Biodiversity baselines: Tracking insects in Kruger National Park with DNA barcodes

Michelle L. D'Souza [a],[*], Michelle van der Bank [b], Zandisile Shongwe [b], Ryan D. Rattray [b], Ross Stewart [b], Johandré van Rooyen [b], Danny Govender [c],[d], Paul D.N. Hebert [a]

[a] *Centre for Biodiversity Genomics, University of Guelph, 50 Stone Road East, Guelph, ON N1G 2W1, Canada*
[b] *African Centre for DNA Barcoding, Department of Botany and Plant Biotechnology, APK Campus, University of Johannesburg, P.O. Box 524, Auckland Park 2006, Johannesburg, South Africa*
[c] *Department of Para-clinical Sciences, University of Pretoria, Onderstepoort 0110, South Africa*
[d] *Scientific Services, Kruger National Park, Skukuza 1350, South Africa*

ABSTRACT

Reflecting their species richness and ecological diversification, insects play a central role in terrestrial ecosystems but difficulties in species-level assignments have restricted large-scale analysis of their community structure. Employing South Africa's largest national park as a model system, we demonstrate that DNA barcoding can break this barrier. A year-long deployment of Malaise traps at 25 sites in Kruger National Park (KNP) generated 1000+ weekly collections containing about 800,000 specimens. Insect biomass averaged 1.05 g per trap-day but varied by up to 2-fold between months, being lower in the dry than wet season. Nearly 370,000 specimens were individually analyzed to reveal 19,730 Barcode Index Numbers (BINs; species proxy), a count equal to 43% of the known insect fauna of southern Africa. There was clear differentiation in insect richness and composition between KNP's two ecoregions, but little among the vegetation types comprising them. The spatial gradient in annual rainfall explained more than half of the variation in compositional similarity among sites with less similarity among samples in the wet season, particularly among those in high rainfall areas. These results suggest that the factors organising insect communities in KNP are not fine-scale vegetation differences, but larger-scale processes associated with ecoregions and rainfall. Estimates of sample coverage indicate that the species not collected are rare, comprising only 4% of the individuals in the community. With a well-parameterized DNA barcode reference library in place, metabarcoding can be used to assess future shifts in the insect fauna of KNP rapidly and inexpensively.

1. Introduction

Protected areas (PAs) play a critical role in sustaining biodiversity and function best when they are large, connected, and span multiple ecoregions (Gray et al., 2016; Smith et al., 2018). As a result, it is critical to consider ecoregion boundaries in the designation of PAs to ensure that key biotic compartments are safeguarded and to assess their resilience to future change. However, the designation of ecoregion boundaries has been based almost entirely upon the assessment of vascular plants and vertebrates (Olson et al., 2001; Dinerstein et al., 2017). As these compartments comprise a tiny fraction of overall biodiversity, there is a clear need for a better understanding of spatial structure in other taxonomic assemblages.

Insects have great potential to advance understanding of ecoregion boundaries due to their species richness and ecological diversification. Moreover, they react faster to environmental changes than plants and vertebrates (Kremen et al., 1993). However, insect communities have rarely been compared at large scale because of the difficulties in their identification. Instead, analysis has targeted indicator species (i.e., 'sentinels') (e.g., Clark and Samways, 1996; McGeoch et al., 2002) or functional groups (e.g., Kaiser et al., 2009; Buschke and Seaman, 2011). Both approaches have weaknesses. The sentinel approach presumes that a subset of taxa can represent the community-at-large, but this is rarely the case (Westgate et al., 2014). The assignment of specimens to functional groups allows inclusivity, but it suppresses any signal linked to the biological differences among the species in each partition. By enabling

direct assessments of species composition, DNA barcoding avoids both limitations (deWaard et al., 2019).

This study employs South Africa's largest national park, the Kruger (KNP), as a model system to demonstrate how DNA barcoding can enable comprehensive comparisons of insect species across large areas such as ecoregions. KNP includes habitats representing two ecoregions, the Limpopo lowveld and the Zambezian Mopane woodlands, covering nearly 20,000 km^2. In contrast to the detailed knowledge of its vertebrate fauna, the insects of KNP are little known (see McGeoch et al., 2011). Best estimates suggest the fauna includes some 40–60% of all South African insects (Braack and Kryger, 2003). As the known insect fauna (43,565 species) for the southern half of Africa is estimated to be at least half of its actual diversity (Scholtz and Holm, 1985; Scholtz and Chown, 1995), the insect fauna of KNP is undoubtedly substantial.

The present study examines insect and other arthropod (hereafter collectively termed 'insect') assemblages collected by Malaise traps. To maximize the likelihood that species with local distributions or narrow seasonal occurrences were collected, at least one trap was deployed year-round in each of KNP's 22 management sections. This study aimed firstly to quantify the effectiveness of this sampling program by comparing species richness and sampling completeness values with regional estimates. A key objective was to develop a comprehensive well-curated DNA barcode reference library for the insects of KNP so metabarcoding can be used to assess future shifts in species composition. Secondly, the spatial and temporal variation in insect biomass, abundance, richness, and compositional similarity was examined to determine the environmental drivers (i.e., climatic gradients, vegetation types, ecoregions) that play a key role in structuring patterns across the landscape. The study had the final aim of assessing trends in the variation of compositional similarity across spatial and temporal distance in relation to the dominant environmental drivers.

2. Materials & methods

2.1. Study sites

A single Malaise trap was deployed at each of 25 sites in KNP ranging from 22.4°–25.5° S and from 30.9–32.0° E (Fig. B1). Mean monthly rainfall varies from 10 mm in the driest winter months (Jun–Aug) to 100 mm in the wettest summer months (Dec–Feb) with mean annual rainfall twice as high in the south-west (740 mm) as in the north (440 mm) (Gertenbach, 1980). Monthly mean temperatures show limited variation, ranging from 17.8 °C in the winter to 26.4 °C in the summer (Zambatis, 2006). The 25 sites included habitats representing two ecoregions (sensu Dinerstein et al., 2017); 14 were in the Limpopo lowveld (hereafter 'lowveld'), found largely in the south, while 11 were in the Zambezian Mopane woodlands (hereafter 'woodlands') in the north (Fig. B1a). The sites also spanned 11 of the 15 vegetation types in KNP recognized by the ecozone classification of Gertenbach (1983) based largely on dominant soil and vegetation patterns (Fig. B1b). Eight of these vegetation types were in the lowveld: Knob Thorn/Marula savannah (4 sites), Sabie/Crocodile thorn thicket (3 sites), Malelane mountain bushveld (2 sites), Lembombo mountain bushveld (1 site), Olifant rugged veld (1 site), Pretoriuskop sourveld (1 site), sandveld (1 site), and thornveld (1 site). By comparison, only three were in the woodlands: Mopane/Bushwillow woodlands (5 sites), Mopane shrubveld (4 sites), and Alluvial plains (2 sites).

2.2. Field sampling and processing protocol

A Townes-style Malaise trap (BT1002, Bugdorm) was deployed at each site (Townes, 1972). Samples were preserved in 95% ethanol and collected on a near-weekly basis (mean ± SE: 7.8 ± 0.12 days) over a 392-day interval (2018-05-04 to 2019-05-31). The ethanol in the collection bottle was refreshed after its contents were emptied into Whirl Pak bags and stored away from direct light. A total of 1082 samples were collected across 56 weeks with an average of 43 samples (range: 21–58) per site (Fig. B2). From this total, 506 samples were selected for processing, representing every second collection at each site. The other samples were stored at −20° C; they remain available for additional barcode analysis or for metabarcoding.

The biomass of each of the 506 samples was determined before its component specimens were sorted, arrayed, labelled, and databased following standard protocols (deWaard et al., 2019). Analysis began by removing the specimens in each sample from ethanol by applying vacuum to a Buchner funnel lined with a 0.45 μm filtration membrane and, once it was drained, the wet weight of the sample was determined with a Sartorius balance (QUINTIX5102-1S). Specimens were then returned to fresh ethanol and stored at −20 °C until they were sorted. Daily biomass was calculated by dividing the sample biomass by the number of days elapsed between bottle deployment and retrieval.

Each specimen in a sample was separately barcoded except for a few very common morphospecies (hereafter 'excess individuals'). For example, when a particular morphospecies dominated a sample, reflecting an emergence of alate termites or a nearby ant colony, only a subset was selected for sequencing. Based on the relationship between biomass and specimen count (Fig. B3), an estimated 56,000 individuals (13%) were not processed. Specimen analysis began with the removal of a leg from large individuals, while small-bodied individuals were placed directly into 96-well plates and recovered after DNA extraction (Porco et al., 2010).

2.3. DNA barcoding protocol

Alkaline lysis was used to generate DNA extracts (Truett et al., 2000) before PCR amplification of the 658 bp barcode region of COI in 384-well plate format. The amplicons from 24 of these plates were pooled to allow the analysis of 9216 specimens at a time. Single molecule, real-time sequencing was carried out on a Sequel platform and each sequence was linked to its source well/specimen by examining the Unique Molecular Identifiers before sequence validation and analysis on mBRAVE (Hebert et al., 2018) (see Appendix A1 & A2 for supplementary lab protocols).

Sequences were then imported to BOLD and assigned to an existing or new Barcode Index Number (BIN) (Ratnasingham and Hebert, 2007, 2013). With very few exceptions, each specimen was assigned to a family, based on its membership in a previously identified BIN, its grouping with other identified BINs of the same family or by morphological inspection of its members. As BIN assignments are dynamic, all data were downloaded on April 1, 2020 and analyses reflect their assignments at that time. All specimen data are compiled in four datasets accessible on BOLD: DS-KMTPD1, DS-KMTPD2, DS-KMTPD3, DS-KMTPD4.

2.4. Data analysis

Analyses were performed with packages and functions in R version 3.6.3 (R Core Team, 2020). Values are presented as means ± SE and ranges (min–max).

2.4.1. Quantifying sampling efforts with estimates of species richness and sample completeness

Species richness was defined as the number of BINs in a sampling unit while the specimen count was the number of individuals (see Appendix A3 for more details on the choice of species diversity measures). Insect communities invariably include many rare species that are either never or infrequently detected. As a result, observed species richness underestimates the real species richness (observed plus undetected) of a region and the degree of underestimation is strongly dependent on sampling effort (Chao and Jost, 2012). Both asymptotic and non-asymptotic statistical approaches were used to estimate species richness and to assess the completeness of sampling in KNP (Chao and Chiu,

2016).

