Overview of the horseflies (Diptera: Tabanidae) of South Africa: assessment of major collections for spatiotemporal analysis

Louwrens P Snyman[1,4*], Luis Neves[1], Laetitia Lempereur[2] and Albé C Bosman[3]

[1] Vectors and Vector-Borne Diseases Research Programme, Department of Veterinary Tropical Diseases, Faculty of Veterinary Sciences, University of Pretoria, Onderstepoort 0110, South Africa.

[2] Laboratory of Parasitology and Parasitic Diseases, Centre for Fundamental and Applied Research for Animal Health (FARAH), Faculty of Veterinary Medicine, University of Liège, Liège 4000, Belgium.

[3] Department Collections and Digitisation: Marine Biodiversity collections, Iziko Museums of Cape Town, Cape Town 8000, South Africa. 4 Durban Natural Science Museum, Durban 4025, South Africa.

*Correspondence to: lokisnyman@gmail.com

Abstract

Tabanidae, commonly known as horseflies, is a large, cosmopolitan family with approximately 4400 described species, of which about 400 occur in southern Africa. Both sexes feed on nectar or pollen and some species are important pollinators. Females usually require a bloodmeal to produce eggs, implicating them as vectors of disease-causing agents. Despite their importance, they have been largely neglected by science, especially in the Afrotropics. In this study, we aimed to elucidate some fundamental spatiotemporal aspects of horsefly biology in South Africa by centralising digitised data from the major natural history museums in the region. Approximately 6 000 records have been digitised and collated into a Specify database from an approximate 17 000 specimens housed in 10 museums in the region. The collections from the 10 institutions, including six collections in neighbouring countries are briefly discussed. Spatiotemporal analysis according to biomes (Albany thicket, desert, forests, fynbos, grassland, Indian Ocean belt, Kalahari savannah sub-biome, Central Bushveld sub-biome, Lowveld savannah sub-biome, Nama Karoo and Succulent karoo) revealed baseline information regarding their biology within the biomes. The majority of the records are from north eastern KwaZulu-Natal, followed by Kruger National Park and areas surrounding Cape Town, indicative of uneven sampling. The database is represented by a total of 195 species of which 83 species were recorded in only one biome, with grassland home to 28% of the biome endemic species. A species list to the region, supplemented by records from previous authors, totals to 203 species occurring in the region. Future studies should aim to digitise the remaining specimens in order to create species pages with brief descriptions, high-resolution photographs, distribution maps and spatiotemporal patterns available online.

Keywords: biome; digitisation; disease vector; insect pollination; museum collection; Specify

Introduction

Tabanidae, commonly known as horseflies, deerflies or clegs, is a large family of lower brachyceran flies comprising approximately 4400 species distributed worldwide (Baldacchino *et al*. 2014). The Afrotropical region is not particularly rich in species, with ca 800 species recorded from the region (Chainey 2017) of which about 400 occur in southern Africa (Taioe *et al*. 2017).

Adults of both sexes rely on nectar and pollen as their primary food source, and some species, especially members of the Pangoniinae, have been implicated as important pollinators (Johnson & Steiner 1995, 1997, 2000; Karolyi *et al*. 2014). Females of most species require a bloodmeal for egg development and are equipped with mouthparts modified to perform both tasks (Karolyi *et al*. 2014). Blood meals are usually taken from mammals, but birds, reptiles and amphibians have also been recorded as hosts (Chainey & Oldroyd 1980). Female tabanids are effective vectors of pathogens and readily transmit viruses, bacteria and protozoans mechanically and act as biological vectors of some filarian nematodes and *Trypanosoma theileri* (Baldacchino *et al*. 2014; Mulandane *et al*. 2020).

Despite their importance as pollinators and threat to human and animal health, they have been neglected subjects of science, especially in southern Africa, with the majority of studies from the region taxonomic in nature (Oldroyd 1952, 1954, 1957; Usher 1965, 1972; Chainey 1983, 1987; Morita 2008). Studies that have had human and animal health as a focus commonly report on other known vectors (biting midges, mosquitoes, ticks and tsetse flies), obscuring the role tabanids might play as disease vectors.

Despite the limited knowledge on the role of tabanids in pollination and disease transmission, substantial horsefly collections are housed by several institutes in the region. Traditionally, these collections were mainly used by taxonomists to classify and elucidate the relationships among taxa. With the advent of digital data sharing, these collections play an increasingly important role in fields in addition to taxonomy, with invaluable contributions to disease ecology, invasive biology, climate change, pollination, biodiversity studies, conservation and land use management (Suarez & Tsutsui 2004; Baird 2010; Smith & Blagoderov 2012). Despite the increased interest in natural history collections, funding for collection houses (museums, universities, research institutes etc.) has ironically reduced, hampering the already mammoth technical and organisational challenges that accompany the curation and digitisation of collections.

This study aims to assess the major tabanid collections in the region by reporting on the number of specimens, progress of digitisation, biodiversity and gap analysis and to initiate a centralised Specify database to act as a foundation for future research on tabanids.

Materials and Methods

Data sources

The collection details were compiled by contacting the institutions with relevant collections listed by Kirk-Spriggs (2017) in the Manual of Afrotropical Diptera with the addition of Rekomitjie Research Station (Zimbabwe), Skukuza Biological Reference Collection (South Africa) and Travassos Santos Dias Collection (Mozambique). Data from the Natural History

Museum, London (NHMUK) supplemented the regional data due to the readily available digitised data as well as the large number of types housed in the museum.

The collections reported on in this study is housed at the following institutions. Names and codens of institutions are based on Evenhuis (2007), asterisk (*) denotes collections not listed in Evenhuis (2007):

South Africa (RSA):

- Albany museum, Makhanda (before 2018 known as Grahamstown) (AMGS)
- Ditsong Museum, Pretoria (TMSA)
- Durban Natural Science Museum (DMSA)
- Iziko Museum, Cape Town (SAMC)
- KwaZulu-Natal Museum, Pietermaritzburg (NMSA)
- National Museum, Bloemfontein (BMSA)
- Rhodes University Collection, Makhanda (RUDZ)
- Skukuza Biological Reference Collection, Kruger National Park (SKNP)*
- South African National Collection of Insects, ARC Biosystematics Institute, Pretoria (SANC)
- Stellenbosch University, Stellenbosch (SUEC)*

Neighbouring countries (NC):

- Rekomitjie Research Station, Zimbabwe (RRS)*
- Natural History Museum of Zimbabwe, Bulawayo (Rhodesia Museum, National Museum of Rhodesia) (NMBZ)
- Botswana National Museum, Gaborone (MAGB)
- Museu de História Natural, Maputo (Museu Dr Álvaro de Castro) (CPMM)
- National Museum of Namibia, Windhoek (Landesmuseum; South West Africa Museum; State Museum) (NMNW)
- Plant Protection Research Institute, Harare (Department of Agriculture, Salisbury) (PPRZ)
- Travassos Santos Dias Collection, Maputo (Eduardo Mondlane University, Mozambique) (TSDC)*

Data preparation

All nomenclature follows that of Tabanid PEET Project and the Smithsonian Entomology Diptera Diversity Digitization Team (http://tabanidae.myspecies.info/). The taxonomy of the subfamilies Adersiinae and Scepsidinae is still largely unresolved. Both subfamilies are still recognised with Adersiinae comprising a single tribe and genus, Adersiini and *Adersia*. Scepsidinae in turn comprise the tribe Scepsidini and the genera *Braunsiomyia* and *Scepsis*. Morita *et al*. (2016) suggested that the members of both subfamilies probably form a single tribe, Scepsidini, within Chrysopsinae. In the analysis presented here, all members are treated as Scepsidini for the sake of brevity and not to suggest any taxonomic changes. The supplementary species list (Table S1) treats the species present in South Africa as it is currently classified on the PEET Project website (http://tabanidae.myspecies.info/).

The majority of records obtained for this study were not digitised for this project, but previously digitised by respective institutions, most of which used Darwin Core Standard, and could thus be combined. The combined data were then georeferenced in GEOLocate (http://www.geo-locate.org/) and cleaned using batch processing in Open Refine (http://openrefine.org/). All records lacking identification, dates or localities were excluded from analysis. The taxonomy was manually edited to conform to the current classification as prescribed by the Tabanid PEET Project website (http://tabanidae.myspecies.info/). All records that remained were subject to analysis and are termed fixed records. All fixed records were manually reviewed for suspicious records (e.g. suspicious species distribution or identification). All suspicious records were either validated by contacting the relevant housing institution for additional data (e.g. taxon determiner) or removed. Fixed records were either sorted into number of records per taxon (subfamily/tribe) per biome, to represent relative abundance or collated into a single taxon record for relative diversity per biome. It is important to note that fixed records are not equal to number of specimens as some institutes will assign a single unique identifier to a series of specimens. To address this issue, relevant institutions were duly contacted if unique/suspicious records were present and all feedback are present in the supporting information.

