae.

Furthermore, Anacker e Strauss (2014) suggest that vicariant speciation is less important in plants than in animals and that sympatric processes may more ubiquitous in plants than once thought. They argue that a group of processes known as budding-speciation in which a progenitor species that is more widely distributed gives rise to a new species with a more restricted distribution leading to overlapping ranges with a high degree of asymmetry.

Therefore, we believe that the evidence for sympatric speciation processes, such as budding

speciation, as a primary mechanism for the origin of species in Saccharinae along with the disjunct distribution of major clades within this subtribe may indicate the biogeography of this group was characterized by ancient long distance dispersions followed by more recent sympatric diversification. This biogeographical hypothesis is in line with evidence that species in the tribe Andropogoneae suffered a recent burst of diversification in the late Miocene, in which extant species are a product of recent allopolyploidization events (Estep et al., 2014).

Biogeography has great influence on the ecological aspects of speciation (Peterson et al., 1999), thus, understanding the processes underlying the evolution of the ecological requirements (niche optima and niche breadth) of the Saccharinae is also essential to understand the macroevolutionary patterns of this group. In this context, the measurement of the range of conditions that define the niche of a species, niche breadth, is an essential concept to investigate biological adaptation, speciation and niche evolution (Sexton et al., 2017). The measurements of climatic niche breadth showed that there is great variation in the amplitude of abiotic conditions tolerated by the species within the "core Saccharinae. S. spontaneum and M. floridulus have a much larger niche breadth than all other species analysed, followed by M. sinensis and S. villosum and all the remaining species are characterized by narrower niches. It seems that most species in Saccharinae are climatic specialists and that the condition of climatic generalists arose twice and independently in Miscanthus and Saccharum.

In that regard, the recent publication of the genome of a haploid lineage of *S. spontaneum* along with the resequencing of 64 individuals from different geographical locations by Zhang et al. (2018) may help to shed light on the causes of this species' ecological diversity. *S. spontaneum* is historically known for its high abiotic stress tolerance and disease resistance, which motivated the introgression of its genes in modern hybrids sugarcane cultivars (Daniels et al., 1987). *S. spontaneum* suffered a reduction in the basic number of chromosomes, when compared to its ancestor, caused by fission of two ancestor chromosomes and fusion with other chromosomes. Additionally, this genome of the species went through two independent rounds of whole genome duplication, most likely due to autopolyploidization (Zhang et al., 2018).

Furthermore, diversity analysis showed that nucleotide diversity in *S. spontaneum* is low throughout most of its genome, however this is not true to rearranged regions, which have considerable amounts of genetic diversity. Functional analysis of these diverse regions showed that they are enriched with genes associated with stress, the NBS-encoding genes. Also, these regions are thought to be under balancing selection, which maintain allelic diversity and therefore may have helped *S. spontaneum* to use a wide set of environments (Zhang et al., 2018). Thus, comparative genomic studies with other Saccharinae species may help us to understand the basis of niche breadth variation in this group.

Niche evolution has been a controversial topic in ecological science. The predominant hypothesis is that phylogenetic niche conservatism (PNC) is a widespread pattern in clades, and it predicts that that ecological niches should evolve in a very slow rate in a way that closely related species would share very similar ecological requirements and that this similarity increases with phylogenetic relatedness (Wiens e Donoghue, 2004). On the other hand, there has been evidence that closely related species should show more divergence than expected by chance (Pyron et al., 2015).

Among the Saccharinae species we investigated, we found that pairwise D values were low on average. Nonetheless, the highest values of D were found in species that belong to the same phylogenetic group and among groups pairs show D values next to zero. In order to test the significance of these values, we performed the equivalency and similarity tests, which test whether one niche is identical to the other and whether the niche of one species can predict the niche of other species better than random points, respectively. Our results show that, in general, the niches of the sugarcane relatives are not identical, however, they are neither more similar or divergent that the expected by random. This pattern can be explained, in part, by the variation in niche breadth and their arrangement within environmental space, in which broader niches enclose smaller niches. The D value, as adapted by Broennimann et al. (2012), compares the occupancy of both species in each cell of the environmental grid. Thus, when a bigger niche encloses a smaller one the difference between occupancy values of the two species is high for the majority of cells in the environmental grid, lowering the value of D. In comparison, the D values obtained under the null hypothesis are considerably higher, leading to the rejection of the null hypothesis in favor of the alternative that niches are significantly non-equivalent.