Asymptotic methods aim to estimate the asymptote of a species accumulation curve and those employing non-parametric estimators are more robust as they make no assumptions about the underlying species abundance distribution of the assemblage. The commonly used non-parametric estimators *Chao1*, and the *first-order* and *second-order jack-knife estimators* were employed to estimate species richness including a bias-corrected version of each designed to improve performance under conditions of low sampling effort (Lopez et al., 2012) (R BAT:alpha. accum). We calculated these estimates using abundance data where estimate values were based on counts for detected rare species, particularly the numbers of singletons and doubletons. Finally, the *combined non-parametric estimator* (CNE), a combination of the Chao1 and *Chao2* estimators was also examined (Ronquist et al., 2020; CNE estimate (Eq. (9), $\hat{S}_{CNE} = S_{obs} + \frac{Q_1(Q_1-1)}{2Q_2+1}$) corrected to $\hat{S}_{CNE} = S_{obs} + \frac{Q_1(Q_1-1)}{2Q_2+2}$).

Non-asymptotic methods aim to control the dependence of the observed species richness values on sampling effort by standardization to a common sample size, either smaller than (rarefaction) or larger than (extrapolation) that observed (Chao et al., 2014). It needs emphasis that the total number of undetected species in a community is difficult to estimate when many rare species are present because extrapolations beyond twice the observed sample size are unreliable (Colwell et al., 2012). In contrast to the difficulty in ascertaining the total species count, the fraction of the community represented by the encountered species can be more accurately estimated. Termed *sample coverage*, this value represents the proportion of the total species abundance curve at a site represented by species that were sampled (Chao and Jost, 2012). Since rare species do not significantly alter this proportion, even when there are many of them, sample coverage can be accurately estimated. Its complement (1 – sample coverage) represents the proportion of total individuals at a site that belong to undetected species.

For comparisons between assemblages, coverage-based rarefaction and extrapolation yield less biased estimates of species richness with less total sampling effort than traditional methods based on sample size (Chao and Jost, 2012). Both sample-size- and coverage-based rarefaction and extrapolation sampling curves were used to obtain a standard estimate of species richness (hereafter 'rarefied species richness') (R

iNEXT:estimateD). For comparisons among sites, species were rarefied to a coverage of 0.83 and a specimen count of 4862, twice the sampling effort of the site with the lowest sample size (8-Skukuza).

2.4.2. Measuring spatial and temporal variation in insect biomass and biodiversity

2.4.2.1. Biomass, abundance, and species richness. General linear models (R stats: lm) were used to identify the main factors impacting three biodiversity metrics across the 25 sites: biomass per trap-day, specimen count, and rarified species richness. These models considered 10 explanatory variables: sampling day, excess individuals, latitude, longitude, vegetation type, ecoregion, annual mean temperature, seasonality of temperature, annual mean rainfall, and seasonality of rainfall. Only the non-asymptotic species richness estimates (i.e., rarefied species richness) were used for modeling as most asymptotic estimates fell beyond the range which yield reliable values (Colwell et al., 2012) (see Table 1).

Excess individuals were considered as a binary variable to account for differences in the biomass and specimen count in those 39% of samples where all individuals were weighed but not all were counted. Estimates of the four latter climatic variables were obtained for each site from WorldClim v2 which is based on mean monthly values from 1970 to 2000 (Fick and Hijmans, 2017). Seasonality was measured as the coefficient of variation in monthly values over the year (e.g., Hijmans and Graham, 2006). Long-term rainfall data from WorldClim was strongly correlated with direct rainfall measurements in the park over the sampling period ($r = 0.79$, $P < 0.001$) (Fig. B4). Months were assigned to the wet (Oct–Apr) or dry (May–Sept) season based on rainfall patterns and the average vegetation response as measured by the normalized difference vegetation index (NDVI) time series data collected in KNP from 2000 to 2012 by Smit et al. (2013).

Environmental variables were standardized to a mean of 0 and a standard deviation of 1 to adjust for differences in the magnitude and spread of values. Diagnostic plots for all models were examined for deviations from normality and for heteroscedasticity and corrected to meet model assumptions. The Akaike Information Criteria corrected for

Table 1
Species richness estimates based on abundance data of insects from 25 sites in Kruger National Park.

Site	N	Specimens	BINs	Chao1	Chao1P	Jack1ab	Jack1abP	Jack2ab	Jack2abP	CNE
01 - Malelane	13	17,297	4033	7717	10,022	6237	8100	7783	10,107	4855
02 - Crocodile Bridge	9	19,862	3991	7877	10,185	6151	7953	7712	9971	14,627
03 - Stolznek	24	10,259	2797	5198	6771	4336	5649	5383	7013	11,437
04 - Pretoriuskop	26	7275	2086	4637	6318	3342	4554	4290	5845	11,079
05 - Lower Sabie	10	31,115	4027	7949	10,169	6155	7874	7707	9859	14,239
06 - Skukuza	22	20,531	3323	6209	7963	5089	6526	6316	8100	12,230
07 - Skukuza	9	4081	1144	2433	3288	1822	2462	2323	3139	5440
08 - Skukuza	26	2431	986	2377	3420	1639	2358	2140	3079	8867
09 - Tshokwane	22	5798	1804	3871	5263	2886	3924	3686	5012	9737
10 - Nwanetsi	19	8304	2205	4351	5711	3438	4513	4318	5668	9424
11 - Kingfisherspruit	22	6438	2095	4686	6403	3363	4595	4322	5905	10,981
12 - Satara	25	15,451	3386	6652	8646	5240	6811	6569	8538	13,734
13 - Houtboschrand	27	8103	1962	3976	5283	3087	4102	3899	5181	8790
14 - Olifants	22	11,567	2930	5487	7136	4536	5899	5639	7333	11,612
15 - Phalaborwa	19	4357	1202	2610	3572	1932	2645	2474	3387	5948
16 - Letaba	26	16,787	2624	4642	5819	3945	4945	4835	6060	8460
17 - Mahlangeni	12	4409	1190	2677	3673	1916	2629	2466	3384	6431
18 - Mooiplaas	26	14,462	2459	4770	6168	3790	4900	4739	6127	9713
19 - Woodlands	25	20,492	3542	6413	8099	5358	6766	6601	8336	11,632
20 - Shangoni	26	6527	1797	3952	5329	2858	3854	3659	4935	9412
21 - Shingwedzi	27	12,055	2441	5269	6946	3818	5033	4861	6408	10,449
22 - Vlakteplaas	26	7918	1841	3778	5013	2894	3841	3662	4860	8231
23 - Punda Maria	15	4775	1197	2901	4007	1936	2674	2516	3475	7814
24 - Pafuri	21	6315	1305	2815	3775	2067	2772	2638	3537	6195
25 - Pafuri	7	6164	1414	2679	3453	2174	2802	2707	3489	5581
Total	506	272,773	19,730	30,212	35,019	27,600	31,991	32,517	37,691	49,998

Species richness estimates include Chao1 (CHAO1), first-order jackknife (JACK1AB), second-order jackknife (JACK2AB), their bias-corrected complements (CHAO1P, JACK1ABP, JACK2ABP respectively) and the combined non-parametric estimator (CNE).

small sample sizes (AICc) was used to determine the degree of support for the various models while analysis of variance was used to test for significant effects (R stats: AICc, anova).

2.4.2.2. Quantifying compositional similarity. The compositional similarity among sites was measured using the *Sørensen* similarity index and visualized by UPGMA clustering of pairwise site similarity (see Appendix A3 for more details on the choice of species diversity measures). Values range from 0 to 1 with the maximum indicating comparisons where all species are shared. In addition, non-metric multidimensional scaling (NMDS) was used to analyze the variation in compositional similarity among sites and cumulatively in KNP among weeks using a dissimilarity matrix based on the Sørensen similarity index optimized for two dimensions (stress = 0.14; non-metric fit, $R^2 = 0.98$) (R MASS: metaMDS). The NMDS ordination was fit for factor averages displaying maximum correlation with corresponding environmental variables to examine those that best fit the variation in community composition extracted by the main ordination axes (R vegan: envfit; $P < 0.001$, 999 permutations). Ten variables were considered: latitude, longitude, annual mean temperature, seasonality of temperature, annual mean rainfall, seasonality of rainfall, vegetation type, ecoregions, months, and seasons.

Scores for the first NMDS axis were extracted as response variables describing community change in the subsequent general linear models. However, environmental variation, interspecific interactions, neutral processes, and stochasticity may result in communities being spatially structured (Borcard et al., 2011). To help discriminate the mechanisms that generate these structures and the scale at which they operate, Moran's Eigenvector Maps (MEMs) were used to model the spatial relationships in community composition. The MEM analysis produces eigenvectors from the geographic coordinates of the 25 sites and a *spatial weighting matrix* that models the spatial interactions among them based on a connectivity and a weighting matrix (Borcard and Legendre, 2002). These matrices were constructed using a data-driven approach by selecting the configuration that resulted in optimal performance of the spatial model as recommended by Dray et al. (2006). MEMs were computed for four types of connectivity (decreasing order of nestedness representing large to finer scale spatial effects) (Fig. B5) and for all combinations of these connectivity matrices with three monotonic weighting functions that decreased with distance (see Dray et al., 2006).

The eigenvectors of each spatial weighting matrix were calculated, and a subset of eigenvalues with significant positive spatial correlation relative to the response variable—that is the NMDS axes—were retained based on forward selection and Moran's I statistic ($P < 0.05$, 999 permutations) to avoid model overfitting (R adespatial: mem.select). The best spatial weighting matrix was then selected corresponding to the lowest values of AICc. The selected MEMs and environmental variables were then combined into a single model to assess their shared influence on variation in community composition. In addition, the variation of the overall model was partitioned between the two factors to assess the contribution of environmental variables independent of any significant underlying spatial structures (R vegan: varpart).

Finally, the overall heterogeneity among sites in KNP was assessed by multiplicatively partitioning the total species richness (gamma, D_γ) into its two independent components: mean species richness per site (alpha, D_α) and the effective number of compositionally distinct sites, vegetation types, and ecoregions (beta, $D_\beta = D_\gamma/D_\alpha$) (R vegan: multipart) (Baselga, 2010; Tuomisto, 2010). Beta diversity has a minimum value of 1 when all species are shared and a maximum value equivalent to the number of sampling units ($n_{site} = 25$, $n_{veg} = 11$, $n_{eco} = 2$) when no species are shared.