Biomes and graphs

In order to describe the composition of species and records within a biome, the number of tribe-specific records/species was divided by the total number records/species of the family within the biome, yielding a taxon-specific proportionate composition (biome composition). In order to compare record/species composition across biomes, the number of tribe-specific records/species within a biome was divided by the total number of tribe specific records/species across all biomes, yielding a taxon-representative proportion in a biome (taxon representation). In order to assess the east-interior-west composition and abundance of the three dominant subfamilies across the region, both species and record data were collated by adding all data from the eastern biomes (Central savannah, Lowveld savannah, Indian Ocean coastal belt and grassland), the central or interior biomes (Albany thicket, Nama Karoo and Kalahari savannah) and the western winter rainfall biomes (fynbos, desert and Succulent Karoo). Forests were omitted due to the presence of this biome within several other biomes throughout the region. Plots with the collated proportionate records and species were generated for the three subsections.

A species was considered as endemic if all records of the species were collected in only one biome. Endemism rates were calculated as the proportion of the total number of species present in the biome vs. the number of biome-endemic species. All collection facilities were contacted if an endemic species was only recorded once in a biome to evaluate the nature of the record (i.e. single specimen, series of specimens, number of collection trips, identification determiner) and evaluated on a case by case basis.

We analysed all records in accordance with the biomes as currently delineated by the South African National Parks (http://bgis.sanbi.org/SpatialDataset/Detail/484). Due to the extreme different rainfall, elevation and temperature fluctuations of the Savannah biome in South Africa, the Kalahari (Duneveld and Eastern Kalahari bioregions), the Central Bushveld savannah bioregion and the Lowveld savannah (Mopane bushveld, Lowveld and sub-escarpment lowveld bioregions) were kept separate as ‘sub-biomes’ (Fig. 1). Lesotho was included as part of the analysis due to the difficulty of removing the land-locked country from the layers used in the GIS analysis. Biome-endemic species were categorised as such if a species was only recorded in a single biome. Mean precipitation and temperature were extracted from Kriticos *et al*. (2012). The species checklist was compared with the list compiled by Usher (1972). The graphs were constructed in R studio (Team 2015) using the base packages (Team 2013), ggplot2 (Wickham 2011) and cowplot (Wilke 2016). Azonal vegetation is scattered throughout South Africa and characterised as not belonging to any specific biome and thus excluded from all graphs.

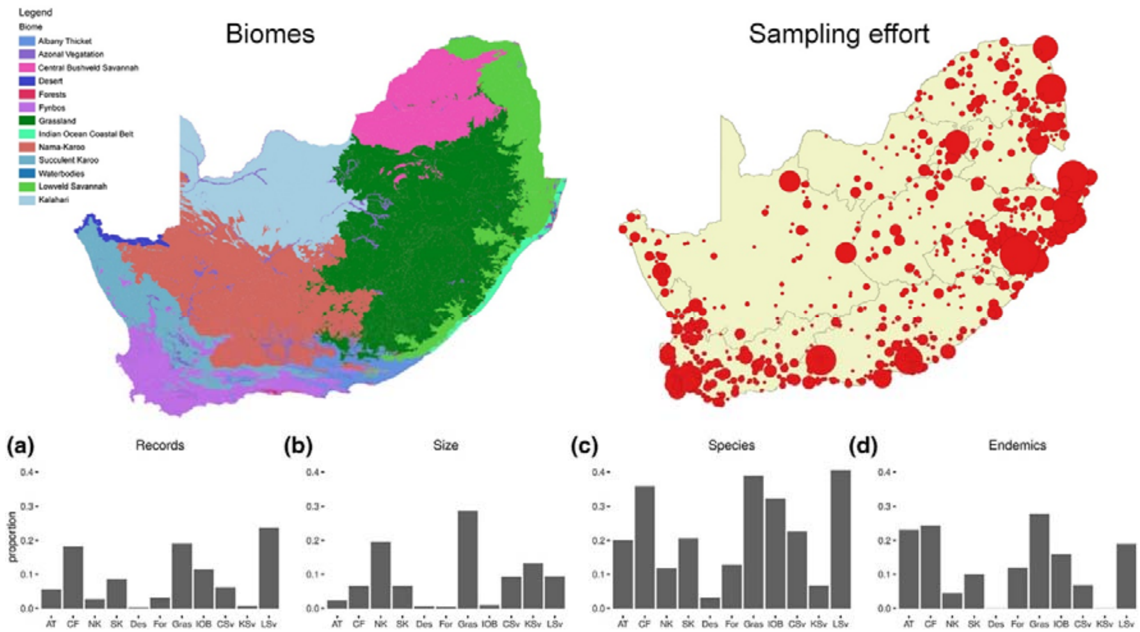

Fig. 1. Top left, South African biomes as delineated by South African National Parks. The savannah biome was separated into the Kalahari savannah (Eastern Kalahari Bushveld savannah and Kalahari Duneveld savannah bioregions), the Central Bushveld savannah bioregion and the Lowveld (Mopane, lowveld and subescarpment lowveld bioregions). Top right, graphic representations of the collections housed in South African museums. Size of the circle represents sampling effort. Bottom left to right, proportions of records, biome size, species and endemic species across biomes. AT, Albany thicket; CB, Central Bushveld savannah; CF, fynbos; D, desert; F, forest; G, grassland; IB, Indian Ocean belt; K, Kalahari savannah; L, Lowveld savannah; NK, Nama Karoo; SK: Succulent Karoo.

Results

Collections

Approximately 34 000 Tabanidae specimens are formally housed in collection institutes throughout southern Africa. The Travassos Santos Dias Collection (±14 500 specimens) (FC Mulandane pers. comm. 2019) and the NMSA collection (±12 000 specimens) comprise over 75% of Tabanidae in all the collection facilities. None of the collections outside South Africa, apart from the small Rekomitjie Research Station collection, have been digitised and therefore could not be used here. Nearly 35% of the South African specimens have been digitised and were centralised into a single Specify database (http://www.sustain.specifysoftware.org) (Table 1). The database is currently privately maintained and managed by the authors of this paper.

Table 1. Summary of the natural history collections of Tabanidae in South Africa

	% digitised	% ID	Total collection	Notes
AMGS	1.72	72.30	989	
BMSA	58.95	58.95	894	
TMSA	46.07	46.07	280	
DMSA	100.00	75.00	323	
SAMC	55.30	55.30	1566	
NMSA	31.47	75.00	12000†	
RUDZ	0.00	50.00	160†	
SANC	23.50	23.50	634	
SKNP	100.00	100.00	193	
SUEC	0.00	58.42	594	98 species
Subtotal (RSA)	33.83	61.5	17 633	
RRS	100.00	100.00	69	
MAGB	0	?	500†	±500 Diptera
CPMM	0	?	2000‡	>50% *T. taeniola*
NMBZ	0	100.00	1430†	
NMNW	0	?	?	
PPRZ	0	?	454‡	73 species
TSDC	0	100.00	14500†	
Subtotal (NC)	0.36	7.91	18 953.00	
Grand totals	16.48	37.76	36 564.00	

South Africa (RSA). Neighbouring countries include Botswana, Mozambique, Namibia and Zimbabwe (NC).

† Estimates.

‡ Figures taken from Kirk-Spriggs (2017).

The Tabanidae collections resources of South Africa and in neighbouring countries are summarised as follows:

Albany Museum, Makhanda (=Grahamstown) (AMGS): The Albany Museum collection comprise five drawers of which one is filled with unsorted/unidentified material. The museum mostly focusses on Hymenoptera and has not been able to digitise the Diptera collection. Specify is used as data management software for the collection. An unknown number of additional Tabanidae specimens are in the predatory wasp collection, which were collected as prey items of those wasps (T Bellingan pers. comm. 2018).

National Museum, Bloemfontein (BMSA): The collection at BMSA comprise a fully digitised subcollection (n = 469), mostly relatively recent series collected by A Kirk-Spriggs and colleagues. This subsection is only identified to family level. The second, unsorted, section (n = 425) of the collection has been donated to the museum by the University of Pretoria and is unsorted, undigitised and in poor condition. This section is only in part identified by the students as part of their studies. Specify is used as data management software for the collection. The catalogued section of the collection is mostly representative of the Afrotropics beyond South Africa. The collection is currently undergoing identification and digitisation by the lead author and will subsequently be incorporated into the Specify database of the institution.

Ditsong Museum, Pretoria (TMSA): The Ditsong collection, formerly the Transvaal Museum, is housed in plastic drawers covered by glass. Approximately half of the collection comprise specimens from the University of Pretoria student collections, often erroneously identified. The specimens, however, are in good condition. The collection is currently undergoing identification and digitisation by the lead author and will subsequently be incorporated into the Specify database of the institution.

