



# Plant diversity conservation in an agricultural frontier in the Brazilian Cerrado

Jéssica Cauana de Oliveira Santana<sup>1</sup> · Marcelo Fragomeni Simon<sup>2</sup>

Received: 3 June 2021 / Revised: 15 December 2021 / Accepted: 2 January 2022 /

Published online: 18 January 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

The conversion of natural landscapes associated with agricultural activities is one of the major causes for the decline of biodiversity, especially in tropical regions. However, conservation planning within these regions is often hampered by insufficient biological data. In this work, we compiled plant diversity data for the agricultural frontier known as Matopiba. This is an area experiencing accelerated habitat loss. Although home to one of the largest remnants of natural vegetation in the Brazilian Cerrado biodiversity hotspot, only scant knowledge of its plant diversity exists. In total, we compiled 24,312 unique angiosperm records belonging to 139 families, 796 genera and 2,517 species. Occurrence data are very sparse. Approximately 39% of the region has <math><0.1</math> record/km<sup>2</sup>, and only two areas have more than 1 record/km<sup>2</sup>. Rarefaction curves and a steady increase in new species discovery (27 new species from 2015–2020) indicate that knowledge of plant diversity in the study area is far from complete. Despite the significant proportion of collection gaps, our results suggest that the region's flora is considerably rich, including 54 endemics and 38 threatened species. However, only 15% of the flora has been evaluated for conservation status. Only 28% of endemic and 53% of threatened species are reported to occur in protected areas. It is expected that this study will encourage a greater sampling effort to fill gaps in floristic knowledge and contribute data to support biodiversity conservation in a key region that is under high pressure by agricultural expansion.

**Keywords** Biodiversity · Endemism · Flora · Matopiba · Threatened species

---

Communicated by Daniel Sanchez Mata.

✉ Jéssica Cauana de Oliveira Santana  
jessica.stna@hotmail.com

Marcelo Fragomeni Simon  
marcelo.simon@embrapa.br

<sup>1</sup> Programa de Pós-Graduação em Botânica, Universidade de Brasília, Brasília, Brazil

<sup>2</sup> Embrapa Recursos Genéticos e Biotecnologia, Brasília, Brazil

## Introduction

Currently, habitat loss and fragmentation are the main drivers of the decline in global biodiversity (Newbold et al. 2015), and their impacts on terrestrial ecosystems are expected to increase even more during this century (Haddad et al. 2015; Wilson et al. 2016), especially in the tropics where the largest biological diversity is concentrated (Collen et al. 2008). In tropical countries, loss of natural habitats is strongly related to changes in land use, especially agricultural activities (Gibbs et al. 2010; Newbold et al. 2015; Lewis et al. 2015). Brazil is considered one of the largest producers of agricultural commodities in the world, and as a megadiverse country, it faces the challenge of reconciling economic growth, largely driven by agriculture, with biodiversity conservation (Lemes et al. 2019).

The highest volumes of agricultural production in Brazil are concentrated in the Cerrado region, which was responsible for 55% of the country's grain harvest in 2019 (Conab 2019). The Cerrado covers 23% of Brazilian territory and is the second largest biome in South America, surpassed in area only by the Amazon (IBGE 2020). In addition, it harbours the tropical savanna with the richest flora in the world (Murphy et al. 2016; Borghetti et al. 2019) and home to more than 13,000 plant species, of which about 5000 are endemic (Flora do Brasil 2020). Despite being considered one of the world's biodiversity hotspots (Myers et al. 2000), the Cerrado is under continuous pressure from deforestation, which has resulted in the loss of approximately 50% of its original vegetation cover (Alencar et al. 2020), posing an increasing threat to its flora (Strassburg et al. 2017; Velazco et al. 2019).

According to the Brazilian National Space Research Institute (INPE 2020), the Cerrado has had 91,800 km<sup>2</sup> of its area deforested in the 2008–2018 surpassing by nearly 30,000 km<sup>2</sup> the deforested area in the entire Brazilian Amazon (62,300 km<sup>2</sup>), despite being 50% of its size (<http://terrabrasilis.dpi.inpe.br>). Recent deforestation rates in the Cerrado can be explained by the expansion of the agricultural frontier, most notably in northern Mato Grosso state, and in a region known as Matopiba, which accounts for the acronyms of four states in the northeastern part of the Cerrado (Maranhão, Tocantins, Piauí and Bahia). In the past few years, Matopiba has been portrayed as the new agricultural frontier of Brazil, given its remarkable potential for large-scale agriculture (e.g., Araújo et al. 2019).

Land conversion in Matopiba has occurred mainly through the replacement of native savannas and grasslands by mechanized agriculture with large-scale production. This process has been facilitated by environmental and topographic conditions favourable to agricultural practices, land availability and fiscal incentives (Bolfe et al. 2016). However, environmental concerns have grown since the Matopiba region also harbours the largest natural remnants of the highly fragmented Cerrado biome (Alencar et al. 2020). Besides important biodiversity repositories and providers of ecosystem services, remnants of native vegetation in Matopiba are vital for the subsistence of local traditional communities (Schmidt and Ticktin 2012; Borges et al. 2016).

A major challenge in conservation planning is the protection of areas that contain as many species as possible (Tjørve 2010). As conservation implies economic and social costs, a viable and efficient strategy involves focusing on areas with greater diversity, rarity and endemism (Cañadas et al. 2014; Enquist et al. 2019). Therefore, the choice of the most appropriate areas for conservation can be optimized with robust information on species occurrence data and the filling of knowledge gaps.

However, biodiversity conservation planning is often hampered by the absence or incomplete and outdated data on species occurrences (Darbyshire et al. 2017). In addition, sampling efforts are generally biased towards regions close to major cities and research

centres (e.g., Sousa-Baena et al. 2014). This practice favours the dissemination of misleading estimates on biodiversity data owing to biases introduced through an excessive representation of both well-studied taxonomic groups and regions (McRae et al. 2017).