2.4.3. Assessing trends in the variation of compositional similarity
The change in compositional similarity over space and time was examined using distance-decay graphs where pairwise Sørensen

similarity values between the 506 samples were regressed onto geographic as well as temporal distance (i.e., time between collection dates) and examined in relation to the dominant environmental drivers in the landscape. When examining trends in similarity with geographic distance, only samples collected in the same week were analyzed ($n = 1243$) while the examination of trends over temporal distance were based on the analysis of samples from each site ($n = 5391$). That is, comparing 506 samples resulted in 127,765 similarity values of which 1243 values are comparisons between samples in the same week and 5391 values are comparisons between samples at the same site.

3. Results

3.1. Sampling completion and sequence recovery

Sequencing success averaged $75 \pm 1.0\%$ (61–83%) across sites with sequences recovered from 272,773 of the 367,743 specimens analyzed from the 506 samples. As such, sequencing success was about 10% lower than in prior studies (Geiger et al., 2016; D'Souza and Hebert, 2018), likely because samples were exposed to high temperatures for up to four months due to the lack of refrigeration, particularly for sites in the north of KNP (Fig. B6a). The percentage of specimens generating a sequence varied among arthropod classes (Fig. B6b) and orders (Fig. B6c) with substantially lower recovery for Arachnida (48%) as well as the orders Ephemeroptera (56%), Symphypleona (55%), Odonata (43%), Poduromorpha (37%), and Phasmatodea (29%).

The sequences were assigned to 19,730 BINs. Insecta comprised 96% of the 272,773 specimens with BINs, while Arachnida and Collembola represented 6976 (2.6%) and 3270 (1.2%) specimens respectively (Fig. B7a). Sequences were recovered from 32 orders of arthropods, but five made up 93% of the total: Diptera (130,628), Hymenoptera (57,212), Hemiptera (28,060), Lepidoptera (22,534), Coleoptera (14,750). Despite being represented by half as many specimens as the Diptera, Hymenoptera were more species rich, making up 35% of the 19,730 BINs (Fig. B7b). In fact, the number of BINs of Hymenoptera (6867) exceeded the known species count for southern Africa (Fig. 1). The same was true for the Psocoptera (148) while the number of BINs for three other orders—Diptera (5786), Embioptera (31), and Thysanoptera (214)—were close to the known counts for the region.

Most BINs (96%) were only identified to a family, but 12% and 3.8% were identified to a genus and species, respectively. Among the 458 families (Fig. B7), the most abundant and species-rich was Cecidomyiidae (30,460 specimens; 2169 BINs). Other abundant and diverse families included the Ceratopogonidae (24,933 specimens; 469 BINs), Platygastridae (15,401 specimens; 1264 BINs), Cicadellidae (15,054 specimens; 611 BINs), and Muscidae (14,801 specimens; 191 BINs).

More species await detection as accumulation curves did not reach an asymptote at any site in either ecoregion, or when all 506 samples were pooled (Fig. 2a). Based on the non-parametric species richness estimates, the median diversity of KNP insect fauna is 32,517 species (27,600–49,998 species), 39% more species than were encountered in the samples that were analyzed (Table 1). However, the sample coverage index indicated that these 'missing taxa' comprise just 4% of the specimens in KNP readily sampled by Malaise traps (Fig. 2c). The same analysis showed how sample coverage rose with increasing analytical effort. For example, sample coverage reached 50% after just 1620 specimens were analyzed and achieved 75% with 8370 specimens.

3.2. Insect biodiversity patterns and dominant environmental drivers

3.2.1. Biomass, specimen, and species richness variation

The 506 samples of insects had a total biomass of 4.08 kg for an average of 1.05 ± 0.04 g per trap-day. The daily trap catch pooled for all 25 sites varied more than 35-fold (0.19–6.70 g per trap-day) across the 392 sampling days with median values ranging from 0.37 g per trap-day in July to 1.75 g per trap-day in December (Fig. 3). Variation in the
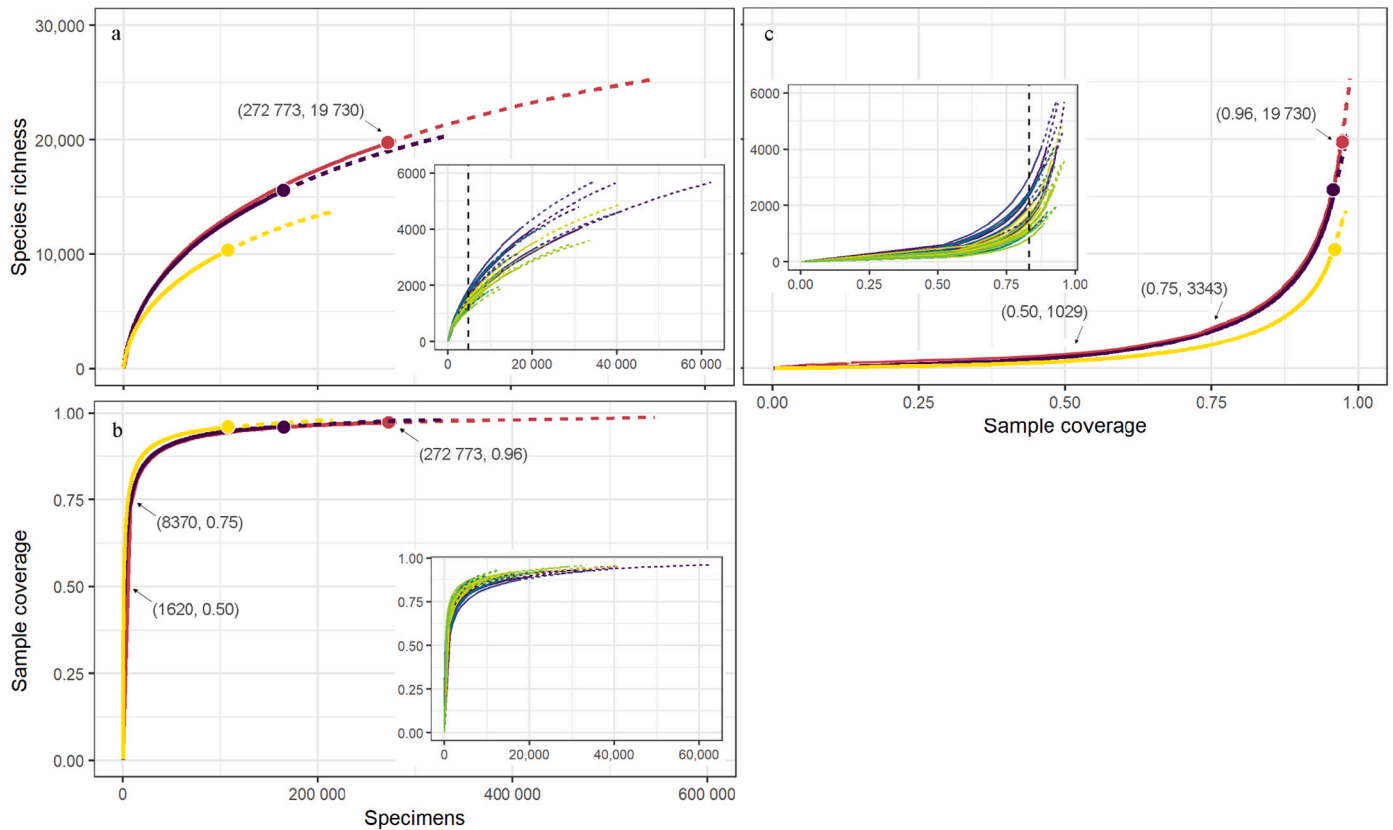
Fig. 2. Sample-size-based (a) and coverage-based (b) rarefaction (solid line) and extrapolation (dashed line) sampling curves as well as sample completeness curve (c) for 272,773 specimens with barcodes from Kruger National Park (red) and separately for the lowveld (purple) and woodlands (yellow) ecoregions. Solid dots represent the observed richness and sample coverage values with 50% and 75% coverage indicated for KNP. Inset graphs outline sampling curves per site in the lowveld (dark blues) and woodlands (light greens). Black dashed line indicates sampling effort for site-level comparisons. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
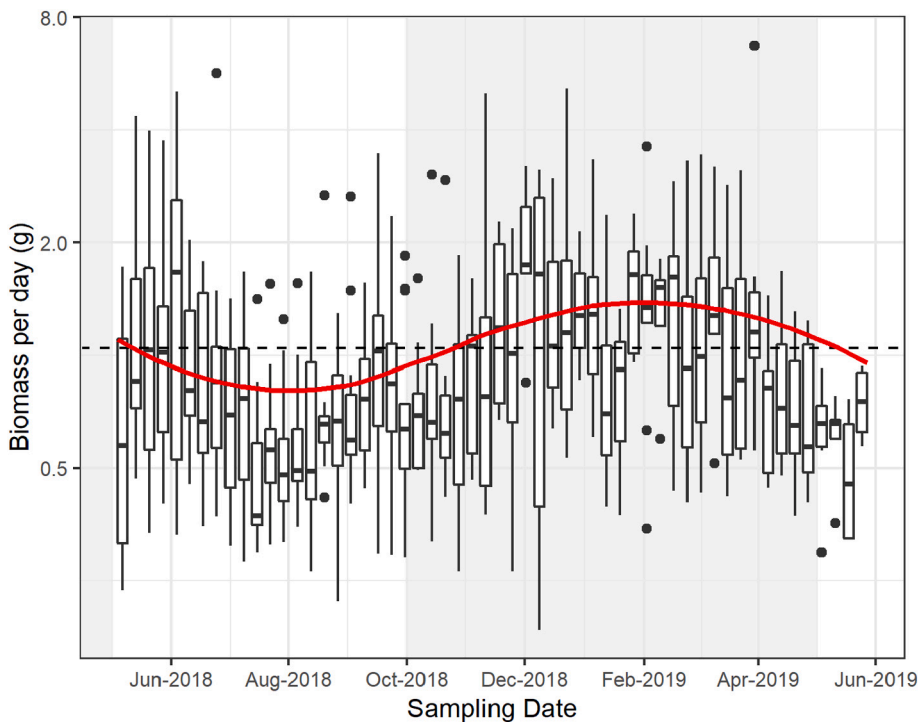
Fig. 3. Variation in biomass (g) per day for the 506 samples collected across the 392 sampling days in Kruger National Park. Biomass per day was calculated by dividing the total biomass of each sample by the duration of collection. Each boxplot represents values for the samples in one of the 56 weeks. The dashed line indicates the average biomass per day considering all samples while the solid red line models the periodic regression ($R^2 = 0.06$) for the cumulative data. Grey shading indicates the wet season (Oct–Apr). The Y-axis is $\log_2$ scaled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pooled biomass per trap-day followed a sinusoidal function ($R^2 = 0.09$) with an amplitude of 0.31 g per trap-day reflecting an overall variation of ~33% around the 1.05 g average (0.79–1.4 g per trap-day, 2-fold variation) with 4–10% change between consecutive months (Fig. 3). Overall, patterns of biomass variation in KNP insects show a decrease in the early dry season (May–Jul) followed by an increase from the late dry season (Aug–Sept) into the early wet season (Oct–Jan) before declining in the late wet season (Feb–Apr). When considering spatial effects in the model, significant differences in biomass per trap-day were apparent among sites ($F_{14, 478} = 16$, $P < 0.001$), vegetation types ($F_{9, 478} = 5.7$, $P < 0.001$), and ecoregions ($F_{1, 478} = 4.5$, $P < 0.05$), jointly explaining 43% of the variation in biomass values (Fig. B8).