Durban Natural Science Museum (DMSA): The collection is a small collection housed in wooden drawers of which 75% is identified. All the specimens are digitised on a Specify database (N Govender pers. comm. 2019).

KwaZulu-Natal Museum, Pietermaritzburg (NMSA): This is by far the most comprehensive collection of Tabanidae in southern Africa, housing approximately 70% and 60% of the Tabanidae specimens in South Africa and southern Africa, respectively. Specify is used as data management software for the collection (K Williams pers. comm. 2018). The comprehensive collection is due to the studies of Oldroyd and the continuation by Pamela Stuckenberg (neé Usher) at the institution as well as the Fritz Zumpt collection (Kirk-Spriggs 2017). The Usher (1972) review was also solely based on selected specimens housed in this collection.

Rhodes University Collection, Makhanda (=Grahamstown) (RUDZ): The collection at Rhodes University comprises approximately 160 specimens of which only the *Philoliche* species, approximately half of the collection, are identified. The entire collection is uncatalogued and unidentified due to the focus on flies relevant to forensic science. Digitisation is ongoing (MH Villet pers. comm. 2018). Specify is used as data management software for the collection.

Iziko Museum, Cape Town (SAMC): The Iziko collection is approximately 80% fully sorted, catalogued and digitised with all records fixed. Additionally, approximately 700 specimens are identified to family level but remain unsorted and uncatalogued until the specimens can be identified (A Mayekiso pers. comm. 2018). Specify is used as data management software for the collection.

South African National Collection of Insects, ARC Biosystematics Institute, Pretoria (SANC): All identified specimens, ca 634, in the SANC collection are fully catalogued and digitised. Instead of assigning a catalogue number per specimen, SANC assigns a catalogue number to each series per species which amounts to 224 digitised records. The majority of the collection is awaiting identification and digitisation.

Skukuza collection, Kruger National Park (SKNP): The Skukuza collection is fully digitised and all records are fixed. The collection is made up by two smaller subcollections. The first subcollection (n = 58) was all identified by PJ Usher. The slightly larger second subcollection comprises 135 specimens, mostly collected by LEO Braack and all identified by JE Chainey (LEO Braack pers. comm. 2017). The collection is well kept and under the care of Ms Guinevere Zambatis.

Stellenbosch University, Stellenbosch (SUEC): The Stellenbosch University has quite a species rich collection comprising 347 specimens identified into 98 species, excluding 247 unsorted and unidentified specimens. Of the identified material, approximately 15% are non-

Afrotropical. A detailed list of drawer content is available (P Addison pers. comm. 2018). Currently, there is no dedicated full-time curator.

Rekomitjie research station, Zimbabwe (RRS): Rekomitjie (also sometimes referred to as Rikometjie or Rekometjie) research station has a small collection currently on long-term loan to Laboratory of Parasitology and Parasitic Diseases, University of Liège, Belgium. The entire collection is housed in a single, double-sided wooden box. The majority of the collection was sampled at the research station in northern Zimbabwe with isolated specimens from elsewhere in Zimbabwe and South Africa. All specimens were identified by RJ Phelps and the majority also collected by him. The entire collection has been catalogued and digitised by the authors of this manuscript.

Natural History Museum of Zimbabwe, Bulawayo (formerly Rhodesia Museum, National Museum of Rhodesia) (NMBZ): The collection is the most comprehensive Tabanidae collection in Zimbabwe with about 1400 specimens housed at the institution. The entire collection is identified and manually accessioned (D Madamba pers. comm. 2017). The collection is currently undergoing digitisation by D Madamba.

Botswana National Museum, Gaborone (MAGB): The entire Diptera collection at the institution seems to be unidentified and small (±500 specimens) (BL Rapalai pers. comm. 2018). Even though the institution has some funding support for digitisation, the focus is on Coleoptera and Hymenoptera which have been prioritised. According to Kirk-Spriggs (2017), Tabanidae forms a prominent part of the collection.

Museu de História Natural, Maputo (Museu Dr Álvaro de Castro) (CPMM): Unfortunately, no contact could be made with the institution. Kirk-Spriggs (2017) described the collection as a small collection, ca 2000 specimens with the majority identified to species; however, over half of the specimens represent a single species, *Tabanus taeniola*.

National Museum of Namibia, Windhoek (Landesmuseum; South West Africa Museum; State Museum) (NMNW): Unfortunately, no curator has been appointed since the departure of Kirk-Spriggs in 2003 (Kirk-Spriggs 2017). To make matters worse, the entomology technician has since retired, leaving the collection completely unattended. The collection has a substantial Tabanidae section which was identified by JE Chainey in the early 2000s. The Tabanidae are possibly, in part, digitised (Kirk-Spriggs pers. comm. 2018).

Plant Protection Research Institute, Harare (Department of Agriculture, Salisbury) (PPRZ): Unfortunately, no contact could be made with the institution. From Kirk-Spriggs (2017), the institution house 454 Tabanidae specimens divided into 73 species, probably mostly from Malawi and Zimbabwe.

Travassos Santos Dias Collection, Eduardo Mondlane University, Mozambique (TSDC): This institution houses the important 'Missão de combate às Tripanossoíases' collection, including the type specimens of the species described by JA Travassos Santos Dias. The collection comprises approximately 14 500 specimens with most of the specimens identified by Travassos Santos Dias. The collection is currently not digitised, but housed in drawers in a temperature-controlled room at the Centro de Biotecnología at the Faculty of Veterinary Sciences under curation of FC Mulandane (FC Mulandane pers. comm. 2018).

Digitised records in South Africa: In total, 5958 records from all the collections are databased, of which 4123 records are from South Africa, included dates and localities and could be used for graphic representation. The following tribes were recorded in South Africa: Adersini (Adersinae); Bouvieromyiini, Chrysopsini, Rhinomyzini (Chrysopsinae); Pangoniini, Philolichini, (Pangoniinae); Scepsidini (Scepcidinae); and Diachlorini, Haematopotini and Tabanini (Tabaninae). A total of 195 species were represented by 3280 complete databased records (including identification). The species list compiled by Usher (1972) was mostly based on the records from the Oldroyd (1952, 1954, 1957) monographs. With the addition of species in the Usher (1972) list not present in the database, a total of 203 species are present in South Africa. A total of 83 species were recorded as unique to a biome, and 24 species not recorded by Usher (1972) are recorded as new to the region. A full species list of the region, including species distribution per biome, is provided as supporting information. Vegetation biomes, instead of climatic regions, as was done by Usher (1972), were used for spatiotemporal analysis. Here, vegetation biomes are considered as a delineation basis for spatiotemporal assessment. Biomes are described according to dominant plant assemblages, but generally conforms to climatic and geological patters which influence the formation of plant assemblages (Mucina and Rutherford 2006). South Africa is currently divided into nine biomes, with scattered water bodies and Azonal vegetation throughout the region. In alphabetical order, the western part of South Africa is occupied by (1) Albany thicket, (2) fynbos, (3) Succulent Karoo and (4) Nama Karoo. A small section in the north-western part of the country is formally recognised as (5) desert. The desert biome is discussed last of the western biomes due to the limited information we have of the biome. (6) Forests are scattered along the escarpment from the southwestern coast up to the Mozambique and Zimbabwe borders in the northeastern part of the country. Due to the small size of many forest patches in South Africa and the low resolution in the GIS analysis due to nonspecific historical labels, species might not have been accurately assigned to this biome. The eastern part of South Africa is further alphabetically divided into (7) grasslands, (8) Indian Ocean Belt and the savannah. Because of the large geographical range and climatic differences, the savannah biome is discussed as the (9) Central savannah, (10) Kalahari savannah and (11) Lowveld savannah. The Tabanidae found in the biomes are discussed below in this numerical order.

1. Albany thicket: The biome is represented by 5.6% of the records and occupies 2.30% of the area of South Africa. Of all species recorded in South Africa, 20% of the species are recorded from the biome representing 23% of the biome-endemic species (Fig. 1). There are 39 species present in the region, of which nine species are uniquely recorded in the region (endemic) yielding a 23% endemism rate (Table 2). This is an important biome for Scepsidini, with all South African species present in the biome, as well as nearly 40% of the Scepsidinae records originating from biome (Fig. 2). The composition of the biome is skewed towards Philolichini, which accounts for 47% of the records and 36% of species belonging to the tribe (Fig. 3). Tabanidae adults are most active from October to February, with an activity peak in December (n_{records} = 111) (Fig. 3).
2. Fynbos: The biome is represented by 18% of the records and occupies 6.5% of the area of South Africa. Of all species recorded in South Africa, 35% occur in the biome, representing 24% of biome-endemic species (Fig. 1). There are 70 species present in the region, of which 17 species are endemic, yielding a 24% endemism rate (Table 2). Seventy percent of Pangoniini records from the region occur in the biome, and one of the two Pangoniini species occur in the biome (Fig. 2). Over a third of the species of Philolichini (57%), Pangoniini (50%), Scepsidini (50%), Diachlorini (41%) and

Bouvieromyiini (40%) are present in the biome (Fig. 2). Philolichini accounts for the most often sampled and species-rich tribe, accounting for 47% of the records and 37% of the species (Fig. 3). Tabanidae adults are active in this biome from September to December, with a peak in October ($n_{\text{records}} = 554$) (Fig. 4).