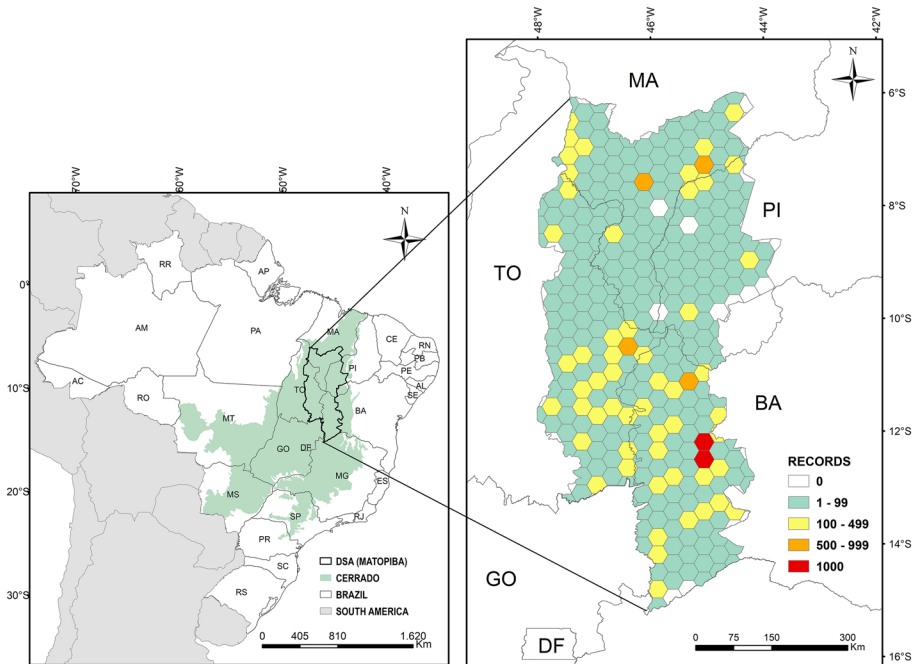
In the Cerrado, floristic inventories are concentrated in the southern and central regions (Sousa-Baena et al. 2014) with an overrepresentation of surveys focused on tree species occurring in savanna habitats (e.g., Ratter et al. 2003; Françoso et al. 2020). This concentration of biological information in a particular region, life form or habitat creates severe gaps in the knowledge of species distribution, hindering the implementation of adequate species conservation strategies. Botanical surveys in the Matopiba region thus far carried out have been restricted to only a few areas with particular preference for woody plants in savanna and grassland habitats (e.g., Antar and Sano 2019). However, a general compilation of the flora for the region as a whole, encompassing all different habitats and life forms, is still missing. Nevertheless, new approaches based on specimen data deposited in large public biodiversity repositories, where millions of species occurrence data are available online, offer new opportunities for assembling comprehensive floristic datasets (e.g. Maldonado et al. 2015). Such approaches can be an efficient way to obtain adequate estimates of species diversity and distribution for a particular region, including areas of conservation concern experiencing increased rates of habitat loss such as the Matopiba. However, limitations regarding data quality remain a major obstacle when using information from large biodiversity repositories (Colli-Silva et al. 2020).

Given ongoing impacts created by the expansion of large-scale agriculture on natural ecosystems in Matopiba, coupled with the lack of robust biodiversity information in the region, we herein asked about (i) the extent of flora sampling in the Matopiba region, (ii) the number of plant species occurring in Matopiba, and (iii) the number of these species endemic to the region, or listed as threatened. To accomplish these goals, we compiled botanical information on plant occurrence in Matopiba based on online public repositories, in order to provide robust data to support effective conservation strategies. The analysis of the data collected enabled us to identify areas of sampling gaps and produce the first compilation of the angiosperm flora in the region, highlighting species of high conservation value, such as threatened and endemic.

## Materials and methods

### Study site

In this study, we adopted a delimitation for the Matopiba region that is located between 6° 15'—15° 15' S and 48° 45'—44° 45' W (Fig. 1) and comprised an area of physiographic and climatic homogeneity of about 300.000 km<sup>2</sup> within the boundaries of 80 municipalities (Supplementary Material S1). As inclusion criteria for this delimitation, we adopted municipalities from Maranhão, Tocantins, Piauí and Bahia states that have most of their territory included within the boundaries of the Cerrado biome (sensu IBGE 2020) and that also have high potential for agricultural activity, including flat areas with deep soils suitable for the implementation of large-scale agriculture. The definition adopted here differs from a broader delimitation of Matopiba adopted in other studies (e.g., Miranda et al. 2014; Dias et al. 2016). We preferred to restrict the study area towards a more biologically homogeneous region in terms of climate and geology, thus concentrating most areas now facing intensive pressure for land use change (Dias et al. 2016; Parente and Ferreira



**Fig. 1** Cerrado biome in Brazil showing the location of the defined study area (DSA) in the Matopiba region (left), which was delimited by the territory of 80 municipalities (see also Supplementary Material S1), and density of occurrence records of angiosperms in the DSA based on 24,312 unique records (right). Each hexagonal grid-cell has an area of 1000 km<sup>2</sup>, and the colours represent the range of botanical records per grid-cell

2018). Therefore, areas with semi-arid climates in the east, as well as lowlands west of the Tocantins River and close to the Amazon border, are excluded from our study site, avoiding an excessively heterogeneous definition in terms of environmental variability and agricultural potential.

The defined study area (hereafter referred as DSA) used here largely coincides with the Cerrado ecoregion known as “Chapadão do São Francisco”, which was recently defined based on biophysical characteristics and encompasses a region with flat terrain in plateaus with altitudes above 500 m (Sano et al. 2019). This area also broadly overlaps with the physiography unit called Espigão Mestre do São Francisco (Cochrane et al. 1985), which is a study area considered in many floristic and biogeographic studies (e.g., Mendonça et al. 2000; Felfili et al. 2001).

The main vegetation types of the DSA are the savannas and grasslands that occur preferably on flat areas on well-drained terrain, while gallery forests and palm swamps (*veredas*) occur in association with water courses. Patches of seasonally dry tropical forests occur scattered throughout the region, often associated with limestone outcrops. The climate is classified as tropical semi-arid/humid with seasonal (summer) rains, annual rainfall ranging from 1000–1400 mm, and mean annual temperatures ranging from 23–26 °C (IBGE 2020). Predominant soil classes are the dystrophic latosols, as well as sandstone/quartzite-derived soils, which have good permeability, but limited fertility and low pH (Santos et al. 2018). In addition, the DSA includes important water sources that drain into major hydrographic basins, which are often used in irrigation by pivots (Landau et al. 2016).

## Data acquisition and analyses

We concentrated our analyses on the angiosperms because they are the most diverse and well-studied group of plants (Magallón and Castillo 2009), as well as a preferred target of biodiversity assessments and conservation planning. In order to build a dataset for the flora of Matopiba, we conducted a search for occurrence records of angiosperms available for the DSA in the SpeciesLink (2021; <http://specieslink.net>) online herbarium repository, which embraces all major herbaria in Brazil. We also searched the Botanical Collection Management System—JABOT (2021; <http://jabot.jbrj.gov.br/v3/consulta.php>) for botanical data of the BRBA and IBGE herbaria, which house relevant collections from our study area but are not available in SpeciesLink. All repositories were last accessed in September 2021.