Samples contained an average of 727 ± 42 specimens (13–11,118) for a total of 367,743 (Fig. B9). The specimen count, after accounting for excess individuals, increased with biomass ($t_{502} = 30$, $P < 0.001$) ($F_{2, 502} = 557$, $P < 0.001$, $R^2 = 0.69$) (Fig. B3). Over the year, the catch per site averaged 14,710 ± 1846 specimens, but there was 12-fold variation (3163–39,173) among traps (Fig. B10). Both the variation in sampling days ($t_{22} = 0.17$, $P = 0.86$) and the number of samples with excess individuals ($t_{22} = 2.0$, $P = 0.06$) had small effects on the number of specimens collected per site, together explaining 22% of the variance ($F_{2, 22} = 3.08$, $P = 0.07$) (Fig. B11). Once the variation in sampling effort among sites was considered by dividing the specimen count by the number of sampling days at each site, daily catches showed more pronounced variation (19–404 specimens, 20-fold) (Fig. B10). Despite this effect, only two sites (7, 25) were considerably affected by this standardization.

While the coverage-based rarefied species richness among sites showed a slightly different pattern from the sample-size-based rarefied species richness and the observed species richness (Fig. B12, B13), the effects of vegetation type and ecoregion were unchanged. Coverage-based rarefied species richness was higher in the lowveld ($F_{1, 14} = 16.4$, $P < 0.01$) with no effect of vegetation type ($F_{9, 14} = 2.4$, $P = 0.07$) (Fig. B14). When examining trends in climatic data from WorldClim, higher annual rainfall was associated with higher rarefied species richness at a site ($F_{1, 23} = 15$, $P < 0.001$) while seasonal variation in rainfall had the opposite effect ($F_{1, 23} = 14.7$, $P < 0.001$) (Fig. 4).

3.2.2. Spatial and temporal drivers of compositional similarity

The proportional number of compositionally distinct units was
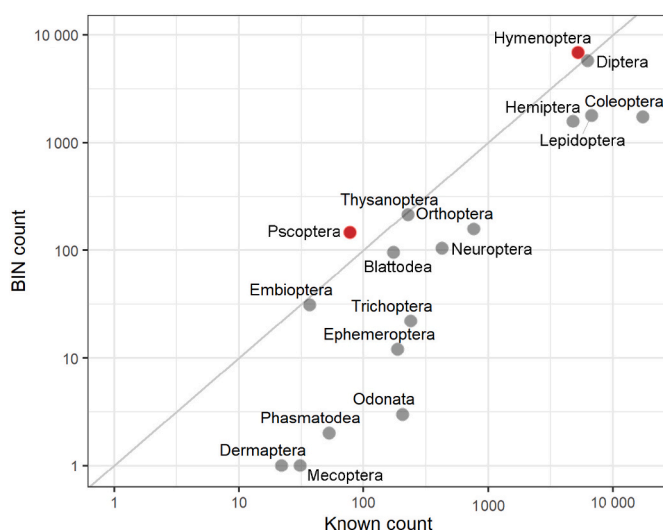
Fig. 1. The number of BINs collected in Kruger National Park versus the number of species known from southern Africa for each of 17 insect orders. Red denotes orders with BIN counts that exceed the known species count for southern Africa. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

higher for ecoregions (1.45 of max 2 units = 45%) than for vegetation types (4.91 of max 11 units = 39%) or sites (8.52 of max 25 units = 31%) indicating beta diversity decreased (or more species were shared) when considering ecoregions than vegetation types than sites as sampling units. On average, 403 ± 44 BINs (range: 111–864) were unique to a site, and, in aggregate, they represented 51% of the 19,730 BINs. Additionally, 69% of the BINs were unique to an ecoregion, and 77% are so far only reported from KNP.

Sites were very dissimilar, with an average pairwise Sørensen similarity index of 0.24 ± 0.004 (Fig. 5a). Within a vegetation type and ecoregion, similarity averaged 0.28 ± 0.01 ($n = 27$) ranging from 0.19 between sites (21, 25) in the Alluvial plains (woodlands) to 0.40 between sites (2, 5) in the Knob thorn/Marula savannah (lowveld). Within different vegetation types in the same ecoregion, between-site similarity averaged 0.25 ± 0.006 ($n = 119$) with a low of 0.13 in the lowveld between sites (2, 23) at opposite ends of the park and a high of 0.39 between sites (16, 19) in the woodlands. Finally, sites in different vegetation types and ecoregions averaged 0.23 ± 0.005 ($n = 154$) ranging from 0.10 between sites (4, 24) at opposite ends of the park and 0.37 between sites (14, 19) in adjacent vegetation types.

Sites within an ecoregion were loosely clustered (envfit $R^2 = 0.31$, $P < 0.001$) but clearly distinguished on the first axis in an NMDS ordination based on their Sørensen-based dissimilarity (Fig. 5b). Although sites were more tightly clustered by vegetation (envfit $R^2 = 0.71$, $P < 0.01$), separation was less obvious, particularly in the woodlands where sites from different vegetation types showed considerable overlap in their insect communities. To further investigate the effect of spatial relationships among sites on community composition, Moran's Eigenvector Maps (MEM) were constructed with the sphere of influence connectivity matrix and with a concave-up weighting function ($1/d^2$, where d is the distance between sites) (Fig. B5). Three of 24 structures were significant ($P < 0.05$), all at broad rather than fine spatial scales, together accounting for 60% of the variation in community composition as represented by the NMDS ordination analysis: MEM1 ($R^2 = 0.28$, $P = 0.002$), MEM2 ($R^2 = 0.14$, $P = 0.005$), and MEM4 ($R^2 = 0.18$, $P = 0.003$) (Fig. 6). MEM1 reflected an association between two sites at Skukuza (7, 8) while MEM2 indicated a similar association for the two sites at Pafuri (24, 25). In addition, two sites juxtaposed on the ecoregion boundary (13, 14) were associated. MEM4 also reflected this edge effect between sites 13 and 14 as well as associations among sites on either side of the ecoregion boundary.

Variation in mean annual rainfall ($t_{20} = -3.0$, $P < 0.01$) was a significant factor in explaining variation in community composition after considering the spatial structures modelled by the three MEMs ($F_{4, 20} = 22$, $P < 0.001$) (Fig. B15), and it was most strongly correlated with MEM4 ($r = -0.72$, $P < 0.001$). Together, annual rainfall and the three MEMs explained 62% of the variance in community composition as captured by the ordination analyses. Variance partitioning indicated that more than half (36%) of this variation was explained jointly by annual rainfall and the three MEMs indicating that the spatial patterning of rainfall in KNP produced a similar spatial structure in community composition. Rainfall-related processes independent of spatial structures accounted for only 7% of the variability, while spatial structures independent of rainfall accounted for 19%.

When the community composition patterns of samples collected across all 25 sites were considered over the full 392-day interval, weeks within a month were tightly clustered in the NMDS ordination (envfit $R^2 = 0.81$, $P < 0.001$) but the nine May samples collected in two separate years were less cohesive (Fig. B16). A seasonal pattern was also evident, with samples clustering into months in a circular pattern and separating between the dry and wet season on the first axis (envfit $R^2 = 0.36$, $P < 0.001$).

3.3. Compositional similarity trends along spatial and temporal distances

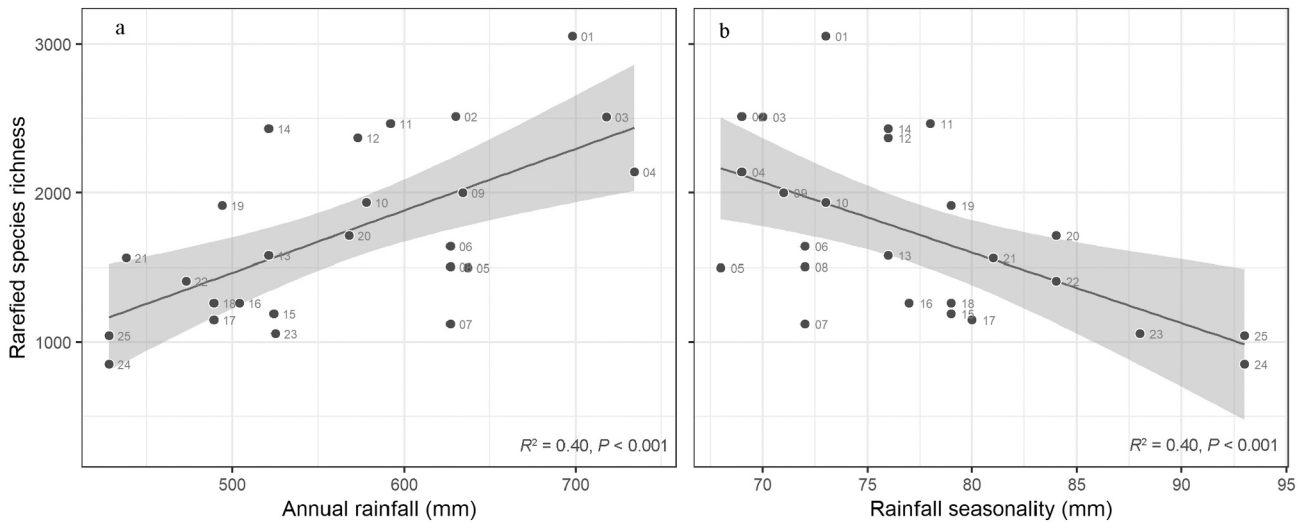Change in the compositional similarity between pairs of samples