3. Succulent Karoo: The biome is represented by 8.6% of the records and occupies 6.6% of the area of South Africa. Of all species recorded in South Africa, 2% occur in the biome, representing approximately 1% of endemic species (Fig. 1). There are 40 species present in the region, of which four species are endemic, yielding a 10% endemism rate (Table 2). A notable proportion of Philolichini (35%), Bouvieromyiini (29%) and Scepsidini (25%) species is present in the biome (Fig. 2). A prominent representation of the Bouvieromyiini (29%) and Diachlorini (23%) records also originate from the biome (Fig. 2). Similar to the fynbos composition, Philolichini accounts for the most often sampled and species-rich tribe accounting for 49% of the records and 40% of the species (Fig. 3). September and October seems to be the peak activity period for adult tabanids in this biome ($n_{\text{records}} = 324$) (Fig. 4).
4. Nama Karoo: The biome is represented by 2.7% of the records and occupies 19.59% of the area of South Africa. Of all species recorded in South Africa, 1.2% occurs in the biome representing 0.4% of endemic species (Fig. 1). There are 23 species present in the region, of which one species, *Rhiglioglossa redunda* Oldroyd, 1957, is endemic, yielding a 4% endemism rate (Table 2). A noteworthy proportion of the species recorded from the biome are Bouvieromyiini (22%), despite the tribe only accounting for 3.8% of the records. Comparatively few records are from this biome despite strong species representation of Rhinomyzini (18%), Philolichini (15%), Bouvieromyiini (14%) and Haematopotini (14%) (Fig. 2). The Nama Karoo is dominated by Philolichini, which accounts for 33% of the records and 30% of the species (Fig. 3). Nearly 20% of the records from this biome are not yet identified. Adults are active from October to March, with a slightly higher number of specimens collected in November ($n_{\text{records}} = 123$) (Fig. 3).
5. Desert: The biome is represented by 0.02% of the records and occupies 0.57% of the area of South Africa. Of all species recorded in South Africa, 3% of the species are recorded from the biome (eight species) with no endemic species (Fig. 1) (Table 2). In comparison with other biomes, the taxon representation in this biome is unremarkable (Fig. 2). The desert biome is dominated by Philolichini, with 56% and 67% of the records and species collected, respectively, belonging to the tribe (Fig. 3). Bouvieromyiini (22% of records, 17% of species) and Tabanini (11% of records, 17% of species) are also worth mentioning as is the 11% of records that are unidentified (Fig. 3). Only 11 fixed specimen records are available for this biome, with nine of them collected in October (Fig. 4)
6. Forest: The biome is represented by 3% of the records and occupies 0.41% of the area of South Africa. Of all species recorded in South Africa, 1.3% occur in the biome, representing 1.2% of endemic species (Fig. 1). There are 25 species present in the region, of which three species are endemic, yielding a 12% endemism rate (Table 2). The taxon representation of this biome is unremarkable, probably due to the lack of accurate coordinates. One of the two Pangoniini species of South Africa is present in this biome (Fig. 2). Most of the records from the biome are Rhinomyzini (39%), followed by Philolichini (18%), unidentified records (16%) and Haematopotini (13%) (Fig. 3). In turn, species proportions favour Philolichini (32%), Tabanini (16%) and Bouvieromyiini and Haemotopotini comprising 12% of the species each (Fig. 3). Most specimens were collected from October to February, with a peak in January ($n_{\text{records}} = 130$) (Fig. 5).

7. Grassland: The biome is represented by 19% of the records and occupies 28.6% of the area of South Africa. Of all species recorded in South Africa, 39% occur in the biome, representing 28% of endemic species (Fig. 1). There are 76 species present in the region, of which 21 species are endemic, yielding a 28% endemism rate, the highest in the region (Table 2). Grassland is seemingly an important biome for Tabaninae, with 59% of Haematopotini and 54% of Tabanini species present in the biome, despite the low proportion of taxon records from the biome (25% and 16%, respectively) (Fig. 2). Other noteworthy tribes well represented in the biome are Rhinomyzini with 33% of the tribe records and 36% species, Chrysopsini with 29% of records and 30% of species, Bouvieromyiini with only 8% of the records but 29% of species and lastly, Philolichini with 16% of the taxon records and 35% of Philolichini species (Fig. 2). Philolichini accounts for the highest proportion of records at 23%, followed by Haematopotini (21%), Rhinomyzini (20%) and Tabanini (16%). Despite the comparatively low records of Tabanini, Tabanini accounts for the highest proportion of species (29%), followed by Haematopotini (22%) and Philolichini at 21% (Fig. 3.). December and January seem to be the months where adult tabanids are most often collected (n_{records} = 622) (Fig. 5).
8. Indian Ocean Coastal Belt: The biome is represented by 11.4% of the records and occupies 0.9% of the area of South Africa. Of all species recorded in South Africa, 32% occur in the biome, representing 16% of endemic species (Fig. 1). There are 63 species present in the region, of which 10 species are endemic, yielding a 16% endemism rate (Table 2). Scepsidini (75%), Rhinomyzini (55%), Chrysopsini (50%) and Tabanini (49%) are species wise well represented in the biome (Fig. 2). Within the biome, both records (34%) and species richness (32%) are dominated by Tabanini (Fig. 3). Tabanid adults are present from October to February, with a slight peak during November to January (n_{records} = 304) (Fig. 5).
9. Central Bushveld savannah: The region is represented by 6.48% of the records and occupies 9.34% of the area of South Africa. Of all species recorded in South Africa, 23% occur in the region, representing 7% of the endemic species (Fig. 1). There are 44 species present in the region, of which three species are endemic, yielding a 7% endemism rate (Table 2). The Chrysopsini is well represented in the biome, with 50% of Chrysopsini species recorded in the biome, despite the comparatively low proportion of records, represented by just over 10% (Fig. 2). This is, however, not reflected in the composition of the biome, with Tabanini and Philolichini better represented by both records (26% and 23%, respectively) and species richness (36% and 25%, respectively) (Fig. 3). Tabanidae adults seem to be active in the biome from September to March, with slightly higher activity during December and January (n_{records} = 217) (Fig. 5).
10. Kalahari savannah: The region is represented by 0.7% of the records and occupies 13.17% of the area of South Africa. Of all species recorded in South Africa, 6.7% occur in the region (Fig. 1). There are 13 species present in the region, of which no species are endemic (Table 2). This biome is poorly representative of both records and species from the region (Fig. 2). The biome composition is dominated by Philolichini records (42%) and species (46%) (Fig. 3). An unclear adult activity pattern is apparent in the biome, with a peak in September followed by relatively uniform activity during October and November, January and February and again, April and May (n_{records} = 93) (Fig. 5).
11. Lowveld savannah: The region is represented by 24% of the records and occupies 9.39% of the area of South Africa. Of all species recorded in South Africa, 41% occur in the region, representing 19% of endemic species (Fig. 1). There are 79 species

present in the region, of which 15 species are endemic, yielding a 19% endemism rate (Table 2). Tabanini is best represented in the biome, with 42% of the records and 66% of the species originating from the biome (Fig. 2). Several other tribes are also well represented with 55% of the Rhinomyzini species present in the biome as well as more than 30% of the Haematopotini, Chrysopsini, Diachlorini and Philolichini species present in the biome (Fig. 2). Within the biome, however, both records (35%) and species (34%) are dominated by Tabanini, with only Philolichini and Haematopotini other noteworthy compositional tribes. November to January seems to be the peak activiy period for adult Tabanidae in the region, with moderate activity during October and February to April (n_{records} = 721) (Fig. 5).

Table 2. Presence of species and unique species (endemic) per biome in South Africa

South African biomes	**Number of species (presence)**	**Number of endemic species**	**Percentage of endemism (%)**
Albany thicket	39	9	23
Fynbos	70	17	24
Nama Karoo	23	1	4
Succulent Karoo	40	4	10
Desert	6	0	0
Forest	25	3	12
Grassland	76	21	28
Indian Ocean CB	63	10	16
Central Bushveld Sv.	44	3	7
Kalahari Sv.	13	0	0
Lowveld Sv.	79	15	19
Total	195	83	42.56

CB, coastal belt; Sv., savannah.