These tests are influenced both by niche optima (mean) and niche breadth (variance); therefore, if two species have identical means and different niche variances, the test result will be that the niches are similar although not identical (Glennon et al., 2014). Therefore, although these metrics have been widely used in comparison among niches, caution is necessary when using them to identify ecological patterns. Rather than simple ecological divergence, our data is more likely to have arisen due to differences in magnitude of niche breadths, especially within clades.

Beyond comparing current niche differences, the increasing availability of molecular phylogenies has allowed the PNC hypothesis to be investigated, in several clades, mainly by two methods: a) the investigation of phylogenetic signal (PS), a linear relationship between evolutionary distance among species and their ecological similarity; b) Testing the fit of *a priori* evolutionary models of trait evolution to quantitative measurements of niche characteristics. In this approach, the fit to specific models could be seen as evidence to the macroevolutionary forces acting on niche evolution (Münkemüller et al., 2015). Although widely used, evidence from PS to test for PNC has been recently criticized because it can result in ambiguous results (Münkemüller et al., 2015).

The fitting of evolutionary models to the two dimensions of the environmental space in which we characterized the climatic niche of Saccharinae species resulted in evidence of different processes acting in the evolution of climatic niches for this group. In the first dimension, more associated to temperature variables, the macroevolutionary model that presented the best fit was a Ornstein-Uhlenbeck (OU) model with two adaptive peaks. OU models are used to model the evolution of traits under selection processes, in which trait values of different species are attracted to one or more adaptative optima (Butler e King, 2004). In Saccharinae, the two optima represented a peak at low temperature conditions and another one at high temperatures. This result, combined with the patterns from modelled geographical distribution, shows that Saccharinae species that occupy high latitudes are likely under selection pressure to adapt to cold climates, whilst species that have a tropical or subtropical distribution are under pressure to endure warm climates.

The existence of a selection force driving adaptation to cold climates in some species of Saccharinae seems paradoxical considering that this subtribe Saccharinae belongs to the tribe Andropogoneae (Kellogg, 2015), which is one of the few clades of grasses that present C4 photosynthesis (Group, 2012). Physiological models predict that C4 photosynthesis is only advantageous over the more widespread C3 photosynthesis in warm climates, in which the rate of photorespiration is higher (Sage, 2004). Indeed, C4 grasses are known to dominate open grasslands located in tropical regions, whilst C3 grasses dominate grasslands in higher latitudes (Griffith et al., 2015). Additionally, comparative phylogenetic studies show that C4 photosynthesis in grasses arose in warm climates with low concentration of CO_2 (Edwards et al., 2010).

However, recent studies have shown that there are no physiological barriers to C4 grasses to occupy cold climates (Long e Spence, 2013; Atkinson et al., 2016). Furthermore, a large comparative analysis comprehending all species of Poaceae has shown that C4 grasses were able to increase their temperature niche breadth, accelerating the migration of these grasses between warm and cold climates and possibilitating the occupation and diversification of theses clades in cold habitats (Watcharamongkol et al., 2018). Besides macroevolutionary patterns, phylogenetic and demographic analysis together with evidence from vegetation dynamics in relation to past climate change showed that *M. sinensis* radiated

from Southern China, from where there was an early migration that colonized Japan. After the glacial maximum, as the climate was getting warmer, *M. sinensis* started migrating towards higher latitudes in China and Korea (Clark et al., 2015).