Our searching criteria was based on the 80 municipalities that comprise the DSA (Supplementary Material S1). Multiple searches were performed by entering the name of each municipality in the search operator “municipality” (SpeciesLink) or “locality” (JABOT). This search resulted in the download of 44,948 records of angiosperms from 106 herbaria (SpeciesLink) and 10,158 records from the BRBA and IBGE herbaria (JABOT). We also obtained 434 occurrence records (observation data) from a dataset on savanna woody species (Françoso et al. 2020) containing surveyed sites that fall within the DSA. This initial compilation comprised 55,540 occurrence records.

After compilation, taxonomic names were checked for spelling errors, and synonyms were updated using the R package *flora* v. 0.3.4 (Carvalho 2020; R Development Core Team 2020), which retrieves data from the *Flora do Brasil 2020* platform. Varieties and subspecies were collapsed to species level. The data were filtered to eliminate specimens with doubtful identifications (incomplete or indeterminate) and botanical duplicates (specimens representing a single collection deposited in several herbaria). Duplicates were detected by manually searching for records with the same collecting date, collector number and collector name, and only one record per collection was kept in the dataset. In case of duplicates with different identifications, we kept the specimen determined by specialists, or considered the latest determination. Also, we excluded species that are non-native to the study region, which were identified with the help of the *flora* package version 0.3.4 (Carvalho 2020). For records lacking original geographic coordinates (c. 34% of all records) we used the centroid of the municipality where the specimen was collected, as provided by SpeciesLink. Although using municipality centroid, which may be some kilometres away from the locality where the specimen was actually collected, could result in less precise estimates of specimen georeferencing, this was not a major issue considering the large geographical scale of our study. Records with imprecise geographic coordinates (i.e., only to degrees or with geographic coordinates not matching collection locality) were manually checked and eventually eliminated. After the cleaning procedures described above, the final dataset contained 24,312 unique records of angiosperms native to the DSA (Supplementary Material S2).

All occurrence records were mapped into a grid consisting of 1000 km<sup>2</sup> hexagonal cells superimposed on the study area boundaries, using ArcGis version 10.3.1 (ESRI) software. The density of species occurrence by grid cell was used to access gaps of botanical knowledge (angiosperm records) in the DSA. Species richness was estimated using rarefaction and extrapolation curves as a function of Hill numbers in the order  $q=0$ , according to the method proposed by Chao et al. (2014), as implemented in the iNEXT package, version 2.0.17 (Hsieh et al. 2016; R Development Core Team 2020).

The curves were generated with 95% confidence intervals for the DSA, as well as for individual states.

Based on our occurrence dataset, we assembled a list of angiosperm species for the DSA, by choosing one representative specimen, preferably determined by a specialist, to represent each species. Species only represented by specimens determined by non-specialists also occurred in our dataset. Those records were checked individually for their taxonomic accuracy, and specimen's images were compared with specimens identified by specialists in biodiversity repositories such as SpeciesLink and Flora do Brasil 2020. A total of 352 species determined by non-specialists or without information on the determinant were excluded from our list due to unreliable identifications (Supplementary Material S2). The classification of the families follows the Angiosperm Phylogeny Group IV system (APG IV 2016). Species nomenclature and synonymy, as well as information on life form and geographical distribution, were revised and corrected based on the Brazilian Flora Checklist (Flora do Brasil 2020), using the flora package (Carvalho 2020).

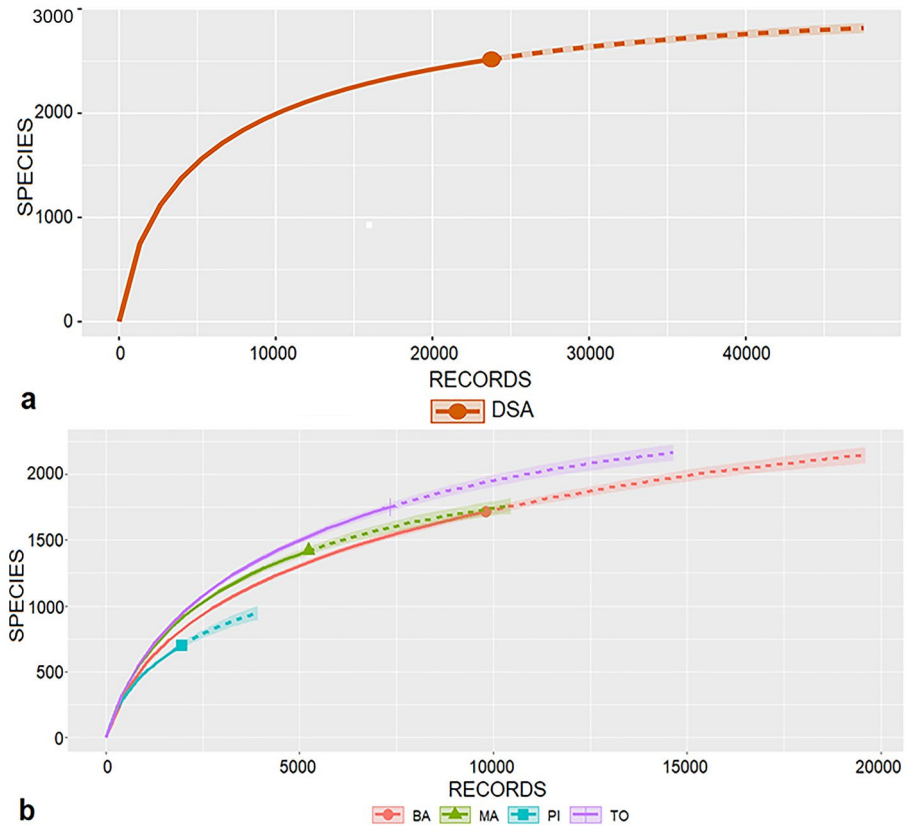
We checked the geographic distribution of each species present in our list in online databases (SpeciesLink and Flora do Brasil 2020) to search for species endemic to the DSA, i.e., species that only occur within the limits of our study area. Some species that occur predominantly in the DSA, but are also recorded along the Serra Geral in Goiás near the border with Bahia, were also included in our list of endemic species. We also searched for the occurrence of threatened species in the DSA based on the Brazilian Red List of the National Centre for Flora Conservation (CNCFlora 2021; <http://cncflora.jbrj.gov.br/portal/pt-br/listavermelha>), which uses threat categories established by the International Union for the Conservation of Nature and Natural Resources (IUCN): Critically Endangered, Endangered and Vulnerable. Subsequently, the endemic and threatened species had their distribution superimposed on the protected areas of the region (<http://mapas.mma.gov.br>) in order to verify if their occurrence lies within areas protected by law. To access recent botanical discoveries in the region, we also conducted searches in the International Plant Names Index ([www.ipni.org](http://www.ipni.org)) database in order to identify species published in the last five years (2015–2020) that occur within the DSA.