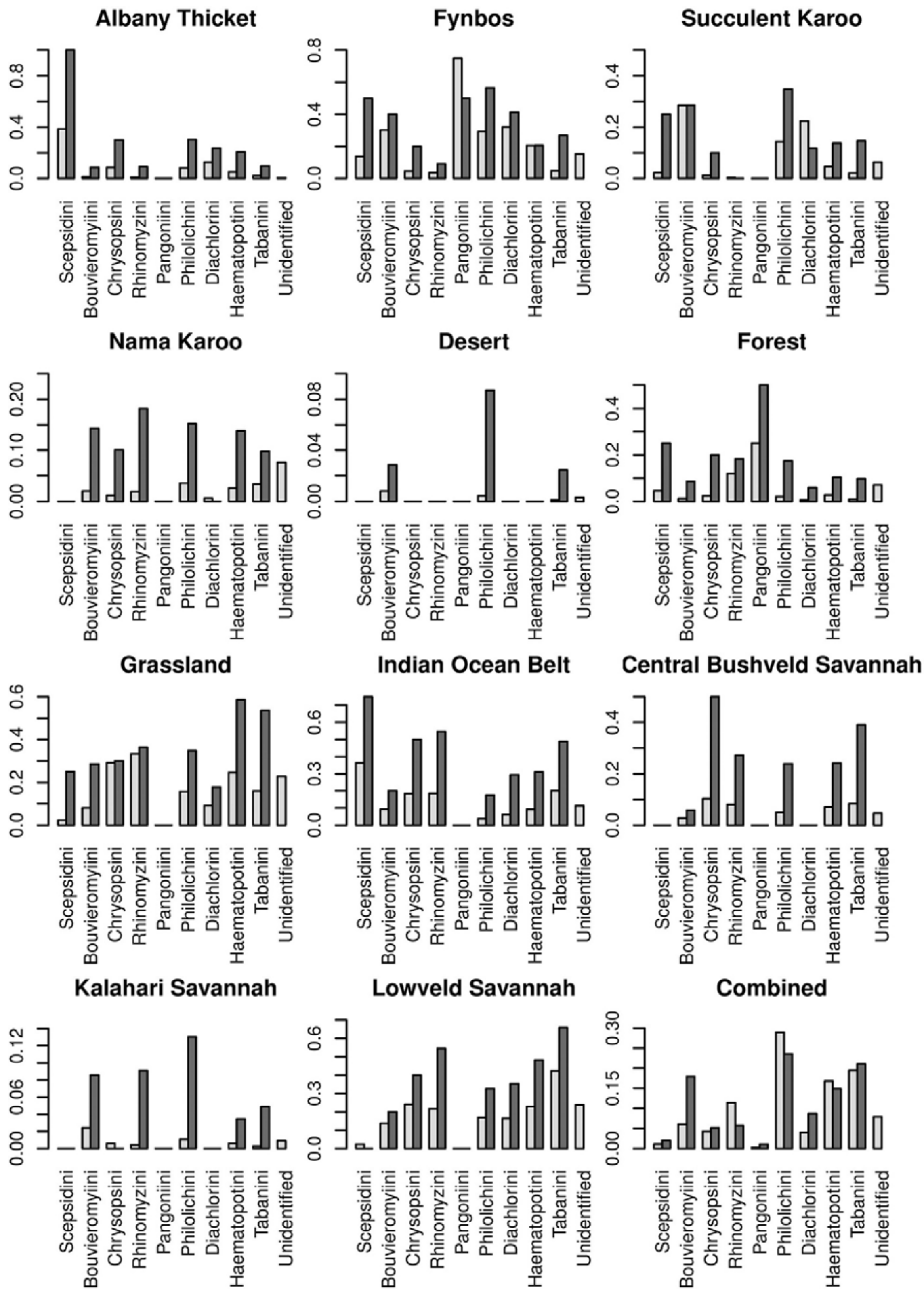

Fig. 2. The taxon representation as proportions of records (abundance) (light grey) and species (richness) (dark grey) per tribe across all biomes.

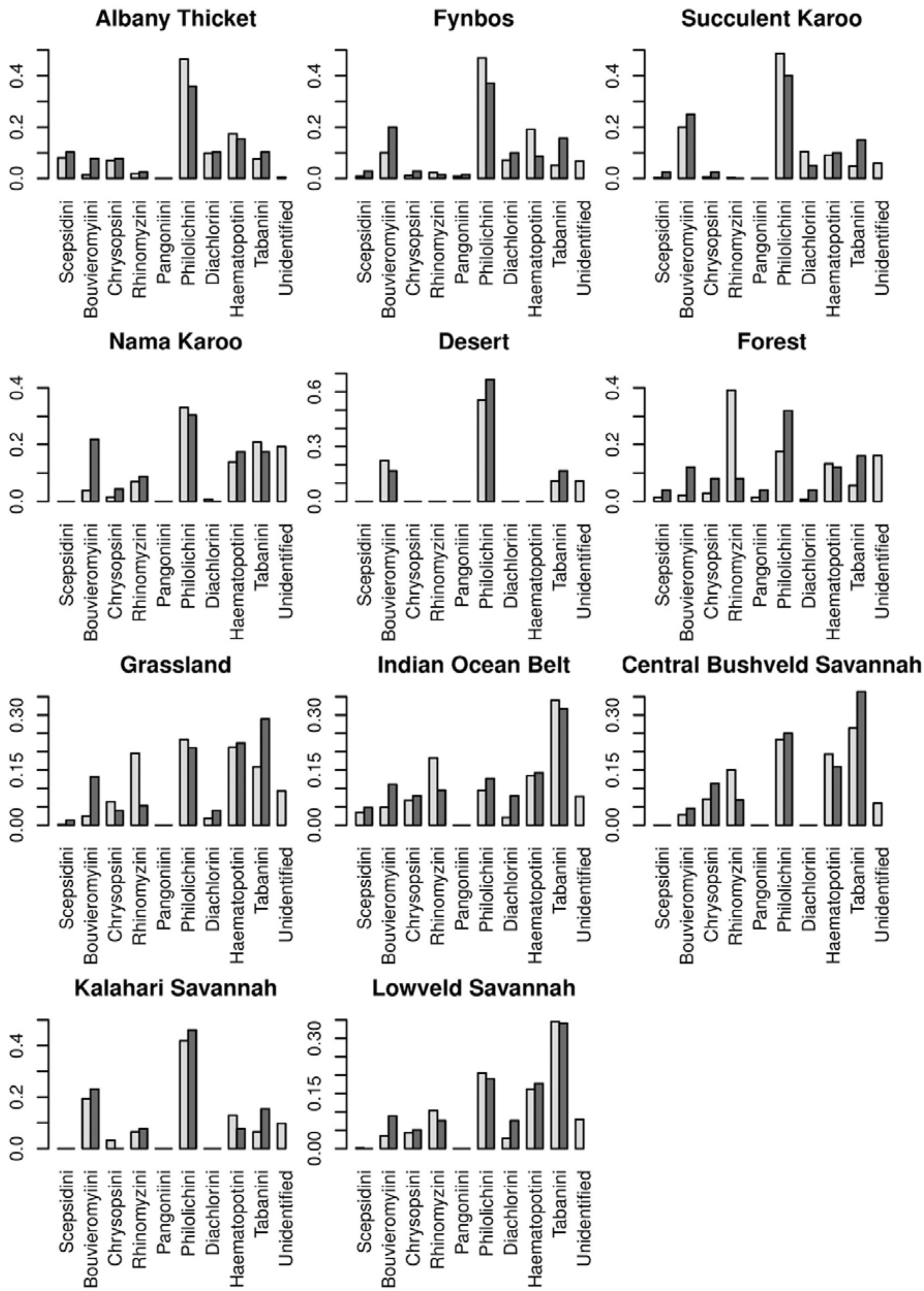

Fig. 3. The biome composition as proportions of records (abundance) (light grey) and species (richness) (dark grey) per tribe within a biome.