Despite the evidence that selection is the force behind the adaptation for colder climates within the groups of Saccharinae that were able to expand to higher latitudes, it is important to take into consideration other processes that could generate similar patterns. The role of biogeographical processes in shaping niche similarities has been investigated. Recent studies have shown that niche similarities can arise under neutral processes of evolution (in the absence of any selection forces) just because of the role of biogeographical processes. Simulations of climatic niche evolution under neutral community dynamics showed that high speciation rates resulted in more frequent co-occurrence between closely related species in the geographical space, which shows high auto-correlation, leading to more similarity between these species in the environmental space (Coelho et al., 2019). Therefore, considering the evidence we found for the prevalence of recent sympatric speciation within the group of "core Saccharinae", we can not discard the possibility that these mechanism could have an influence on the similarities in species' niche optima related to temperature.

On the other hand, the evolution of niche optima PC2, more related to aridity index and evapotranspiration, was unequivocally best modelled by a Brownian-Motion-like model. The BM model is a random walk model that can be seen as a special case of the OU model where the strongness of selection is zero (Münkemüller et al., 2015). BM models are normally associated with drift, but it has been shown that BM patterns can be obtained by models where selection optima change very quickly (Hansen, 1996). In BM models, we expect that closely related species show a similar niche and that when phylogenetic distance increases, the ecological differentiation also increases. In fact, it is clear that there is a relation between phylogenetic groups and niche similarity for the species analyzed in the PC2. However when we consider the Saccharinae as a group, there is no evidence that selection forces acted to stabilize the climatic niches towards an optimum related to hydric stress. On the contrary, the variation of this trait in Saccharinae seem to be only a secondary product of the diversification of the group.

2.5 Conclusions

In this study, we were able to use ENMs to predict the potential geographical distribution of 15 species of sugarcane wild relatives. Additionally, ordination techniques were used to characterize and compare the climatic niches of these species. We found that minimum temperature and aridity are the main climatic variables that influence the ecology of sugarcane wild relatives. Also, most sugarcane wild relatives present their niche optimum in cold climates, meaning that these species are a valuable genetic resource for breeding cold tolerance to commercial cultivars of sugarcane and *Miscanthus*. The range of climatic conditions presented great variation within clades and most species studied are restricted to a narrower set of environmental conditions. Two species (M. floridulus and S. spontaneum) are climate generalists and are able to occur in broader range of climates. The differences in niche breadth and the arrangement of climatic niches on the environmental space resulted in a pattern of non-identical niches. However, there was not enough evidence to support either niche similarity or divergence.

We used available genetic data to reconstruct a phylogeny of 9 species, from the 15 included in this study. This phylogeny was used to investigate the geographical and ecological aspects of the evolution of this group. We found that closely related species tend to occur in sympatry and that different clades are allopatrically distributed. This pattern resulted in a significant negative correlation between the time of divergence events and sympatry, which suggests that non-allopatry is predominant within this clade. Furthermore, we found evidence for two different macroevolutionary processes having acted on the evolution of niche optima in sugarcane relatives. In relation to temperature, natural selection may underlie the divergence of sugarcane relatives in two ecological groups of species, one adapted to cold climates and the other adapted to warm climates. On the other hand, optima related to aridity seemed to arise under a neutral process.

Finally, we were able to demonstrate the importance of analyzing available biodiversity data in the light of ecological and evolutionary theory. However, even with our results, there is still a knowledge gap about the distribution and the phylogenetic relationships of sugarcane wild relatives, despite their economic relevance. This lack of information hinders the conservation of potential genetic resources to crops and a more complete understanding of ecological mechanisms underlying the diversification of this group.