## Results

The distribution of angiosperm occurrence records in the study region is very sparse (Fig. 1) with about 39% of grid cells showing less than 10 records. In general, the DSA has 0.08 record/km<sup>2</sup>, and only three grid cells have a density of > 1 record/km<sup>2</sup>. Individually, the states with the highest sampling effort are Bahia (0.10 record/km<sup>2</sup>) and Tocantins (0.08), followed by Maranhão (0.07) and Piauí (0.03).

The estimate of species richness based on the rarefaction curve of occurrence records indicates that the angiosperm flora of the DSA is still underestimated with an extrapolated increase of around 25% in the number of species, if sampling effort is doubled (Fig. 2A). The same applies to the states analysed individually with Piauí showing the lowest sampling effort and estimates of species richness (Fig. 2B).

We found 2,517 species in the study region, belonging to 796 genera and 139 families (the species list, including species authorities, is presented in Supplementary Material S3). Families with highest species richness were Fabaceae (330), Poaceae (196), Asteraceae (152), Cyperaceae (103), Rubiaceae (101), Euphorbiaceae (85), Malvaceae (84), Malpighiaceae (75), Melastomataceae (67) and Apocynaceae (62). The top ten



**Fig. 2** Rarefaction (continuous lines) and extrapolation (dashed lines) curves representing the number of angiosperm species sampled in the defined study area (DSA) in the Matopiba region (**a**) and in each state that composes it, individually (**b**). The 95% confidence intervals (shaded area that accompanies the lines) were obtained using the bootstrap method

species-rich genera were *Paspalum* (53), *Chamaecrista* (37), *Mimosa* (36), *Rhynchospora* (33), *Cyperus* (22), *Eugenia* (22), *Miconia* (22), *Byrsonima* (21), *Croton* (21), and *Hyptis*, *Ipomoea* and *Polygala* (all with 20). The ten species with the highest number of occurrences were *Qualea parviflora* (139), *Vochysia gardneri* (138), *Pouteria ramiflora* (134), *Connarus suberosus* (122), *Hirtella ciliata* (121), *Tachigali subvelutina* (121), *Exellodendron cordatum* (115), *Dimorphandra gardneriana* (114), *Rourea induta* (110), and *Eugenia puniceifolia* (109), all trees or large shrubs.

We recorded 54 species endemic to the DSA, of which only 15 are reported to occur in protected areas (Table 1; Supplementary Material S4). We highlight the occurrence of *Keratochlaena rigidifolia* (Poaceae), a species that belongs to a monospecific genus endemic to the DSA. We also report in our compilation 38 threatened species (Supplementary Material S4; Table 1; Fig. 3) included in the Red List of the National Centre for Flora Conservation, seven of which are critically endangered (CR), 17 endangered (EN) and 14 vulnerable (VU); 20 of these species occur in protected areas. Our survey on the taxonomic literature between 2015 and 2020 found 27 new species collected in the DSA (Supplementary Material S5), of which 23 are considered endemic to the region.



**Table 1** Summary information on angiosperm diversity compiled for the defined study area (DSA) in the agricultural frontier of the Matopiba in the Brazilian Cerrado

Data	Values
Occurrence records obtained from online public repositories (before data cleaning)	55,540
Occurrence records after data cleaning (elimination of records with doubtful taxonomic identification, non-native species, imprecise locality, and duplicated specimens)	24,312
Number of species recorded	2,517
Number of species endemic to DSA	54
Number of endemic species occurring in protected areas	15 (28%)
Number of species evaluated for conservation status by CNC Flora	377 (15%)
Critically endangered	7
Endangered	17
Vulnerable	14
Data deficient	30
Not threatened or least concern	309
Number of threatened species occurring in protected areas	20 (53%)