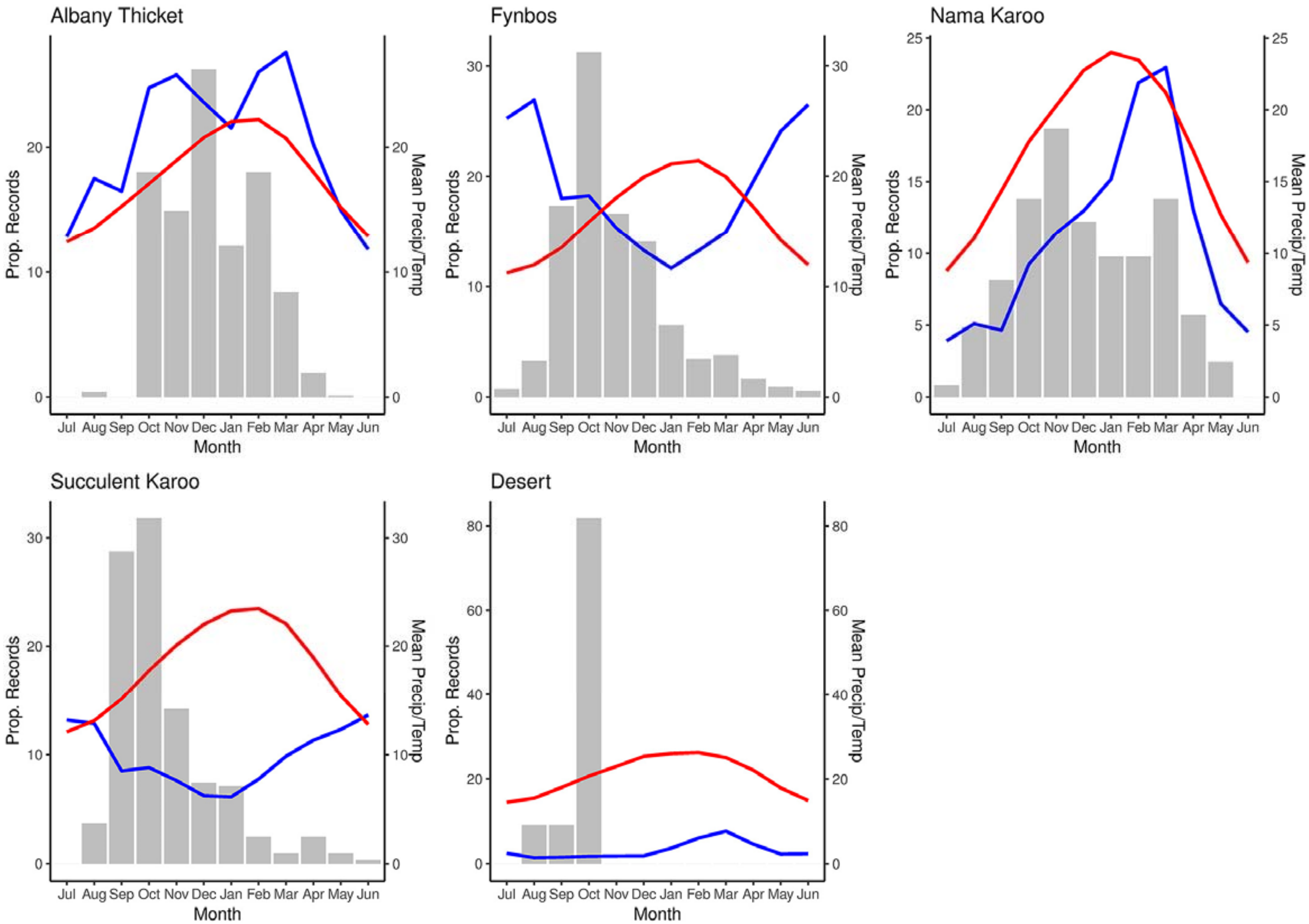

Fig. 4. Proportions of records of Tabanidae specimens collected presented over a year period per biome. The red line indicates mean temperature, and the blue line indicates mean precipitation. Mean precipitation was multiplied by a factor of 0.5 to keep proportions relative.

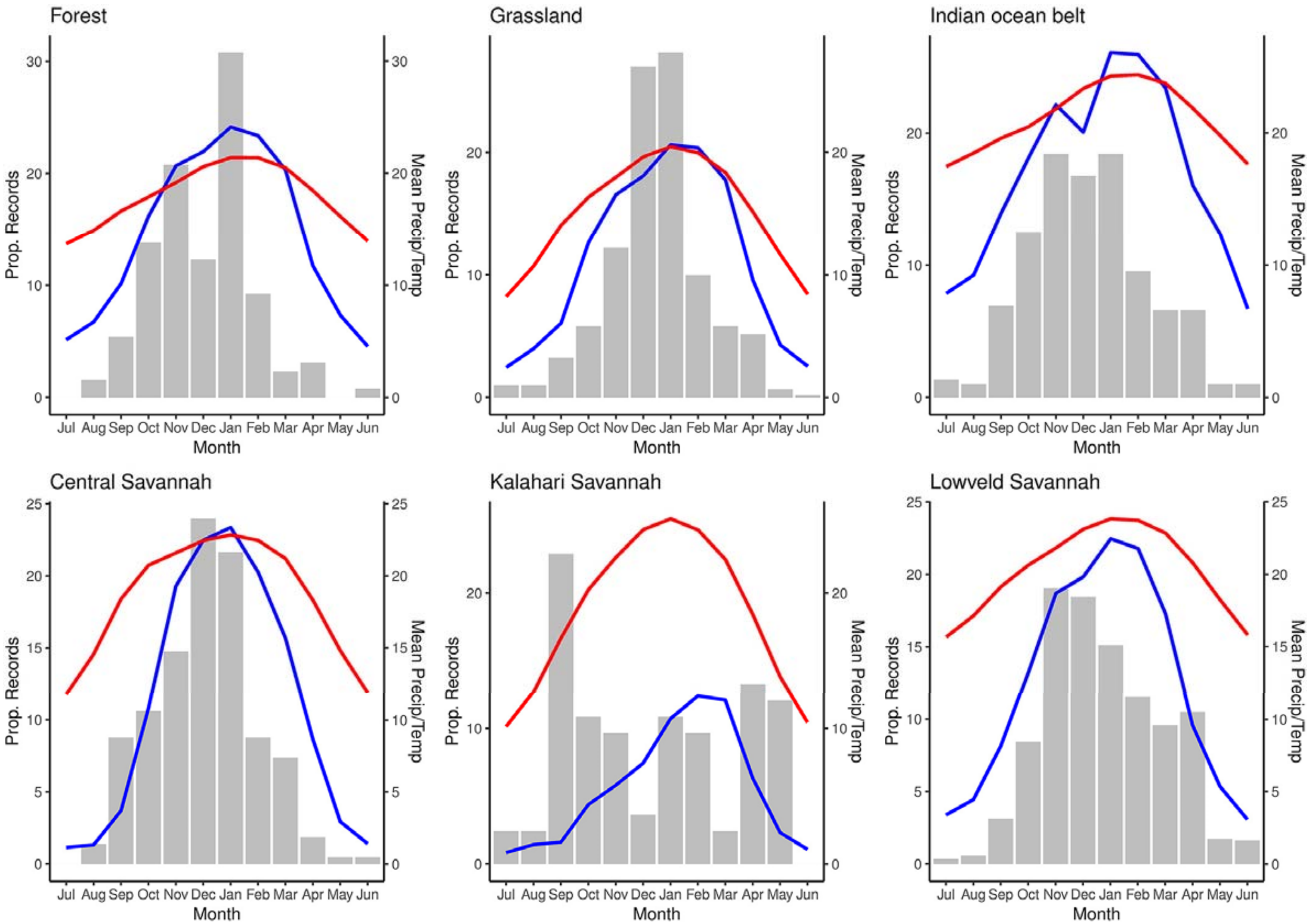

Fig. 5. Proportions of records of Tabanidae specimens collected presented over a year period per biome or savannah region. The red line indicates mean temperature, and the blue line indicates mean precipitation. Mean precipitation was multiplied by a factor of 0.2 to keep proportions relative.

Discussion

South African museums and collection facilities have made substantial advances in digitisation of natural history collections. In neighbouring countries, attempts to digitise collections are ongoing; however, little digitised data are currently available for analysis. From over 5500 records, a collection bias in areas surrounding major cities and popular National Parks (e.g. Kruger National Park) exists. Hull *et al*. (1998) explained, in a study focussing on collections of Buprestidae in South Africa, that Pretoria, Skukuza (KNP) and Rustenburg, for example, are popular collection areas for museum curators from Gauteng, patterns evident in this study as well. Except for Gauteng, collections seem to be focussed on the eastern Mozambican/Swaziland border and coastal areas. This leaves the large inland regions comparatively unexplored. These are important factors when assessing hotspots and coldspots, which might purely reflect sampling bias. The undersampling in these areas has been highlighted before and is not unique to Tabanidae (Hull *et al*. 1998). The Travassos Santos Dias Collection in Maputo is of extreme importance and should be digitised as it might yield substantial insights into tropical Africa as well as Mozambique, a country with a high vector-borne disease burden, some of which are transmitted by Tabanidae (e.g. *Trypanosoma* spp.) (Mulandane *et al*. 2018).