References

- Amalraj, V. A. e Balasundaram, N. (2006). On the taxonomy of the members of 'Saccharum complex'. Genetic Resources and Crop Evolution, 53(1):35–41.
- Anacker, B. L. e Strauss, S. Y. (2014). The geography and ecology of plant speciation: range overlap and niche divergence in sister species. *Proceedings of the Royal Society B: Biological Sciences*, 281(1778):20132980.
- Atkinson, R. R., Mockford, E. J., Bennett, C., Christin, P.-A., Spriggs, E. L., Freckleton, R. P., Thompson, K., Rees, M., e Osborne, C. P. (2016). C 4 photosynthesis boosts growth by altering physiology, allocation and size. *Nature Plants*, 2(5):16038.
- Barraclough, T. G. e Vogler, A. P. (2000). Detecting the geographical pattern of speciation from specieslevel phylogenies. *The American Naturalist*, 155(4):419–434.
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberón, J., e Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222(11):1810–1819.
- Beaulieu, J. M. e O'Meara, B. (2014). Ouwie: analysis of evolutionary rates in an ou framework. *R* package version, 1.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., e Drummond, A. J. (2014). Beast 2: a software platform for bayesian evolutionary analysis. *PLoS* computational biology, 10(4):e1003537.
- Bouckaert, R. R. e Drummond, A. J. (2017). bmodeltest: Bayesian phylogenetic site model averaging and model comparison. *BMC evolutionary biology*, 17(1):42.
- Broennimann, O., Fitzpatrick, M. C., Pearman, P. B., Petitpierre, B., Pellissier, L., Yoccoz, N. G., Thuiller, W., Fortin, M.-J., Randin, C., Zimmermann, N. E., et al. (2012). Measuring ecological niche overlap from occurrence and spatial environmental data. *Global ecology and biogeography*, 21(4):481– 497.
- Brosse, N., Dufour, A., Meng, X., Sun, Q., e Ragauskas, A. (2012). Miscanthus: a fast-growing crop for biofuels and chemicals production. Biofuels, Bioproducts and Biorefining, 6(5):580–598.
- Butler, M. A. e King, A. A. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*, 164(6):683–695.

- Cai, Q., Aitken, K., Fan, Y., Piperidis, G., Jackson, P., e McIntyre, C. L. (2005). 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.
- Cai, W., Borlace, S., Lengaigne, M., Van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M. J., Wu, L., et al. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature climate change*, 4(2):111.
- Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., Guarino, L., Harker, R. H., Jarvis, A., Maxted, N., et al. (2016). Global conservation priorities for crop wild relatives. *Nature plants*, 2(4):16022.
- Clark, L. V., Dzyubenko, E., Dzyubenko, N., Bagmet, L., Sabitov, A., Chebukin, P., Johnson, D. A., Kjeldsen, J. B., Petersen, K. K., Jørgensen, U., et al. (2016). Ecological characteristics and *in situ* genetic associations for yield-component traits of wild *Miscanthus* from eastern Russia. *Annals of botany*, 118(5):941–955.
- Clark, L. V., Stewart, J. R., Nishiwaki, A., Toma, Y., Kjeldsen, J. B., Jørgensen, U., Zhao, H., Peng, J., Yoo, J. H., Heo, K., et al. (2015). Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in japan indicates a gradient of bidirectional but asymmetric introgression. Journal of Experimental Botany, 66(14):4213–4225.
- Clayton, W. D., Renvoize, S. A., et al. (1986). Genera graminum. grasses of the world. Genera graminum. Grasses of the World., 13.
- Coelho, M. T. P., Rodrigues, J. F. M., Diniz-Filho, J. A. F., e Rangel, T. F. (2019). Biogeographical history constrains climatic niche diversification without adaptive forces driving evolution. *Journal of Biogeography*, 46(5):1020–1028.
- Cook, B. I., Smerdon, J. E., Seager, R., e Coats, S. (2014). Global warming and 21 st century drying. *Climate Dynamics*, 43(9-10):2607–2627.
- Daniels, J., Roach, B., e Heinz, D. (1987). Sugarcane improvement through breeding. *Elsevier, Amster*dam.
- Dempewolf, H., Baute, G., Anderson, J., Kilian, B., Smith, C., e Guarino, L. (2017). Past and future use of wild relatives in crop breeding. *Crop Science*, 57(3):1070–1082.
- Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'amen, M., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., et al. (2017). ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40(6):774–787.
- Dong, H., Clark, L. V., Lipka, A. E., Brummer, J. E., Głowacka, K., Hall, M. C., Heo, K., Jin, X., Peng, J., Yamada, T., et al. (2019). Winter hardiness of *Miscanthus* (iii): Genome-wide association and genomic prediction for overwintering ability in *Miscanthus sinensis*. *GCB Bioenergy*.
- Ebrahim, M. K., Vogg, G., Osman, M. N., e Komor, E. (1998). Photosynthetic performance and adaptation of sugarcane at suboptimal temperatures. *Journal of Plant Physiology*, 153(5-6):587–592.
- Edwards, E. J., Osborne, C. P., Strömberg, C. A., Smith, S. A., Consortium, C. G., et al. (2010). The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science*, 328(5978):587–591.