Fig. 4. The association between annual rainfall (a) and rainfall seasonality (b) and the coverage-based rarefied species richness for insects at the 25 sites in Kruger National Park. Site indicated by the plotted number as shown in Fig. 6. Seasonality was measured as the coefficient of variation in monthly values over the year.
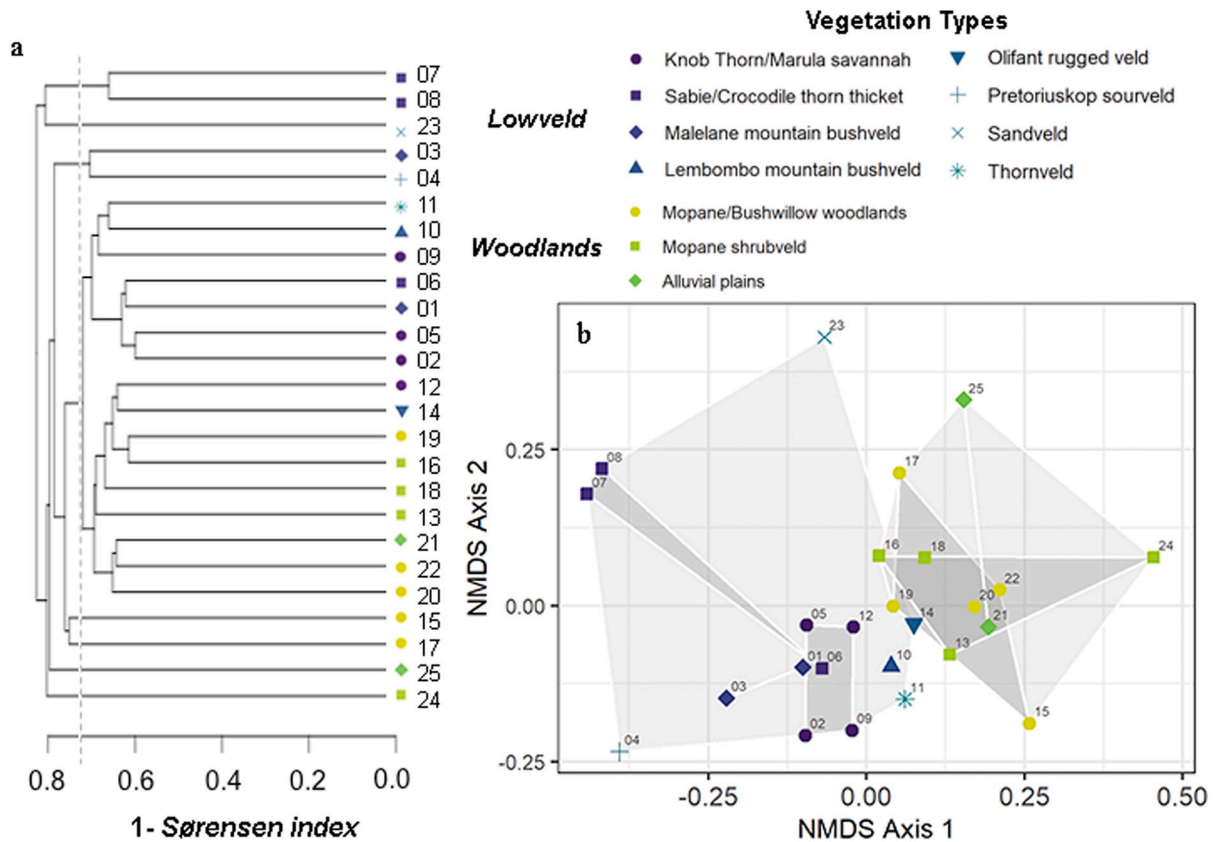

Fig. 5. Dissimilarity values (1 - Sørensen index) of insect communities between 25 sites in Kruger National Park. a) Dendrogram of Unweighted Pair Group Method with Arithmetical Mean (UPGMA) with dashed line indicating the average pairwise dissimilarity value (0.76). A dissimilarity value of 1 indicates no species overlap between sites. b) Non-metric Multidimensional Scaling (NMDS) ordination with light grey hull indicating clustering of sites per ecoregion and darker grey hull per vegetation type.

collected in the same week decreased significantly with geographic distance in the woodlands ($F_{1, 609} = 6.4, P < 0.05$) but not in the lowveld ($F_{1, 630} = 1.4, P = 0.23$) although the latter sites averaged 100 km further apart (Fig. 7a).

When pairwise samples at all site were compared over the year, the change in community composition with increasing temporal distance was best modelled with a quadratic rather than linear function with similarity highest for samples separated by either brief or long intervals and lowest for samples collected about 200 days apart ($t_{5385} = 26, P < 0.001$) (Fig. 7b). Sites in the woodlands displayed a narrower curvature indicating a faster rate of change in community composition with increasing time lag ($t_{5385} = 4.2, P < 0.001$), largely reflecting differences between the seasons ($t_{5385} = -6.3, P < 0.001$) ($F_{5, 5385} = 848, P < 0.001$). During the wet season, faunal similarly between samples was
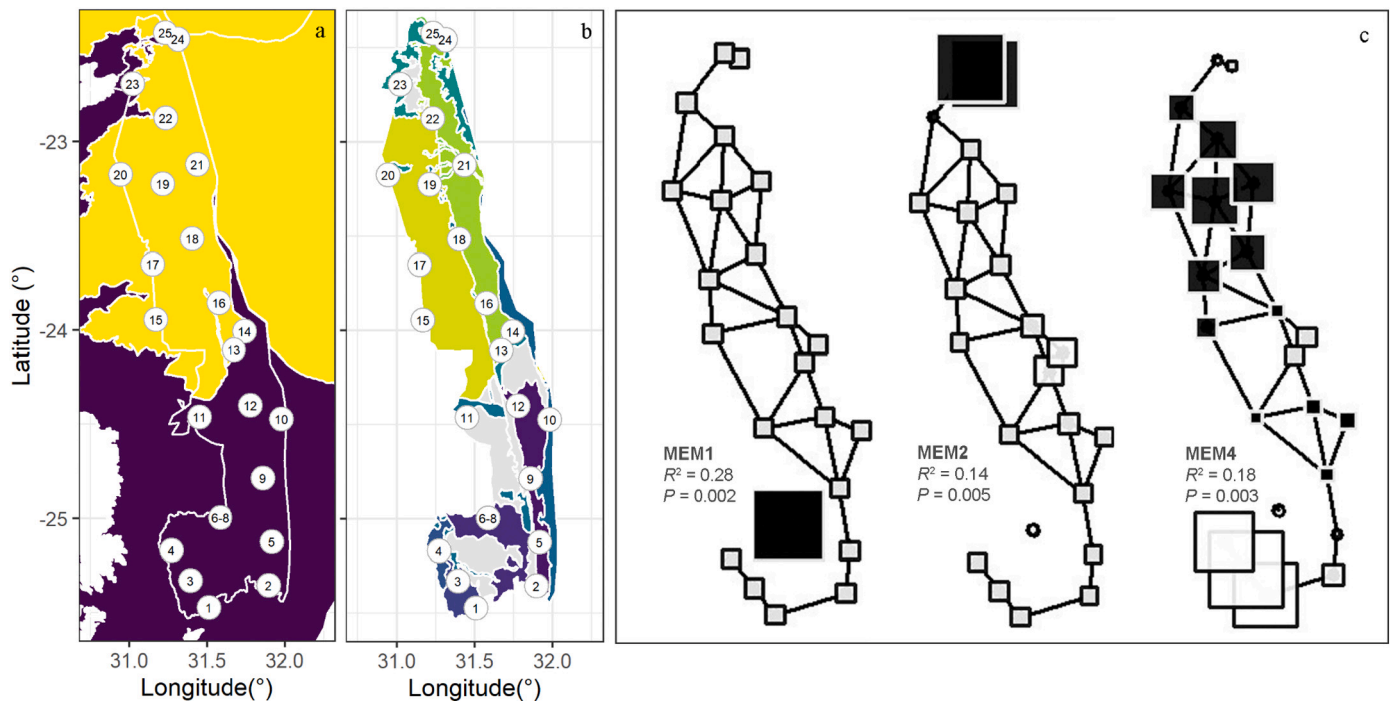
Fig. 6. Locations of the 25 Malaise traps deployed in the two ecoregions (a) and 11 vegetation types (b) represented in Kruger National Park (see Fig. B1 for details). Moran's eigenvector maps (MEM) (c) illustrating three significant spatial structures in the community composition data as described by the NMDS ordination in Fig. 3b. Each square represents the values in each eigenvector or site score for each of the 25 sites, with squares of similar size and colour indicating sites with similar community composition along the respective Moran eigenvector axis.

higher in the woodlands than in the lowveld (Fig. B17). When the effect of seasons on ecoregions was considered, a significant interaction (t_{5383} = 3.5, $P < 0.001$) accounted for the ecoregion difference (t_{5385} = 1.5, P = 0.12) ($F_{7, 5383}$ = 613, $P < 0.001$). In addition, samples collected in May 2018 and 2019 were 3-fold lower in similarity than those collected in May of the same year indicating a decline in similarity over the year or, more likely, a slight seasonal shift.

4. Discussion

Current ecoregion boundaries for terrestrial environments are almost solely based on the study of plant and vertebrate communities (Olson et al., 2001). As insects represent the most diverse compartment of terrestrial life, the inclusion of information on their patterns of diversity will help to refine ecoregion boundaries, supporting the conservation efforts based on them. The present study employed Kruger National Park (KNP) as a model to demonstrate how DNA barcoding enables large-scale evaluations of the spatial and temporal patterns of insect species.