South Africa is home to a rich Tabanidae fauna, with a total of 203 species recorded from the digitised data and the Usher (1972) study. Biome-endemic species account for nearly 43% (83 from 195 species) of the species present in the collated database, highlighting the importance of conservation-based studies assessing the conservation status of these flies and possibly the conservation status of the plants they pollinate. Despite all the effort in cleaning and georeferencing the digitised data, identification mistakes are possible, which might exaggerate or underestimate the actual species number and biome-endemic species rates Usher (1972) focussed mostly on records from the monographs of the late Oldroyd (1952, 1954, 1957). This might in turn have led to an underestimation of species richness in South Africa. Additional disparities between this study and that of Usher include the treatment of *Haematopota spectabilis* Oldroyd 1952 and *H. obscura* Loew, 1858 as synonyms of *H. ocellata*, Wiedeman, 1819 due to the difficulty in identification of the seemingly polymorphic group. Lastly, the seven species are absent from the database but included by Usher (1972) are *Rhigioglossa (Perisilvius) algoensis* (Oldroyd, 1957), *R. (P). cuneata* (Loew, 1858), *Philoliche (Buplex) brincki* Oldroyd, 1957, *Philoliche (Ommatiosteres) nitida* Usher, 1968, *Cydistomyia (Amanella) atra* (Oldroyd, 1954), *Limata facialis* Oldroyd, 1954 and *Tabanus denshamii* Austen, 1908. Reasons for this is unclear since Usher was responsible for identifying the material in the NMSA collection that makes up the majority of the records analysed here and should therefore be present in the data (J Midgley pers. comm. 2019). However, since Usher used additional records form the Oldroyd monographs, these records could be housed in institutions not in southern Africa. As an additional note, *Rhigioglossa (Perisilvius) argentae* (Oldroyd, 1957), *Atylotus diurnus* (Walker, 1850) and *Tabanus nyasae* (Ricardo, 1900) was only found in Azonal vegetation and could not confidently be assigned to a specific biome.

The South African fauna seems to be divided along an east–west axis (Fig. 6). Pangoniinae is better represented in winter western areas (Succulent Karoo, desert and fynbos) with higher richness and abundance. The drier interior (Kalahari Savannah, Nama Karoo and Albany thicket) has an intermediate representation of Pangoniinae, with the subfamily not well represented in the summer rainfall, eastern biomes (Central savannah, Lowveld savannah, Indian Ocean coastal belt and grassland). Tabaninae in turn, seem to follow the inverse and

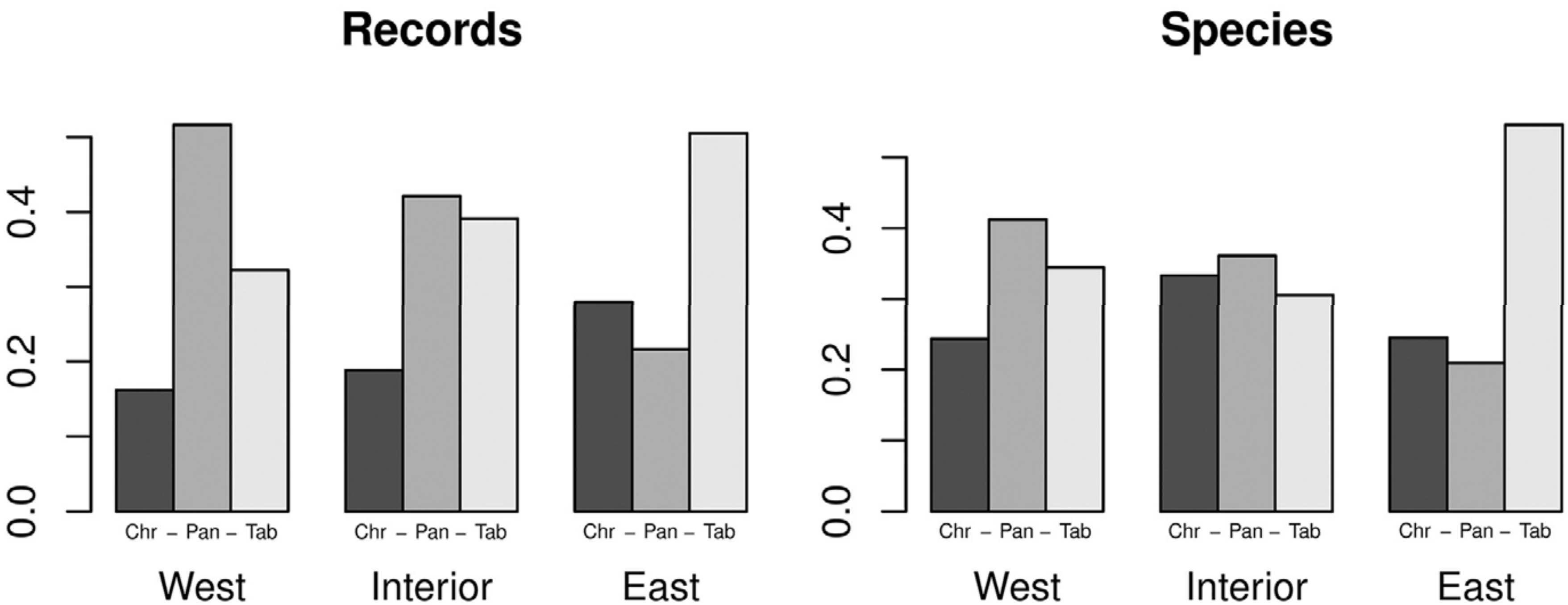

Fig. 6. Proportions of record and species when collated into the western winter rainfall biomes (fynbos, desert and Succulent Karoo), the interior (Albany thicket, Nama Karoo and Kalahari savannah) and the eastern summer rainfall biomes (Central savannah, Lowveld savannah, Indian Ocean coastal belt and grassland) indicating an inverse relationship between Pangoniinae and Tabaninae with both abundance and richness. Chr., Chrysopsinae; Pan., Pangoniinae; Tab., Tabaninae.

are best represented in the eastern section of South Africa. No clear pattern is present for Chrysopsinae, with the abundance (records) of the subfamily following a similar pattern to Tabanidae, increasing in abundance eastward, but with the highest species richness in the dry interior.

This is not an unusual distribution pattern, with the Myrmeleontidae (Neuroptera, antlions) exhibiting a similar faunal division (Mansell 1990). Mansell (1990) and Mansell and Erasmus (2002) postulated that the eastern faunal component is species often widespread into Africa. The western and interior component in turn is home to xerophilous species that diversified in the winter rainfall areas. The authors also described a general central fauna, generally low in endemicity and species richness, similar to what is seen here (Mansell 1990; Mansell & Erasmus 2002). The biodiversity that is seen in the Lowveld and Indian Ocean Coastal Belt biomes, east of the escarpment, is probably due to a continuous influx and efflux to and from the interior of Africa via the eastern tropical corridor. Even though these biomes seem exceptionally rich, the species composition will probably be very similar to much of eastern Africa. The diversity and high endemism seen in the grasslands and winter rainfall regions, however, will probably be quite unique due to species isolation over time. With increased digitised data availability, these hypotheses can be tested by increasing the regional focus.

The grassland biome is exceptionally rich in South Africa. The biome is home to nearly 40% of the species present in the region, and of those, nearly 30% are endemic. This may be an overestimation as the biome is relatively understudied, with possible synonyms and/or erroneously identified specimens, and needs to be investigated. However, most of the species described after the work of Usher (1972) are endemics to the biome and might be indicative of relatively unexplored and unconfirmed taxa present in the region (Manning 1991).

Historically, the grassland biome is at risk with as low as 1.6% of the biome formerly protected, despite covering between 15% and 28% of South Africa (Hull *et al*. 1998; Bond and Parr 2010). Grasslands are surprisingly rich in plant diversity with the number of plant species per 1000 m^2 exceeding that of the fynbos biome, famous for its high biodiversity (Bond & Parr 2010). In South Africa, the grassland biome is third, after fynbos and savannah, in total number of plant species. Additionally, a push–pull effect between grasslands and both forests and savannah might be a reason for species spillover accounting for the rich horsefly fauna in the biome.

Long-tongued flies, including Pangoniinae species, have been implicated in a series of important pollination syndromes in the grassland, savannah and fynbos biomes (Johnson & Steiner 1995, 1997, 2000; Potgieter *et al*. 1999; Goldblatt and Manning 2000). Several of these syndromes are quite specialised, and conservation of the pollinators will be central for the conservation of the flora. A comparative study on the biodiversity hotspots of tabanids vs. formally protected areas is needed to reveal the conservation status of this important ecosystem servicing taxon.

Similarly, Tabanidae is an economically important taxon that negatively affects agriculture, tourism and animal and human health due to their role in disease transmission (Parra-Henao *et al*. 2008; Baldacchino *et al*. 2014; Mulandane *et al*. 2018, 2020). In southern Africa, few studies have investigated the occurrence of horsefly-transmitted parasites, despite the known status of tabanids as vectors of *Trypanosoma* in the region (Taioe *et al*. 2017; Mulandane *et al*. 2018, 2020). Similar studies are needed to understand the role that horseflies play in disease transmission, which may positively impact management policies in the region to benefit animal and human health, and in so doing, economic losses can be mitigated.