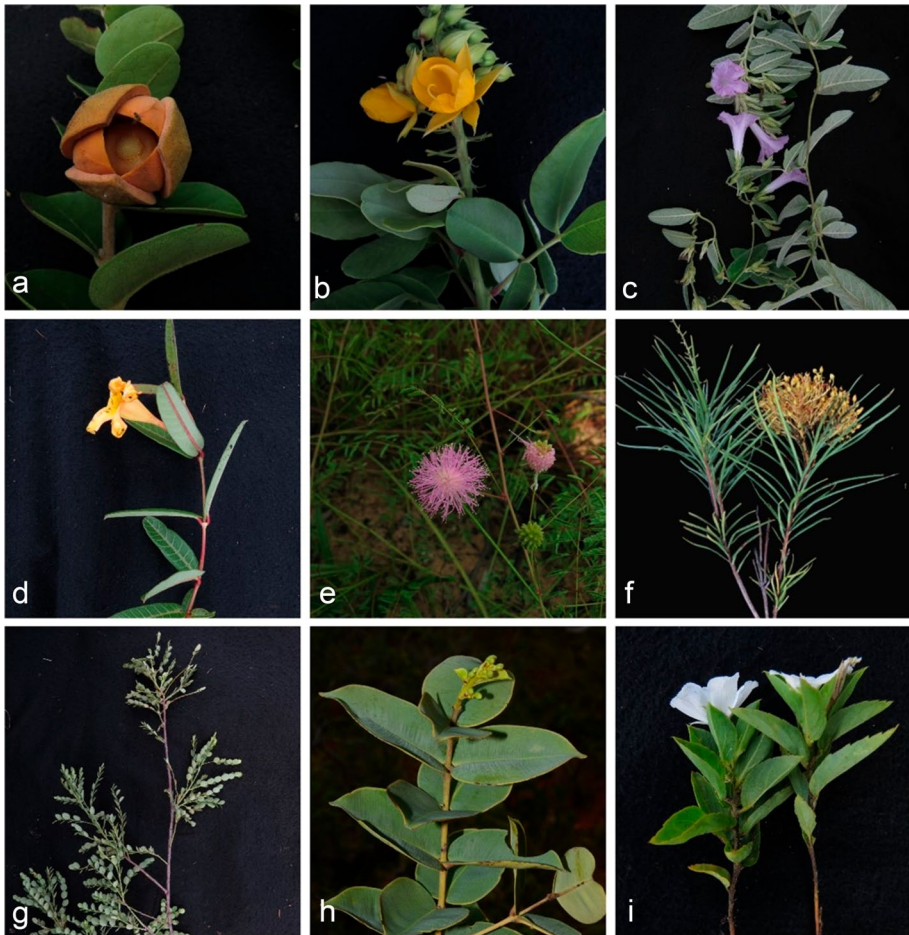
## Discussion

Owing to its vast stock of mostly unoccupied land, Matopiba appears to be a key region, both in terms of agricultural potential and conservation of biodiversity and natural resources. Therefore, reconciling agricultural and economic activities with preservation of biodiversity is central for the sustainable development of that region. Conservation planning in the DSA should rely, among other factors, on the availability of accurate biological data. However, this is not the case for the DSA since most of its territory is poorly sampled. Overall, the floristic sampling effort reported here (0.08 record/km<sup>2</sup>) is well below Brazil's average (0.41 record/km<sup>2</sup>; Sousa-Baena et al. 2014) and far from what is considered an adequate collection effort in tropical regions (1–3 records/km<sup>2</sup>; Campbell 1989).

Some of the better-inventoried areas in the DSA (red spots in Fig. 1B) correspond to the municipalities of Barreiras and São Desidério (Bahia), all with more than 1 record/km<sup>2</sup>. Other intensive botanical surveys carried out in the DSA were associated with the environmental licensing of hydroelectric dams (Estreito Dam, Maranhão/Tocantins; Medeiros et al. 2012), the implementation of large agricultural projects (Balsas, Maranhão; Aquino et al. 2007), and inventories with scientific purposes in western Bahia (Mendonça et al. 2000; Santana et al. 2010; Ribeiro et al. 2020), Maranhão (Conceição and Santos 2014; Silva-Moraes et al. 2019) Tocantins (Proença et al. 2002; Antar and Sano 2019), and Piauí (Castro et al. 1998).

Our compilation revealed that the DSA harbours a considerably rich flora (2,517 species), which represents 20% of the Cerrado angiosperm flora (Flora do Brasil 2020). Although presenting a considerable size, the compilation produced here is comparable to the number of species recorded in the Distrito Federal (3634 species; Flora do Brasil 2020), which is ca. 50 times smaller than that of the DSA, while having the best-known flora in the Cerrado region. Therefore, our dataset certainly underestimates the total number of species in the DSA. With increasing sampling effort, mainly in the gap areas, we estimate that the regional compilation could well surpass 5000 species. We highlight that the list presented here is based on a careful selection of herbarium specimens identified





**Fig. 3** Endemic and threatened species found in the defined study area (DSA) in the Matopiba region: *Annona gardneri* (a) *Chamaecrista coradinii* (b) *Ipomoea maranhensis* (c) *Mandevilla abortiva* (d) *Mimosa carolina* (e) *Ouratea acicularis* (f) *Phyllanthus allemii* (g) *Qualea hannekesaskiarum* (h) and *Turnera macrosperma* (i). Photo credits: Marcelo Simon (a–g, i) and Maurício Figueira (h). See more details in Supplementary Material S4

by specialists, and that records with inaccurate/suspicious identifications have been discarded. It is possible, however, that species only represented by old collections lacking precise locality data, which were not retrieved from public online repositories by our search strategy (municipality), may be absent from our list. For example, *Goyazia villosa* (Gesneriaceae), which was collected only once in 1841, was not retrieved by our initial search, but was later incorporated into our dataset once we occasionally came across some records available online. Limitations involving the use of specimens with imprecise locality data, such as those found in old herbarium collections, have been regarded as one of the major shortfalls in biodiversity documentation (Colli-Silva et al. 2020).

In addition to geographic gaps, we also found that collections are strongly biased towards woody species since all top ten species with the highest number of records in our dataset are trees or large shrubs. However, the rich herb-shrub layer that dominates

grasslands and savannas and makes up the largest portion of the Cerrado's plant diversity (Amaral et al. 2017), besides representing about 80% of the endemic species of the DSA, remains poorly sampled. Therefore, sampling of this component, particularly the endemic species known from a single or few collections (eg. *Goyazia villosa*), should be intensified and incorporated into broader biodiversity surveys and conservation efforts.

The DSA flora harbours a suite of woody species considered typical of the northeastern part of the Cerrado. The region largely overlaps with the northeastern Cerrado floristic province of Ratter et al. (2003) and also occupies parts of the northeastern and northwestern districts proposed by Françoso et al. (2020). Six out of the ten most frequent species in our database, except for *Cenostigma macrophyllum*, *Eugenia puniceifolia*, *Exellodendron cordatum* and *Vochysia gardneri*, are cited as among the 35 most constant woody species in the northeastern province (Vieira et al. 2019). In relation to the herb-shrub flora, the DSA is contained within phytogeographic region 6 (central-north) of Amaral et al. (2017).

Overall, evidence from previous studies on the woody and herb-shrub flora, together with our results, suggest that DSA is part of a floristic region located in the northeastern periphery of the Cerrado, which is characterized by a considerable fraction of regionally exclusive species associated with a suite of climatic variables (Vieira et al. 2019; Françoso et al. 2020). The geographic location of the DSA facilitates the interchange of floristic elements with adjacent Amazonia (north) and Caatinga (east) biomes, which could have been influenced by vegetation dynamics caused by climatic oscillations during the Quaternary (Vieira et al. 2019). Typical Amazon taxa in the DSA flora can be mainly found in wet forests associated with water courses, while Caatinga elements often occur in seasonally dry forests on limestone outcrops.

Most taxa endemic to the DSA occur in savannas and grasslands on high altitude relief (> 700 m). This corroborates the findings that centres of endemism within the Cerrado are correlated with open habitats in higher altitudes areas, such as plateaus and chapadas (Simon and Proença 2000; Echternacht et al. 2011). In a recent survey on Cerrado endemic flora, Vidal et al. (2019) identified two endemism centres (Northern and Central Bahia western plateau) within the DSA. Curiously, none of the endemic species listed here was mentioned in the Vidal et al. (2019) endemic survey.