- Estep, M. C., McKain, M. R., Diaz, D. V., Zhong, J., Hodge, J. G., Hodkinson, T. R., Layton, D. J., Malcomber, S. T., Pasquet, R., e Kellogg, E. A. (2014). Allopolyploidy, diversification, and the miocene grassland expansion. *Proceedings of the National Academy of Sciences*, 111(42):15149–15154.
- Evans, D., Joshi, S., et al. (2017). A new origin of sugarcane: the undiscovered species. In *Proceedings* of the Annual Congress-South African Sugar Technologists' Association, number 90, pages 163–166. South African Sugar Technologists' Association.
- Evans, D. L., Joshi, S. V., e Wang, J. (2019). Whole chloroplast genome and gene locus phylogenies reveal the taxonomic placement and relationship of *Tripidium* (panicoideae: Andropogoneae) to sugarcane. *BMC Evolutionary Biology*, 19(1):33.
- FAOSTAT, F. (2017). Food and agriculture data. Food and Agriculture Organization: Roma, Italy.
- Fitzpatrick, B. M. e Turelli, M. (2006). The geography of mammalian speciation: mixed signals from phylogenies and range maps. *Evolution*, 60(3):601–615.
- Glennon, K., Ritchie, M. E., e Segraves, K. (2014). Evidence for shared broad-scale climatic niches of diploid and polyploid plants. *Ecology letters*, 17(5):574–582.
- Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M., e Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *Science*, 327(5967):812–818.
- Godsoe, W. (2010). Regional variation exaggerates ecological divergence in niche models. Systematic Biology, 59(3):298–306.
- Griffith, D., Still, C., Cotton, J., e Powell, B. (2015). Comparing soil and bison delta13c to field estimates of c4 plant abundances in north america. In *AGU Fall Meeting Abstracts*.
- Group, G. P. W. (2012). New grass phylogeny resolves deep evolutionary relationships and discovers c4 origins. New Phytologist, 193(2):304–312.
- Guisan, A., Petitpierre, B., Broennimann, O., Daehler, C., e Kueffer, C. (2014). Unifying niche shift studies: insights from biological invasions. *Trends in ecology & evolution*, 29(5):260–269.
- Heibl, C., Calenge, C., e Heibl, M. C. (2018). Package 'phyloclim'.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., e Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of* the Royal Meteorological Society, 25(15):1965–1978.
- Hodkinson, T. R., Chase, M. W., Lledó, D. M., Salamin, N., e 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.
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., e Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46:523–549.
- Huang, Y., Wu, J., Wang, P., Lin, Y., Fu, C., Deng, Z., Wang, Q., Li, Q., Chen, R., e Zhang, M. (2015). Characterization of chromosome inheritance of the intergeneric BC2 and BC3 progeny between *Saccharum spp.* and *Erianthus arundinaceus*. *PLoS One*, 10(7):e0133722.