4.1. Ecoregions and other dominant drivers

Despite the host specificity of many herbivorous insects and, by extension, their specialized parasitoids (Krüger and McGavin, 1998), demographic processes affecting insect populations coupled with their active dispersal (i.e., occupancy dynamics) might blur connections to local plant communities (Hortal et al., 2010). Our study supports this conclusion as vegetation type did not predict differences in species richness or variation in community composition among sites. However, insect communities did differ between the ecoregions within KNP. Studies in arid environments of Argentina and New Mexico with lower species richness (~400) revealed that insect communities were structured by ecoregions and were impacted by soil heterogeneity and precipitation patterns (Lightfoot et al., 2008; González-Reyes et al., 2012). A larger-scale evaluation (200 million point occurrences) revealed stronger linkages between ecoregion boundaries and the species

composition of plants and vertebrates than insects (Smith et al., 2018). Further analysis of these data indicated that compositional turnover between ecoregions was more strongly linked to abiotic factors than vegetation (i.e., rainfall variability > mean annual temperature > mean annual rainfall > vegetation (as measured by NDVI)) (Smith et al., 2020). In addition, ecoregions were found to be most distinct in tropical areas with high temperatures and seasonal rainfall (Smith et al., 2020). However, the latter conclusion was only based on a few insect lineages (ants, bees, butterflies, moths, selected beetle families) and excluded many of the most diverse groups (Diptera, parasitic Hymenoptera) (Smith et al., 2018, 2020). Despite this fact, the overall patterns mirror those observed in our study which surveyed all insect lineages.

Our results align with theoretical models that suggest at coarse resolution (e.g., ecoregion scale), the spatial distribution of insects may be more impacted by abiotic factors affecting the growth rate of populations than by the biotic interactions regulating fitness (Soberón, 2010). While the latter factor can influence species distributions, it seems most important at small scales. For example, two congeneric species of fig wasps with similar ecological requirements can coexist in a landscape despite their physiological and behavioural differences, and strong competition for their shared plant host (Warren et al., 2010). Often, only extreme specialists, such as the leaf miner *Cameraria* sp. (Cornelissen and Stiling, 2009), are constrained by biotic factors even at large scales (Hortal et al., 2010).

Physiological constraints and range shifts linked to climate can impact species distributions at the ecoregion scale (Hortal et al., 2010). For example, total rainfall and its seasonality are important drivers of environmental heterogeneity in KNP (MacFadyen et al., 2016) which impacts insect abundance and diversity, a fact confirmed by our study. Rainfall patterns had differing effects on faunal similarity between the two ecoregions. Insect communities at sites in the woodlands with lower annual rainfall and greater fluctuations in rainfall were more similar than those in the lowlands, particularly during the wet season. This pattern may reflect increased insect dispersal in the woodlands in response to greater surface water availability and more synchronized
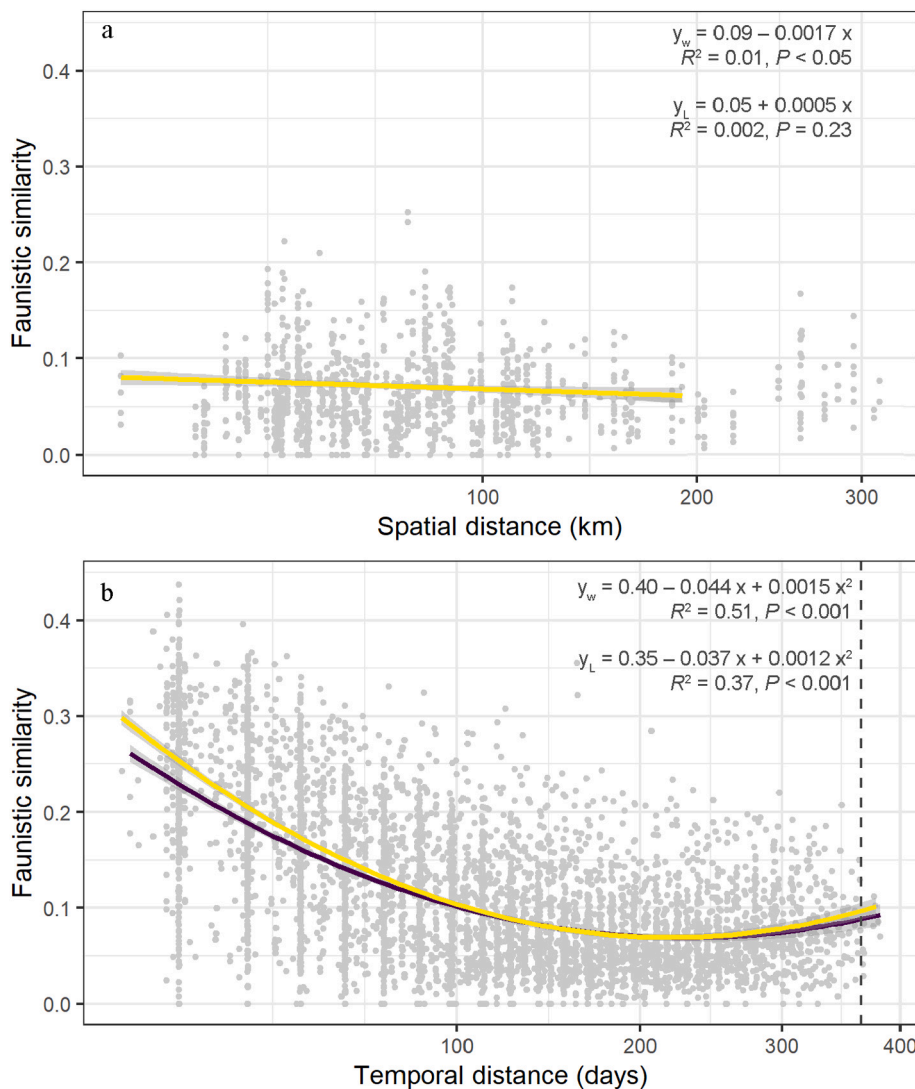
Fig. 7. Pairwise similarity values (Sørensen index) for 506 Malaise trap samples of insects over distance in space (a) and time (b) in the lowveld (purple) and woodlands (yellow). A similarity value of 0 indicates no species overlap between samples. The x-axes are square root transformed to reduce the impact on the model of fewer measurements at larger distances and long intervals. Dashed line indicates 365 days—samples a year apart. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

vegetation flush during the wet season. An increase in dispersal could also explain the higher rate of decay over distance in the woodlands. Of the six most common woody tree species in KNP, the woodlands are evenly dominated by mopane (*Colophospermum mopane*) trees that flush their leaves between Oct–Nov, the first months of the wet season (Kiker et al., 2014; Makhado et al., 2018). The trees more common in the lowveld, such as silver cluster leaf (*Terminalia sericea*), red bush willow (*Combretum apiculatum*), and knobthorn (*Senegalia nigrescens*), are more heterogeneously distributed (Kiker et al., 2014) and diverse in their phenology (Novellie, 1989; Scholes et al., 2003). Although mean rainfall in KNP shows multi-decadal oscillation with a nested seasonal cycle, these patterns have shifted over the last 40 years (MacFadyen et al., 2018). While the north-south and east-west gradients in rainfall remain, their spatial complexity has increased. As these changes will certainly impact insect communities, ongoing sampling will be critical to ascertain the response and resilience of insect communities in the KNP to shifting climates.

4.2. Placing Kruger insect assemblages in perspective

A recent study that examined long-term (10+ years) trends in global insect communities found no evidence that abundance shifts at a local or regional scale were associated with the extent of change in temperature or precipitation (van Klink et al., 2020). However, an overall 9% decline

in insect abundance per decade was apparent with stronger losses in unprotected than protected sites, and in temperate than tropical areas (van Klink et al., 2020). These patterns are based on the compilation of 166 surveys of insect abundance and biomass across 1676 sites but only two were conducted in Africa (Crosa et al., 2001; Valtonen et al., 2013), indicating the pressing need for more information on trends in African insect communities.

Because our study is the first to quantify insect biomass in KNP, it provides no direct basis for assessing trends in insect abundance. However, it does provide the baseline data needed to track future changes in the KNP while also providing information that helps to contextualize reports of major insect declines in other regions (Wagner, 2020). For example, a study of 63 PAs in Germany noted a 75% decline in insect biomass from 1989 to 2016 (Hallmann et al., 2017). Interestingly, the average daily biomass of our KNP samples (1 g) was 50% lower than recent German samples (2 g). Similarly, KNP samples were 30% lower in catch measured by specimens per trap-day than in similar studies on Swedish insects (Karlsson et al., 2020; Ronquist et al., 2020). An estimated 370 specimens per trap-day were caught at 55 sites across six habitats in Sweden, while traps in the KNP averaged 110 specimens per trap-day. Although traps of similar design (bi-colour Townes) were used in all three studies, small differences in trap construction can affect catch, so there is a need for the adoption of a standard trap to allow regional comparison.

Our study revealed as much as 35-fold variation in insect biomass among traps deployed in a single ecoregion, variation that could obscure long-term trends. Additionally, the seasonal variation in biomass and species count at single sites as well as compositional similarity among weeks was substantial, making clear the need for sampling programs to extend for at least a year (Lovell et al., 2010). In fact, year to year variation can further complicate the quantification of long-term shifts in insect biomass (Didham et al., 2020a). For example, biomass values in this study may have been low because samples were collected during a year with below average rainfall, soon after the extreme drought of 2016 (Malherbe et al., 2020).

The coupling of sampling programs using Malaise traps with DNA barcoding to identify specimens enable taxonomically comprehensive comparisons of insect faunas (Perez et al., 2015). Samples from KNP were dominated by species of Diptera and Hymenoptera and, despite undersampling, the BIN counts for these orders closely corresponded with the known species number for the whole of southern Africa. Although more than twice as many specimens of Diptera were collected in KNP, hymenopterans were more diverse, a pattern observed in Pakistan, Saudi Arabia (Ashfaq et al., 2018), and Sweden (Ronquist et al., 2020). However, surveys in Canada (Hebert et al., 2016), Costa Rica (Janzen et al., 2020), Egypt (Ashfaq et al., 2018), Germany (Geiger et al., 2016), and Greenland (Wirta et al., 2015) revealed more dipteran species. Collectively these studies indicate that these two orders dominate insect diversity in contrast to the long-held presumption that beetles are the most diverse insect order (Stork et al., 2015).

4.3. DNA barcodes for tracking insect diversity and dynamics

Approximately 100,000 species of insects are known from sub-Saharan Africa (Miller and Rogo, 2001) versus 43,565 species from southern Africa (Scholtz and Holm, 1985; Scholtz and Chown, 1995). The present study delivered records for nearly 20,000 species from KNP and suggests that another 10,000 species await collection. This work involved the barcode analysis of nearly 300,000 specimens, a substantial undertaking. However, the analysis of just 1620 specimens would have provided coverage for insect species representing 50% our specimens while analyzing 8370 specimens would bring barcode coverage for species representing 75% of all specimens collected. Reflecting this fact, highly effective identification systems can be developed with modest analytical effort.