Besides endemism, we found 38 threatened species occurring in the DSA, including six endemics (*Chamaecrista coradinii*, *Diplusodon gracilis*, *Hyptidendron conspersum*, *Janusia christianeae*, *Ouratea acicularis* and *Schultesia irwiniana*). However, it is important to mention that only 15% of the 2,517 species recorded in the region were evaluated for the degree of threat by Brazil's red list authority CNCFlora (2021). Considering the number of species that were not assessed, especially in areas of high anthropic pressure, we can infer that the number of threatened species in the DSA is clearly underestimated. Therefore, we urge that a more comprehensive survey be carried out in the region in order provide sufficiently robust data to allow reliable assessments of species conservation status, particularly for the endemics, in view of the tendency toward habitat loss and fragmentation forecast for the DSA (Barbosa-Silva and Antar 2020). We found that only 28% of endemic species and 53% of threatened species are recorded in protected areas in the region. However, these values might be underestimated since our knowledge of species occurrences within protected areas in the DSA, as well in other Brazilian reserves, is very limited (Oliveira et al. 2017).

Recent botanical surveys associated with taxonomic work have revealed the DSA to be a promising source of new species discoveries with an average of approximately four new species published every year since 2015 (IPNI 2021). This trend is likely to continue as more areas are surveyed and herbarium specimens are made available to specialists.

Among the 27 new species that were published in the last six years, only *Dioscorea compacta*, *Schizachyrium vallsii*, and *Stigmaphyllon occidentale* are not endemic to the DSA, suggesting that most species discoveries are expected to be restricted to the region. It is important to note that more intensive surveys in surrounding areas can reveal that some species referred here as endemic actually have a broader geographical distribution, and therefore may no longer be considered endemic to the DSA.

## Conclusion

Although it should be supplemented with additional surveys, our compilation represents a relevant reference for the angiosperm flora of Matopiba, containing a reliable dataset of plant occurrence data that can be used in downstream biodiversity analyses. Plant diversity data gathered here should be used in planning conservation actions in the study area, which is under ongoing habitat loss. Expansion of agricultural activities on native vegetation should consider the existence of plant species of high conservation value in the region. Deforestation in areas where populations of endemic and threatened species occur should be avoided in order to minimize the impact of habitat loss on plant populations.

Future studies should focus on inventorying the flora in the gap areas in order to provide better data on composition and distribution of species, especially those that are rare, endemic and threatened and, hence, considered to be of high conservation value. Floristic inventories covering diverse habitats and all life forms should be carried out in protected areas to know how much of DSA's flora is represented within protected areas.

It is also vital to increase the number of species assessed for their conservation status, especially those reported as endemic, since only a small fraction of species have been evaluated so far. Lastly, it is important to improve support for research institutions and herbaria in order to foster taxonomic and floristic studies in the region, which would result in more robust biodiversity datasets for conservation planning. This is particularly relevant since many new species, which are being found at high rates in the DSA, could be lost before being collected and formally described. We hope that our results will encourage the development of floristic studies in the region, helping to fill gaps in scientific knowledge and subsidize conservation planning in this complex landscape matrix that includes the largest remnants of the Cerrado biodiversity hotspot and an expanding agricultural frontier.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10531-022-02356-2>.

**Acknowledgements** We thank the Embrapa–Cenargen staff for assistance during herbarium work and field expeditions, as well as Glocimar Pereira-Silva, Mateus Martins, Maurício Figueira and Bianca Schindler for contributing floristic data. We also thank Juliana Rando (Univ. Federal do Oeste da Bahia) for providing data from the BRBA herbarium and Renata Françaço for sharing a database of Cerrado woody plants. We are indebted to Carol Barradas and Marco Borges (ICMBio) and the staff of the Estação Ecológica Serra Geral do Tocantins for their generous support during fieldwork and authorization for plant collections (SISBIO 61941-1). The first author thanks the Coordination for the Improvement of Higher-Level Personnel (CAPES, finance code 88882.387125/2019-01) for the scholarship during her PhD in the Programa de Pós Graduação em Botânica, University of Brasília.

**Author contributions** All authors contributed to the study conception and design, data collection, material preparation and analysis. The first draft of the manuscript was written by Jéssica Cauana de Oliveira Santana and the author Marcelo Fragomeni Simon commented and revised on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Partial financial support was received from Coordination for the Improvement of Higher-Level Personnel—CAPES, finance code 88882.387125/2019–01.

**Data availability** All data generated or analysed during this study are included in this published article and its supplementary information files.

**Code availability** Not applicable.

## Declarations

**Conflict of interest** The authors declare no conflicts of interest.

**Ethical approval** Not applicable.

**Consent to participate** The authors consent to participate.

**Consent for publication** The authors consent for publication.

## References

- Alencar AZ, Shimbo J, Lenti F et al (2020) Mapping three decades of changes in the Brazilian savanna native vegetation using Landsat data processed in the google earth engine platform. *Remote Sensing* 12(6):924. <https://doi.org/10.3390/rs12060924>
- Amaral AG, Munhoz CB, Walter BMT et al (2017) Richness pattern and phytogeography of the Cerrado herb–shrub flora and implications for conservation. *J Veg Sci* 28(4):848–858. <https://doi.org/10.1111/jvs.12541>
- Antar GM, Sano PT (2019) Angiosperms of dry grasslands and savannahs of Jalapão, the largest conserved Cerrado area in Brazil. *Rodriguésia* 70:1–9. <https://doi.org/10.1590/2175-7860201970070>
- Aquino FDG, Walter BMT, Ribeiro JF (2007) Woody community dynamics in two fragments of “cerrado” stricto sensu over a seven-year period (1995–2002), MA, Brazil. *Braz J Bot* 30(1):113–121. <https://doi.org/10.1590/S0100-84042007000100011>
- Araújo MLS, Sano EE, Bolfe ÉL et al (2019) Spatiotemporal dynamics of soybean crop in the Matopiba region, Brazil (1990–2015). *Land Use Policy* 80:57–67. <https://doi.org/10.1016/j.landusepol.2018.09.040>
- Barbosa-Silva RG, Antar GM (2020) Description vs deforestation-*Couepia brevistaminea* (Chrysobalanaceae), a new species on the frontier of agricultural expansion in the Brazilian savanna. *Phytotaxa* 471(1):38–46. <https://doi.org/10.11646/phytotaxa.471.1.4>
- Bolfe ÉL, Victória DDC, Contini E et al (2016) Matopiba em crescimento agrícola: Aspectos territoriais e socioeconômicos. *Revista de Política Agrícola* 25(4):38–62
- Borges SL, Eloy L, Schmidt IB, Barradas ACS, Dos Santos IA (2016) Manejo do fogo em veredas: novas perspectivas a partir dos sistemas agrícolas tradicionais no Jalapão. *Ambiente & Sociedade* 19(3):275–300. <https://doi.org/10.1590/1809-4422ASOC20150020R1V1932016>
- Borghetti F, Barbosa E, Ribeiro L et al (2019) South American Savannas. In: Scogings PF, Sankaran (orgs) Savanna woody plants and large herbivores. John Wiley & Sons, New Jersey, pp 110–140
- Campbell DG (1989) The importance of floristic inventory in the tropics. In: Campbell DG, Hammond D (eds) Floristic inventory of tropical countries: the status of plant systematics, records, and vegetation, plus recommendations for the future. New York Botanical Garden, New York, pp 5–30
- Cañadas EM, Fenu G, Peñas J (2014) Hotspots within hotspots: Endemic plant richness, environmental drivers, and implications for conservation. *Biol Cons* 170(1):282–291. <https://doi.org/10.1016/j.biocon.2013.12.007>
- Carvalho G (2020) Flora: Tools for interacting with the Brazilian Flora 2020. R package version 0.3.4. Available at: <http://CRAN.R-project.org/package=flora>
- Castro AAJF, Martins FR, Fernandes AG (1998) The woody flora of cerrado vegetation in the state of Piauí, northeastern Brazil. *Edinb J Bot* 55(3):455–472. <https://doi.org/10.1017/S0960428600003292>
- CNCFlora - Centro Nacional de Conservação da Flora (2021) Available at: <http://cncflora.jbrj.gov.br/portal/pt-br/listavermelha>. Accessed 10 Jan 2021