- Johnson, A. L., Govindarajulu, R., e Ashman, T.-L. (2014). Bioclimatic evaluation of geographical range in Fragaria (rosaceae): consequences of variation in breeding system, ploidy and species age. Botanical Journal of the Linnean Society, 176(1):99–114.
- Kantar, M. B., Sosa, C. C., Khoury, C. K., Castañeda-Álvarez, N. P., Achicanoy, H. A., Bernau, V., Kane, N. C., Marek, L., Seiler, G., e Rieseberg, L. H. (2015). Ecogeography and utility to plant breeding of the crop wild relatives of sunflower (*Helianthus annuus* l.). Frontiers in Plant Science, 6:841.
- Kar, S., Zhang, N., Nakashima, T., Villanueva-Morales, A., Stewart, J. R., Sacks, E. J., Terajima, Y., e Yamada, T. (2019). Saccharum× Miscanthus intergeneric hybrids (miscanes) exhibit greater chilling tolerance of C4 photosynthesis and postchilling recovery than sugarcane (Saccharum spp. hybrids). Global Change Biology. Bioenergy, 11(11):1318–1333.
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., e Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific data*, 4:170122.
- Kellogg, E. (2015). Poaceae. The families and genera of vascular plants, 13:1-416.
- Khoury, C. K., Castañeda-Alvarez, N. P., Achicanoy, H. A., Sosa, C. C., Bernau, V., Kassa, M. T., Norton, S. L., van der Maesen, L. J. G., Upadhyaya, H. D., Ramírez-Villegas, J., et al. (2015). Crop wild relatives of pigeonpea [cajanus cajan (l.) millsp.]: Distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. *Biological Conservation*, 184:259–270.
- Kim, C., Wang, X., Lee, T.-H., Jakob, K., Lee, G.-J., e Paterson, A. H. (2014). Comparative analysis of *Miscanthus* and *Saccharum* reveals a shared whole-genome duplication but different evolutionary fates. *The Plant Cell*, 26(6):2420–2429.
- Long, S. P. e Spence, A. K. (2013). Toward cool C4 crops. Annual Review of Plant Biology, 64:701–722.
- McCouch, S., Baute, G. J., Bradeen, J., Bramel, P., Bretting, P. K., Buckler, E., Burke, J. M., Charest, D., Cloutier, S., Cole, G., et al. (2013). Agriculture: feeding the future. *Nature*, 499(7456):23.
- Meyer, M. H. (2004). Miscanthus: ornamental and invasive grass. HortScience, 39(4):792B-792.
- Meyer, R. S. e Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature reviews genetics*, 14(12):840–852.
- Moore, P. H. e Botha, F. C. (2013). Sugarcane: physiology, biochemistry and functional biology. John Wiley & Sons.
- Mukherjee, S. K. (1957). Origin and distribution of Saccharum. Botanical Gazette, 119(1):55-61.
- Münkemüller, T., Boucher, F. C., Thuiller, W., e Lavergne, S. (2015). Phylogenetic niche conservatismcommon pitfalls and ways forward. *Functional ecology*, 29(5):627–639.
- Nakazato, T., Warren, D. L., e Moyle, L. C. (2010). Ecological and geographic modes of species divergence in wild tomatoes. *American Journal of Botany*, 97(4):680–693.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., et al. (2001). Terrestrial ecoregions of the world: A new map of life on eartha new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 51(11):933–938.

- Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., FitzJohn, R. G., Alfaro, M. E., e Harmon, L. J. (2014). geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics*, 30(15):2216–2218.
- Peterson, A., Soberón, J., e Sánchez-Cordero, V. (1999). Conservatism of ecological niches in evolutionary time. Science, 285(5431):1265–1267.
- Petitpierre, B., Kueffer, C., Broennimann, O., Randin, C., Daehler, C., e Guisan, A. (2012). Climatic niche shifts are rare among terrestrial plant invaders. *Science*, 335(6074):1344–1348.
- Phillips, S. J., Anderson, R. P., e Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4):231–259.
- Pignon, C. P., Spitz, I., Sacks, E. J., Jørgensen, U., Kørup, K., e Long, S. P. (2019). Siberian Miscanthus sacchariflorus accessions surpass the exceptional chilling tolerance of the most widely cultivated clone of Miscanthus × giganteus. GCB Bioenergy, 11(7):883–894.
- Pigot, A. L. e Tobias, J. A. (2015). Dispersal and the transition to sympatry in vertebrates. Proceedings of the Royal Society B: Biological Sciences, 282(1799):20141929.
- Piperidis, G., Piperidis, N., e D'Hont, A. (2010). Molecular cytogenetic investigation of chromosome composition and transmission in sugarcane. *Molecular Genetics and Genomics*, 284(1):65–73.
- Pyron, R. A., Costa, G. C., Patten, M. A., e Burbrink, F. T. (2015). Phylogenetic niche conservatism and the evolutionary basis of ecological speciation. *Biological Reviews*, 90(4):1248–1262.
- Ram, B., Sreenivasan, T., Sahi, B., e Singh, N. (2001). Introgression of low temperature tolerance and red rot resistance from *Erianthus* in sugarcane. *Euphytica*, 122(1):145–153.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., e Suchard, M. A. (2018). Posterior summarization in bayesian phylogenetics using Tracer 1.7. Systematic Biology, 67(5):901–904.
- Ray, D. K., Mueller, N. D., West, P. C., e Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PloS one*, 8(6):e66428.
- Sacks, E. J., Jakob, K., e Gutterson, N. I. (2013). High biomass *Miscanthus* varieties. US Patent App. 13/513,173.
- Sage, R. F. (2004). The evolution of C4 photosynthesis. New phytologist, 161(2):341-370.
- Schoener, T. W. (1970). Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, 51(3):408–418.
- Sexton, J. P., Montiel, J., Shay, J. E., Stephens, M. R., e Slatyer, R. A. (2017). Evolution of ecological niche breadth. Annual Review of Ecology, Evolution, and Systematics, 48:183–206.
- Skeels, A. e Cardillo, M. (2019). Reconstructing the geography of speciation from contemporary biodiversity data. The American Naturalist, 193(2):240–255.
- Soreng, R. J., Peterson, P. M., Romaschenko, K., Davidse, G., Teisher, J. K., Clark, L. G., Barberá, P., Gillespie, L. J., e Zuloaga, F. O. (2017). A worldwide phylogenetic classification of the Poaceae (Gramineae) ii: An update and a comparison of two 2015 classifications. *Journal of Systematics and Evolution*, 55(4):259–290.