The power of DNA barcoding to reveal insect diversity is clear when comparing these counts to those from two recent surveys reliant on morphological analysis. The study of 50,000 specimens collected from 560 km^2 of savannah habitat in three South African reserves revealed 700 species (Lovell et al., 2010) while a study across 55 sites in the Drakensberg region identified 800 species among 13,000 specimens (Hamer, 2010).

South Africa is thought to be the world's third most biodiverse nation with many of these species found in grassland and savanna habitats that have been heavily impacted by agricultural practices. Accurate, rapid assessments of species diversity are critical to evaluate the extent and pace of species loss in these settings, but such baseline data are lacking due to the taxonomic impediment (McGeoch, 2011). Consequently, ongoing monitoring efforts have sacrificed taxonomic resolution. For example, the South African Grassland Scoring System (SAGraSS) uses functional feeding groups to assess the diversity of taxa and ecosystem integrity (Kaiser et al., 2009). Our study demonstrates how DNA barcoding can enable large-scale biomonitoring programs that retain species-level resolution.

4.4. The road ahead

Despite declarations of an 'insect Armageddon' (Jarvis, 2018), we lack the long-term, comprehensive monitoring programs needed to quantify change in insect abundance and diversity (Didham et al.,

2020b; Montgomery et al., 2020). Studies on a local scale have certainly raised cause for concern, but inferences from this work are constrained by the diversity of protocols, timelines, and data quality (Desquilbet et al., 2020). Extrapolations that employ the local loss of species to predict global extinction are simply too great a stretch (Dornelas and Daskalova, 2020). However, it is perfectly feasible to track shifts in insect populations on a planetary scale by coupling a regimented sampling protocol with standardized data analysis (Hobern, 2020). Malaise traps are an ideal sampling device, but their design must be uniform to support comparisons among sites and through time. Specimens must be harvested regularly, their biomass ascertained, and their species composition determined through DNA barcoding and metabarcoding. What might it cost? Presuming volunteers harvest specimens, a program deploying 100 traps (25 in cities, 25 in agroecosystems, 50 in national parks) could be implemented in South Africa for $0.50 million (M) annually ($0.05 M–sampling equipment and preservatives; $0.05 M–two staff to organize sampling; $0.30 M–weigh/metabarcode each sample; $0.10 M–barcode selected specimens). The activation of similar programs in other nations would create a system tracking trends in abundance and diversity of insects among habitats, within nations and across nations in near real-time. This is precisely the kind of information (Arribas et al., 2021) needed to establish conservation priorities and to direct policy.

5. Conclusion

The analysis of 367,743 insect specimens collected at 25 sites in KNP revealed 19,730 species. Species assemblages were clearly differentiated between ecoregions and were structured most strongly by variation in rainfall. By providing comprehensive, spatially and seasonally explicit data on insect biodiversity in KNP, this study has delivered the baseline data needed to assess future change. Next steps involve extending analysis to South Africa's other 19 national parks and ultimately to all 4000 global national parks to obtain the baseline information needed to assess insect communities and to pave the way for the global biomonitoring systems needed to forecast future biodiversity change.

CRediT authorship contribution statement

Michelle L. **D'Souza**: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing – Original draft, Writing – review & editing, Visualization, Project administration.
Michelle van der Bank: Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision, Funding acquisition.
Zandisile Shongwe: Methodology, Investigation.
Ryan D. Rattray: Conceptualization, Methodology, Investigation, Writing – review & editing, Project administration.
Ross Stewart: Methodology, Investigation, Data Curation, Writing – review & editing, Project administration.
Johandré van Rooyen: Investigation.
Danny Govender: Conceptualization, Methodology, Validation, Resources, Writing - Review & Editing, Supervision.
Paul D.N. Hebert: Conceptualization, Methodology, Validation, Resources, Writing - Review & Editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Reference Data

Data available in data repository Dryad at dx.doi.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biocon.2021.109034.

References

Arribas, P., Andújar, C., Bidartondo, M.I., Emerson, B.C., 2021. Connecting high-throughput biodiversity inventories: opportunities for a site-based genomic framework for global integration and synthesis. Mol. Ecol. 00, 1–16. https://doi.org/10.1111/mec.15797.

Ashfaq, M., Sabir, J.S., El-Ansary, H.O., Hebert, P.D.N., 2018. Insect diversity in the Saharo-Arabian region: revealing a little-studied fauna by DNA barcoding. PLoS One 13, e0199965. https://doi.org/10.1371/journal.pone.0199965.

Baselga, A., 2010. Multiplicative partition of true diversity yields independent alpha and beta components; additive partition does not. Ecology 91, 1974–1981. https://doi.org/10.1890/09-0320.1.

Borcard, D., Legendre, P., 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. Ecol. Model. 153, 51–68. https://doi.org/10.1016/s0304-3800(01)00501-4.

Borcard, D., Gillet, F., Legendre, P., 2011. Numerical Ecology with R. Springer, New York.

Braack, L.E.O., Kryger, P., 2003. Insects and savannah heterogeneity. In: Du Toit, J.T., Rogers, K.H., Biggs, H.C. (Eds.), The Kruger Experience; Ecology and Management of Savannah Heterogeneity. Island Press, Washington, pp. 261–275.

Buschke, F., Seaman, M., 2011. Functional feeding groups as a taxonomic surrogate for a grassland arthropod assemblage. Afr. Invertebr. 52, 217–228. https://doi.org/10.5733/afin.052.0112.

Chao, A., Chiu, C., 2016. Nonparametric estimation and comparison of species richness. eLS 1–11. https://doi.org/10.1002/9780470015902.a0026329.

Chao, A., Jost, L., 2012. Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size. Ecology 93, 2533–2547. https://doi.org/10.1890/11-1952.1.

Chao, A., Gotelli, N., Hsieh, Sander E., Ellison, A., 2014. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. Ecol. Monogr. 84, 45–67. https://doi.org/10.1890/13-0133.1.

Clark, T., Samways, M., 1996. Dragonflies (Odonata) as indicators of biotope quality in the Kruger National Park, South Africa. J. Appl. Ecol. 33, 1001. https://doi.org/10.2307/2404681.

Colwell, R.K., et al., 2012. Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology 5, 3–21.

Cornelissen, T., Stiling, P., 2009. Spatial, bottom-up, and top-down effects on the abundance of a leaf miner. Ecography 32 (3), 459–467. https://doi.org/10.1111/j.1600-0587.2008.05590.x.

Crosa, G., Yaméogo, L., Calamari, D., Nabé, K., 2001. Effects of larvicide treatment on invertebrate communities of Guinean rivers, West Africa. Hydrobiologia 458, 151–158. https://doi.org/10.1023/a:1013169120020.

Desquilbet M., Gaume L., Grippa M., … Goulson D., 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. Science 370: eabd8947. doi: https://doi.org/10.1126/science.abd8947.

deWaard, J., Levesque-Beaudin, V., deWaard, S., Hebert, P.D.N., 2019. Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. Genome 62, 85–95. https://doi.org/10.1139/gen-2018-0093.

Didham, R., Basset, Y., Collins, M., Hassall, C., 2020a. Interpreting insect declines: seven challenges and a way forward. Insect Conservation and Diversity 13 (2), 103–114. https://doi.org/10.1111/icad.12408.

Didham, R., Barbero, F., Collins, M., Stewart, A., 2020b. Spotlight on insects: trends, threats, and conservation challenges. Insect Conservation and Diversity 13 (2), 99–102. https://doi.org/10.1111/icad.12409.

Dinerstein, E., Olson, D., Joshi, A., Saleem, M., 2017. An ecoregion-based approach to protecting half the terrestrial realm. Bioscience 67, 534–545. https://doi.org/10.1093/biosci/bix014.

Dornelas, M., Daskalova, G.N., 2020. Nuanced changes in insect abundance. Science 368 (6489), 368–369. https://doi.org/10.1126/science.abb6861.

Dray, S., Legendre, P., Peres-Neto, P., 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Model. 196, 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015.

D'Souza, M., Hebert, P.D.N., 2018. Stable baselines of temporal turnover underlie high beta diversity in tropical arthropod communities. Mol. Ecol. 27 (10), 2447–2460. https://doi.org/10.1111/mec.14693.

Fick, S.E., Hijmans, R.J., 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37 (12), 4302–4315.

Geiger, M., Moriniere, J., Hausmann, A., Rulik, B., 2016. Testing the Global Malaise Trap program - how well does the current barcode reference library identify flying insects in Germany? Biodiversity Data Journal 4, e10671. https://doi.org/10.3897/bdj.4.e10671.

Gertenbach, W.P.D., 1980. Rainfall patterns in the Kruger National Park. Koedoe 23, 35–43. https://doi.org/10.4102/koedoe.v59i1.1427.

Gertenbach, W.P.D., 1983. Landscapes of the Kruger National Park. Koedoe 26, 9–121. https://doi.org/10.4102/koedoe.v26i1.591.

González-Reyes, A.X., Corronca, J.A., Arroyo, N.C., 2012. Differences in alpha and beta diversities of epigeous arthropod assemblages in two ecoregions of northwestern Argentina. Zool. Stud. 51 (8), 1367–1379.

Gray, C., Hill, S., Newbold, T., Scharlemann, J., 2016. Local biodiversity is higher inside than outside terrestrial protected areas worldwide. Nat. Commun. 7, 12306. https://doi.org/10.1038/ncomms12306.

Hallmann, C., Sorg, M., Jongejans, E., Kroon, H., 2017. More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS One 12 (10), e0185809. https://doi.org/10.1371/journal.pone.0185809.

Hamer, M., 2010. African invertebrates in the international year of biodiversity. Afr. Invertebr. 51 (2), 223–230. https://doi.org/10.5733/afin.051.0201.

Hebert PDN, Ratnasingham S, Zakharov EV, … deWaard JR (2016) Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 371:20150333. doi: 0.1098/rstb.2015.0333.

Hebert, P.D.N., Braukmann, T., Prosser, S., Zakharov, E., 2018. A Sequel to Sanger: amplicon sequencing that scales. BMC Genomics 19, 219. https://doi.org/10.1186/s12864-018-4611-3.

Hijmans, R., Graham, C., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob. Chang. Biol. 12 (12), 2272–2281. https://doi.org/10.1111/j.1365-2486.2006.01256.x.

Hobern D (2020) BIOSCAN: DNA barcoding to accelerate taxonomy and biogeography for conservation and sustainability. *Genome* 0(ja). doi:https://doi.org/10.1139/gen-2020-0009.