- Chao A, Gotelli NJ, Hsieh TC et al (2014) Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol Monogr* 84(1):45–67. <https://doi.org/10.1890/13-0133.1>
- Cochrane TT, Sanchez LG, Porras JA et al (1985). Land in tropical America. Embrapa Cerrados, Planaltina, Distrito Federal
- Collen B, Ram M, Zamin T, McRae L (2008) The tropical biodiversity data gap: addressing disparity in global monitoring. *Tropical Cons Sci* 1(2):75–88. <https://doi.org/10.1177/194008290800100202>
- Colli-Silva M, Reginato M, Cabral A et al (2020) Evaluating shortfalls and spatial accuracy of biodiversity documentation in the Atlantic Forest, the most diverse and threatened Brazilian phyto-geographic domain. *Taxon* 69:567–577. <https://doi.org/10.1002/tax.12239>
- Companhia Nacional de Abastecimento - Conab (2019) Acompanhamento da safra brasileira. Available at: <https://www.conab.gov.br>. Accessed 10 Jan 2020
- Conceição GM, Santos RM (2014) Diversidade florística das diferentes fisionomias de cerrado do Parque Estadual do Mirador, Maranhão, Brasil. *Brazilian Geographical Journal: Geosci Hum Res Medium* 5(1):139-156
- Darbyshire I, Anderson S, Asatryan A et al (2017) Important Plant Areas: revised selection criteria for a global approach to plant conservation. *Biodivers Conserv* 26(8):1767–1800. <https://doi.org/10.1007/s10531-017-1336-6>
- Dias LC, Pimenta FM, Santos AB et al (2016) Patterns of land use, extensification, and intensification of Brazilian agriculture. *Glob Change Biol* 22(8):2887–2903. <https://doi.org/10.1111/gcb.13314>
- Echternacht L, Trovó M, Oliveira CT et al (2011) Areas of endemism in the Espinhaço range in Minas Gerais. *Brazil Flora* 206(9):782–791. <https://doi.org/10.1016/j.flora.2011.04.003>
- Enquist BJ, Feng X, Boyle B et al (2019) The commonness of rarity: global and future distribution of rarity across land plants. *Sci Adv* 5(11):eaaz0414. <https://doi.org/10.1126/sciadv.aaz0414>
- Felfili JM, Sevilha AC, Silva Júnior MC (2001) Comparação entre as unidades fisiográficas Chapada Pratinha, Veadeiros e Espigão Mestre do São Francisco. In: Felfili JM, Silva Júnior MC (orgs) Biogeografia do bioma Cerrado: estudo fitofisionômico da Chapada do Espigão Mestre do São Francisco. Universidade de Brasília, Brasília, pp 80–152.
- Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro. Available at: <http://floradobrasil.jbrj.gov.br/>. Accessed 10 Sep 2021
- Françoso RD, Dexter KG, Machado RB et al (2020) Delimiting floristic biogeographic districts in the Cerrado and assessing their conservation status. *Biodivers Conserv* 29(5):1477–1500. <https://doi.org/10.1007/s10531-019-01819-3>
- Gibbs HK, Ruesch AS, Achard F et al (2010) Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc Natl Acad Sci* 107(38):16732–16737. <https://doi.org/10.1073/pnas.0910275107>
- Haddad NM, Brudvig LA, Clobert J et al (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci Adv* 1(2):e1500052. <https://doi.org/10.1126/sciadv.1500052>
- Hsieh TC, Ma KH, Chao A (2016) iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol Evol* 7(12):1451–1456. <https://doi.org/10.1111/2041-210X.12613>
- IBGE - Instituto Brasileiro de Geografia e Estatística (2020) Available at: <https://www.ibge.gov.br>. Accessed 10 December 2020
- INPE - Instituto Nacional de Pesquisas Espaciais (2020) Available at: <http://terrabrasilis.dpi.inpe.br>. Accessed 10 Jan 2020
- IPNI - International Plant Names Index (2021) The Royal Botanic Gardens, Kew, Harvard University Herbaria & Libraries and Australian National Botanic Gardens. Available at: <http://www.ipni.org>. Accessed 10 Sept 2021
- APG IV (2016) An update of the angiosperm phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* 181(1):1–20. <https://doi.org/10.1111/boj.12385>
- JABOT (2021). Sistema de Gerenciamento de Coleções Botânicas. Available at: <http://jabot.jbrj.gov.br/v3/consulta.php>. Accessed 10 Sept 2021
- Landau EC, Guimarães DP, de Sousa DL (2016) Expansão geográfica da agricultura irrigada por pivôs centrais na Região do Matopiba entre 1985 e 2015. Embrapa Milho e Sorgo, Sete Lagoas, Minas Gerais
- Lemes L, de Andrade AFA, Loyola R (2019) Spatial priorities for agricultural development in the Brazilian Cerrado: may economy and conservation coexist? *Biodivers Conserv* 29:1683–1700. <https://doi.org/10.1007/s10531-019-01719-6>
- Lewis SL, Edwards DP, Galbraith D (2015) Increasing human dominance of tropical forests. *Science* 349(6250):827–832. <https://doi.org/10.1126/science.aaa9932>
- Magallón S, Castillo A (2009) Angiosperm diversification through time. *Am J Bot* 96(1):349–365. <https://doi.org/10.3732/ajb.0800060>