- Syfert, M. M., Castañeda-Álvarez, N. P., Khoury, C. K., Särkinen, T., Sosa, C. C., Achicanoy, H. A., Bernau, V., Prohens, J., Daunay, M.-C., e Knapp, S. (2016). Crop wild relatives of the brinjal eggplant (*Solanum melongena*): poorly represented in genebanks and many species at risk of extinction. *American Journal of Botany*, 103(4):635–651.
- Trabucco, A. e Zomer, R. J. (2009). Global aridity index (global-aridity) and global potential evapotranspiration (global-pet) geospatial database. *CGIAR Consortium for Spatial Information*.
- Wallace, J. G., Rodgers-Melnick, E., e Buckler, E. S. (2018). On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annual review of genetics*.
- Wang, X.-H., Yang, Q.-H., Li, F.-S., He, L.-L., e He, S.-C. (2009). Molecular identification of Saccharum spp. × Erianthus fulvus hybrids using sequence-characterized amplified region markers. Crop science, 49(3):864–870.
- Warren, D. L., Cardillo, M., Rosauer, D. F., e Bolnick, D. I. (2014). Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution*, 29(10):572–580.
- Warren, D. L., Glor, R. E., e Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution: International Journal of Organic Evolution, 62(11):2868–2883.
- Watcharamongkol, T., Christin, P.-A., e Osborne, C. P. (2018). C4 photosynthesis evolved in warm climates but promoted migration to cooler ones. *Ecology Letters*, 21(3):376–383.
- Watson, L., Dallwitz, M. J., et al. (1992). The grass genera of the world. CAB international.
- Welker, C. A., Souza-Chies, T. T., Longhi-Wagner, H. M., Peichoto, M. C., McKain, M. R., e 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.
- Wiens, J. J. e Donoghue, M. J. (2004). Historical biogeography, ecology and species richness. Trends in Ecology & Evolution, 19(12):639–644.
- Wu, J., Huang, Y., Lin, Y., Fu, C., Liu, S., Deng, Z., Li, Q., Huang, Z., Chen, R., e 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.
- Zabel, F., Putzenlechner, B., e Mauser, W. (2014). Global agricultural land resources-a high resolution suitability evaluation and its perspectives until 2100 under climate change conditions. *PloS One*, 9(9):e107522.
- Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., e 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, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., et al. (2018). Allele-defined genome of the autopolyploid sugarcane Saccharum spontaneum l. Nature Genetics, 50(11):1565.
- Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., et al. (2019). Coordinate cleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, 10(5):744–751.