Conclusion

Tabanidae, despite their importance as disease vectors and pollinators, have received very little attention in a nontaxonomic context. In South Africa, digitised collections can be an important tool to realise aspects of their biology and indicate possible research avenues. South Africa seems to have a rich tabanid fauna, but basic taxonomic and ecological studies are needed to assess their potential as indicators of biodiversity, ecosystem health, plant distribution patterns and potential vectors of disease. This study may provide a base for such studies and hopefully stimulate research interest in this important family.

Acknowledgements

We would like to thank all the curators and technicians from the various institutes for their willingness to help, most notably the staff from NMSA and BMSA. This work was in part supported by funding from The National Research Foundation (grant 107861, awarded to Marinda Oosthuizen), Belgian Directorate-General for Development Co-operation Framework Agreement (FA4 DGD-ITM 2017-2021) awarded to the Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria. The first author would like to acknowledge a DST/NRF freestanding award (116711) as a funding source. Three anonymous reviewers are duly thanked for valuable comments that significantly enhanced the quality of the manuscript.

Funding Information

- National Research Foundation. Grant Numbers: 107861, 116711
- Belgian Directorate-General for Development Co-operation Framework Agreement. Grant Number: FA4 DGD-ITM 2017-2021

References

Baird RC. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics* 7, 130– 136.

Baldacchino F, Desquesnes M, Mihok S, Foil LD, Duvallet G & Jittapalapong S. 2014. Tabanids: neglected subjects of research, but important vectors of disease agents! *Infection, Genetics and Evolution* 28, 596– 615.

Bond WJ & Parr CL. 2010. Beyond the forest edge: ecology, diversity and conservation of the grassy biomes. *Biological Conservation* 143, 2395– 2404.

Chainey JE. 1983. Afrotropical Tabanidae (Diptera): the genus *Philoliche* Wiedemann, subgenus *Ommatiosteres* Enderlein. *Annals of the Natal Museum* 25, 453– 474.

Chainey JE. 1987. Afrotropical Tabanidae (Diptera): the genus *Rhigioglossa* Wiedemann, 1828 (including Mesomyia Macquart, 1850, as a subgenus). *Annals of the Natal Museum* 28, 137– 159.

Chainey JE. 2017. Tabanidae (horse flies, deer flies and clegs). In: Manual of Afrotropical Diptera. Volume 2. Nematocerous Diptera and lower Brachycera. Suricata 5 (eds AH Kirk-Spriggs & BJ Sinclair), pp. 893– 913. SANBI Graphics & Editing, Pretoria.

Chainey JE & Oldroyd H. 1980. Family Tabanidae. In: Catalogue of the Diptera of the Afrotropical Region (ed RW Crosskey), pp. 275– 306. British Museum (Nat. History), London.

Evenhuis, NL. 2007. The insect and spider collections of the world website. Available at: Available from URL: http://hbs.bishopmuseum.org/codens/ [Accessed: May 2019].

Goldblatt P & Manning JC. 2000. The long-proboscid fly pollination system in Southern Africa. *Annals of the Missouri Botanical Garden* 87, 146– 170.

Hull HE, Freitag S, Chown SL & Bellamy CL. 1998. Identification and evaluation of priority conservation areas for Buprestidae (Coleoptera) in South Africa, Lesotho, Swaziland and Namibia. *African Entomology: Journal of the Entomological Society of Southern Africa* 6, 265– 274.

Johnson SD & Steiner KE. 1995. Long-proboscid fly pollination of two orchids in the Cape Drakensberg mountains, South Africa. *Plant Systematics and Evolution* 195, 169– 175.

Johnson SD & Steiner KE. 1997. Long-tongued fly pollination and evolution of floral spur length in the *Disa draconis* complex (Orchidaceae). *Evolution* 51, 45– 53.

Johnson SD & Steiner KE. 2000. Generalization versus specialization in plant pollination systems. *Trends in Ecology & Evolution* 15, 140– 143.

Karolyi F, Colville JF, Handschuh S, Metscher BD & Krenn HW. 2014. One proboscis, two tasks: adaptations to blood-feeding and nectar-extracting in long-proboscid horse flies (Tabanidae, *Philoliche*). *Arthropod Structure & Development* 43, 403– 413.

Kirk-Spriggs AH. 2017. Introduction and brief history of Afrotropical dipterology. In: Manual of Afrotropical Diptera. Volume 1. Introductory chapters and keys to Diptera families. Suricata 4 (eds AH Kirk-Spriggs & BJ Sinclair), pp. 1– 425. SANBI Graphics & Editing, Pretoria.

Kriticos DJ, Webber BL, Leriche A *et al*. 2012. CliMond: global high-resolution historical and future scenario climate surfaces for bioclimatic modelling. *Methods in Ecology and Evolution* 3, 53– 64.

Manning JC. 1991. A new species of *Rhigioglossa* (*Mesomyia*) from the Natal Drakensberg (Diptera: Tabanidae). *Annals of the Natal Museum* 32, 163– 166.

Mansell MW. 1990. Biogeography and relationships of southern African Myrmeleontidae (Insecta: Neuroptera). *Advances in Neuropterology. Proceedings of the Third International Symposium on Neuropterology. (3-4 February 1988,* Berg en Dal, Kruger National Park, South Africa), (1986), pp. 181– 190.

Mansell MW & Erasmus BFN. 2002. Southern African biomes and the evolution of Palparini (Insecta: Neuroptera: Myrmeleontidae). *Acta Zoologica Academiae Scientiarum Hungaricae* 48, 175– 184.

Morita SI. 2008. A revision of the *Philoliche aethiopica* species complex (Diptera: Tabanidae). *African Invertebrates* 49, 129– 158.

Morita SI, Bayless KM, Yeates DK & Wiegmann BM. 2016. Molecular phylogeny of the horse flies: a framework for renewing tabanid taxonomy. *Systematic Entomology* 41, 56– 72.

Mucina L & Rutherford MC. 2006. The vegetation of South Africa, Lesotho and Swaziland. South African National Biodiversity Institute, Pretoria.

Mulandane FC, Fafetine J, Van Den Abbeele J *et al*. 2018. Resistance to trypanocidal drugs in cattle populations of Zambezia Province, Mozambique. *Parasitology Research* 117, 429– 436.

Mulandane FC, Snyman LP, Brito DR *et al*. 2020. Evaluation of the relative roles of the Tabanidae and Glossinidae in the transmission of trypanosomosis in drug resistance hotspots in Mozambique. *Parasites & Vectors* 13, 1– 16.

Oldroyd H. 1952. The Horseflies of the Ethiopian Region. I. Haematopota and Hippocentrum. British Museum of Natural History, London.

Oldroyd H. 1954. The Horseflies of the Ethiopian Region. Vol. II. Tabanus and Related Genera. British Museum of Natural History, London.

Oldroyd H. 1957. The Horseflies of the Ethiopian Region. Vol. III. Subfamiles Chrysopsinae, Scepsidinae and Pangoniinae, and a Revised Classification. British Museum of Natural History, London.

Parra-Henao G, Alarcón-Pineda EP & Lopez-Valencia G. 2008. Ecology and parasitological analysis of horse flies (Diptera: Tabanidae) in Antioquia, Colombia. *Caldasia* 30, 179– 188.

Potgieter CJ, Edwards TJ, Miller RM & Van Staden J. 1999. Pollination of seven *Plectranthus* spp. (Lamiaceae) in southern Natal, South Africa. *Plant Systematics and Evolution* 218, 99– 112.

Smith VS & Blagoderov V. 2012. Bringing collections out of the dark. *ZooKeys* 209, 1– 6.

Suarez AV & Tsutsui ND. 2004. The value of museum collections for research and society. *Bioscience* 54, 66.

Taioe MO, Motloang MY, Namangala B *et al*. 2017. Characterization of tabanid flies (Diptera: Tabanidae) in South Africa and Zambia and detection of protozoan parasites they are harbouring. *Parasitology* 144 (9), 1162– 1178.

Team R. 2015. RStudio: integrated development for R. RStudio, Inc., Boston, MA URL Available from URL: http://www.rstudio.com, **42**, 14.

Team RC. 2013. R: a language and environment for statistical computing. URL Available from URL: https://www.r-project.org/

Usher PJ. 1965. Records and descriptions of Tabanidae from Southern Africa (Diptera). *Annals of the Natal Museum* 18, 27– 87.

Usher PJ. 1972. A review of the South African horsefly fauna (Diptera: Tabanidae). *Annals of the Natal Museum* 21, 459– 507.

Wickham H. 2011. ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics* 3, 180– 185.

Wilke CO. 2016. cowplot: streamlined plot theme and plot annotations for 'ggplot2'. CRAN Repos.