- Maldonado C, Molina CI, Zizka A et al (2015) Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecol. Biogeogr* 24:973–984. <https://doi.org/10.1111/geb.12326>
- McRae L, Deinet S, Freeman R (2017) The diversity-weighted living planet index: controlling for taxonomic bias in a global biodiversity indicator. *PLoS ONE* 12(1):e0169156. <https://doi.org/10.1371/journal.pone.0169156>
- Medeiros MB, Walter BMT, Cavalcanti TB et al (2012) Recursos genéticos vegetais: florística, fitossociologia, resgate e conservação ex situ. In: Netto ALC (org) Estreito: o novo cenário da água. E-Papers/Ceste, Rio de Janeiro, pp 89–114
- Mendonça RC, Felfili JM, Fagg CW et al (2000) Florística da região do Espigão Mestre do São Francisco, Bahia e Minas Gerais. *Boletim do Herbário Ezechias Paulo Heringer* 6:38–94
- Miranda EE, Magalhães LA, Carvalho CA (2014) Proposta de delimitação territorial do MATOPIBA. Embrapa Territorial, Campinas, São Paulo
- Murphy BP, Andersen AN, Parr CL (2016) The underestimated biodiversity of tropical grassy biomes. *Philos Trans R Soc Lond, B Biol Sci* 371(1703):20150319. <https://doi.org/10.1098/rstb.2015.0319>
- Myers N, Mittermeier RA, Mittermeier CG et al (2000) Biodiversity hotspots for conservation priorities. *Nature* 403(6772):853–858. <https://doi.org/10.1038/35002501>
- Newbold T, Hudson LN, Hill SL et al (2015) Global effects of land use on local terrestrial biodiversity. *Nature* 520(7545):45–50. <https://doi.org/10.1038/nature14324>
- Oliveira U, Soares-Filho BS, Paglia AP et al (2017) Biodiversity conservation gaps in the Brazilian protected areas. *Sci Rep* 7(1):1–9. <https://doi.org/10.1038/s41598-017-08707-2>
- Parente L, Ferreira L (2018) Assessing the spatial and occupation dynamics of the Brazilian pasturelands based on the automated classification of MODIS images from 2000 to 2016. *Remote Sensing* 10(4):606. <https://doi.org/10.3390/rs10040606>
- Proença C, Sampaio AB, Silva LH, Milhomens LC et al (2002) Relatório de Botânica. In: Arruda MB, Von Behr M (org) Jalapão: expedição científica e conservacionista. IBAMA, Brasília, pp 21–28
- R Development Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- Ratter JA, Bridgewater S, Ribeiro JF (2003) Analysis of the floristic composition of the Brazilian cerrado vegetation III: comparison of the woody vegetation of 376 areas. *Edinburgh J Botany* 60(1):57. <https://doi.org/10.1017/S0960428603000064>
- Ribeiro JF, Walter BMT, Mendonça RC et al (2020) Caracterização florística e potencial de uso das espécies vasculares ocorrentes na região da Trijunção. In: Brandão RA et al (orgs) História Natural do Sertão da Trijunção do Nordeste, Centro-oeste e Sudeste do Brasil. Per Se, São Paulo, pp 55–87
- Sano EE, Rodrigues AA, Martins ES et al (2019) Cerrado ecoregions: a spatial framework to assess and prioritize Brazilian savanna environmental diversity for conservation. *J Environ Manage* 232:818–828. <https://doi.org/10.1016/j.jenvman.2018.11.108>
- Santana OA, Carvalho Júnior OAD, Gomes RAT et al (2010) Distribuição de espécies vegetais nativas em distintos macroambientes na região do oeste da Bahia. *Espaço e Geografia* 13(2):181–223
- Santos HG, Jacomine PKT, Anjos LHC et al (2018). Sistema brasileiro de classificação de solos, 5th edn. Embrapa, Brasília
- Schmidt IB, Ticktin T (2012) When lessons from population models and local ecological knowledge coincide—Effects of flower stalk harvesting in the Brazilian savanna. *Biol Cons* 152:187–195. <https://doi.org/10.1016/j.biocon.2012.03.018>
- Silva-Moraes HG, Cordeiro I, Figueiredo N (2019) Flora and floristic affinities of the Cerrados of Maranhão State, Brazil. *Edinburgh J Botany* 76(1):1–21. <https://doi.org/10.1017/S0960428618000215>
- Simon MF, Proença C (2000) Phytogeographic patterns of *Mimosa* (Mimosoideae, Leguminosae) in the Cerrado biome of Brazil: an indicator genus of high-altitude centers of endemism? *Biol Cons* 96(3):279–296. [https://doi.org/10.1016/S0006-3207\(00\)00085-9](https://doi.org/10.1016/S0006-3207(00)00085-9)
- Sousa-Baena MS, Garcia LC, Peterson AT (2014) Knowledge behind conservation status decisions: data basis for “Data Deficient” Brazilian plant species. *Biol Cons* 173:80–89. <https://doi.org/10.1016/j.biocon.2013.06.034>
- SpeciesLink (2021). Available at: <http://specieslink.net>. Accessed 10 Sept 2021
- Strassburg BB, Brooks T, Feltran-Barbieri R et al (2017) Moment of truth for the cerrado hotspot. *Nat Ecol & Evol* 1(4):1–3. <https://doi.org/10.1038/s41559-017-0099>
- Tjørve E (2010) How to resolve the SLOSS debate: lessons from species-diversity models. *J Theor Biol* 264(2):604–612. <https://doi.org/10.1016/j.jtbi.2010.02.009>
- Velazco SJE, Villalobos F, Galvão F, De Marco JP (2019) A dark scenario for Cerrado plant species: effects of future climate, land use and protected areas ineffectiveness. *Divers Distrib* 25(4):660–673. <https://doi.org/10.1111/ddi.12886>

- Vidal JDD, Souza AP, Koch I (2019) Impacts of landscape composition marginality of distribution soil fertility and climatic stability on the patterns of woody plant endemism in the Cerrado. *Glob Ecol Biogeogr* 28(7):904–916. <https://doi.org/10.1111/geb.12901>
- Vieira LT, Castro AA, Coutinho JM et al (2019) A biogeographic and evolutionary analysis of the flora of the North-eastern cerrado, Brazil. *Plant Ecol Divers* 2(5):475–488. <https://doi.org/10.1080/17550874.2019.1649311>
- Wilson MC, Chen XY, Corlett RT et al (2016) Habitat fragmentation and biodiversity conservation: key findings and future challenges. *Landscape Ecol* 31:219–227. <https://doi.org/10.1007/s10980-015-0312-3>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.