Hortal, J., Roura-Pascual, N., Sanders, N., Rahbek, C., 2010. Understanding (insect) species distributions across spatial scales. Ecography 33 (1), 51–53. https://doi.org/10.1111/j.1600-0587.2009. 06428.x.

Janzen, D.H., Hallwachs, W., Pereira, G., Hasegawa, M., 2020. Using DNA-barcoded Malaise trap samples to measure impact of a geothermal energy project on the biodiversity of a Costa Rican old-growth rain forest. Genome 63, 407–436. https://doi.org/10.1139/gen-2020-0002.

Jarvis, B., 2018. The insect apocalypse is here. The New York Times Magazine. Accessed at: https://www.nytimes.com/2018/11/27/magazine/insect-apocalypse.html.

Kaiser, W., Avenant, N.L., Haddad, C.R., 2009. Assessing the ecological integrity of a grassland ecosystem: the applicability and rapidity of the SAGraSS method. Afr. J. Ecol. 47, 308–317. https://doi.org/10.1111/j.1365-2028.2008.00962.x.

Kiker G, Scholtz R, Smit I, Venter F (2014) Exploring an extensive dataset to establish woody vegetation cover and composition in Kruger National Park for the late 1980s. *Koedoe* 56(1): Art. #1200, 10 pages. doi.org/10.4102/koedoe.v56i1.1200.

van Klink, R., Bowler, D.E., Gongalsky, K.B., Chase, J.M., 2020. Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. Science. 368 (6489), 417–420. https://doi.org/10.1126/science.aax9931.

Karlsson, D., Forshage, M., Holston, K., Ronquist, F., 2020. The data of the Swedish Malaise Trap project, a countrywide inventory of Sweden's insect fauna. Biodiversity Data Journal 8, e56286. https://doi.org/10.3897/BDJ.8.e56286.

Kremen, C., Colwell, R.K., Erwin, T.L., Sanjayan, M.A., 1993. Terrestrial arthropod assemblages: their use in conservation planning. Conserv. Biol. 7, 796–808. https://doi.org/10.1046/j.1523-1739.1993.740796.x.

Krüger, O., McGavin, G., 1998. Insect diversity of Acacia canopies in Mkomazi game reserve, north-east Tanzania. Ecography 21 (3), 261–268. https://doi.org/10.1111/j.1600-0587.1998.tb00563.x.

Lightfoot, D., Brantley, S., Allen, C., 2008. Geographic patterns of ground-dwelling arthropods across an ecoregional transition in the North American Southwest.

Western North American Naturalist 68 (1), 83–102. https://doi.org/10.3398/1527-0904(2008)68[83:gpogaa]2.0.co;2.

Lopez, L.C.S., de Fracasso, M.P.A., Mesquita, D.O., ART, Palma, Riul, P., 2012. The relationship between percentage of singletons and sampling effort: a new approach to reduce the bias of richness estimates. Ecol. Indic. 14, 164–169.

Lovell, S.J., Hammer, M.L., Slotow, R.H., Herbert, D., 2010. Assessment of sampling approaches for a multi-taxa invertebrate survey in a South African savanna-mosaic ecosystem. Austral Ecology 35 (4), 357–370. https://doi.org/10.1111/j.1442-9993.2009.02052.x.

MacFadyen, S., Hui, C., Verburg, P.H., Van Teeffelen, A.J.A., 2016. Quantifying spatiotemporal drivers of environmental heterogeneity in Kruger National Park, South Africa. Landsc. Ecol. 31, 2013–2029. https://doi.org/10.1007/s10980-016-0378-6.

MacFadyen, S., Zambatis, N., Teeffelen, A., Hui, C., 2018. Long-term rainfall regression surfaces for the Kruger National Park, South Africa: a spatio-temporal review of patterns from 1981 to 2015. Int. J. Climatol. 38 (5), 2506–2519. https://doi.org/10.1002/joc.5394.

McGeoch, M.A., et al., 2011. Conservation and monitoring of invertebrates in terrestrial protected areas. Koedoe 53.

Makhado R, Potgieter MJ, Luus-Powell WJ (2018) *Colophospermum mopane* leaf production and phenology in southern Africa's savanna ecosystem—A review. *Insights of Forest Research* 2(1):84–90 doi:10.36959/948/464.

Malherbe, J., Smit, I., Wessels, K., Beukes, P., 2020. Recent droughts in the Kruger National Park as reflected in the extreme climate index. African Journal of Range & Forage Science 37 (1), 1–17. https://doi.org/10.2989/10220119.2020.1718755.

McGeoch, M.A., Van Rensburg, B.J., Botes, A., 2002. The verification and application of bioindicators: a case study of dung beetles in a savanna ecosystem. J. Appl. Ecol. 39, 661–672. https://doi.org/10.1046/j.1365-2664.2002.00743.x.

McGeoch M, Sithole H, Samways M, … Hamer M (2011) Conservation and monitoring of invertebrates in terrestrial protected areas. *Koedoe* 53(2), Art. #1000, 13 pages. doi: https://doi.org/10.4102/koedoe.v53i2.1000.

Miller, S., Rogo, L., 2001. Challenges and opportunities in understanding and utilisation of African insect diversity. Cimbebasia 17, 197–218.

Montgomery GA, Dunn RR, Fox R, … Wagner DL (2020) Is the insect apocalypse upon us? How to find out. *Biol Conserv* 241, 108327. doi:1016/j.biocon.2019.108327.

Novellie, P., 1989. Tree size as a factor influencing leaf emergence and leaf fall in *Acacia nigrescens* and *Combretum apiculatum* in the Kruger National Park. Koedoe 32 (1), 95–99. https://doi.org/10.4102/koedoe.v32i1.466.

Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Kassem, K.R., 2001. Terrestrial ecoregions of the world: a new map of life on earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience 51 (11), 933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2.

Perez KHJ, Sones JE, deWaard JR, Hebert PDN (2015) The Global Malaise Program: assessing global biodiversity using mass sampling and DNA barcoding. In Scientific abstracts from the 6th International Barcode of Life Conference. *Genome* 58(5):266. doi:https://doi.org/10.1139/gen-20150087.

Porco, D., Rougerie, R., Deharveng, L., Hebert, P.D.N., 2010. Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied arthropods in a high-throughput context: the example of Collembola. Mol. Ecol. Resour. 10, 942–945. https://doi.org/10.1111/j.1755-0998.2010.2839.x.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the Barcode of Life Data system (http://www.barcodinglife.org). Mol. Ecol. Notes 7, 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Ratnasingham, S., Hebert, P.D.N., 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PLoS One 8, e66213. https://doi.org/10.1371/journal.pone.0066213.

Ronquist, F., Forshage, M., Häggqvist, S., Gärdenfors, U., 2020. Completing Linnaeus's inventory of the Swedish insect fauna: only 5,000 species left? PLoS One 15 (3), e0228561. https://doi.org/10.1371/journal. pone.0228561.

Scholes, R.J., Bond, W.J., Eckhardt, H.C., 2003. Vegetation dynamics in the Kruger ecosystem. In: Du Toit, J.T., Rogers, K.H., Biggs, H.C. (Eds.), The Kruger Experience; Ecology and Management of Savannah Heterogeneity. Island Press, Washington, pp. 243–261.

Scholtz, C.H., Chown, S.L., 1995. Insects in southern Africa: how many species are there? S. Afr. J. Sci. 91, 124–126.

Scholtz, C.H., Holm, E., 1985. Insects of Southern Africa. Butterworth-Heinemann, Durban.

Smit I, Riddell E, Cullum C, Koedoe P (2013) Kruger National Park research supersites: establishing long-term research sites for cross-disciplinary, multiscaled learning. *Koedoe* 55(1), Art. #1107, 7 pages. doi:10.4102/koedoe.v55i1.1107.

Smith, J., Letten, A., Ke, P., Daily, G., 2018. A global test of ecoregions. Nature Ecology & Evolution 2 (12), 1889–1896. https://doi.org/10.1038/s41559-018-0709-x.

Smith, J., Hendershot, J., Nova, N., Daily, G., 2020. The biogeography of ecoregions: descriptive power across regions and taxa. J. Biogeogr. 00, 1–14. https://doi.org/10.1111/jbi.13871.

Soberón, J., 2010. Niche and area of distribution modeling: a population ecology perspective. Ecography 33, 159–167. https://doi.org/10.1111/j.1600-0587.2009.06074.x.

Stork, N.E., McBroom, J., Gely, C., Hamilton, A.J., 2015. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. Proc. Natl. Acad. Sci. 112, 7519–7523.

Townes, H., 1972. A light-weight Malaise trap. Entomological News 83, 239–247.

Truett, G.E., Heeger, P., Mynatt, R.L., Warman, M.L., 2000. Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT). BioTechniques 29, 52–54.

Tuomisto, H., 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity Ecography 33 (1), 2–22. https://doi.org/10.1111/j.1600-0587.2009.05880.x.

Valtonen, A., Molleman, F., Chapman, C., Carey, J., Ayres, M., Roininen, H., 2013. Tropical phenology: bi-annual rhythms and interannual variation in an Afrotropical butterfly assemblage. Ecosphere 4 (3), 1–28. https://doi.org/10.1890/es12-00338.1.

Wagner, D., 2020. Insect declines in the Anthropocene. Annu. Rev. Entomol. 65 (1), 457–480. https://doi.org/10.1146/annurev-ento-011019-025151.

Warren, M., Robertson, M., Greeff, J., 2010. A comparative approach to understanding factors limiting abundance patterns and distributions in a figtree-fig wasp mutualism. Ecography 33, 148–158. https://doi.org/10.1111/j.1600-0587.2009.06041.x.

Westgate, M., Barton, P., Lane, P., Lindenmayer, D., 2014. Global meta-analysis reveals low consistency of biodiversity congruence relationships. Nat. Commun. 5, 3899. https://doi.org/10.1038/ncomms4899.

Wirta, H., Várkonyi, G., Rasmussen, C., Roslin, T., 2015. Establishing a community-wide DNA barcode library as a new tool for arctic research. Mol. Ecol. Resour. 16 (3), 809–822. https://doi.org/10.1111/1755-0998.12489.

Zambatis, N., 2006. Average Monthly and Seasonal Temperatures (°C) of the Kruger National Park. Unpublished report, SANParks, Skukuza. Available at: http://www.sanparks.co.za/parks/kruger/conservation/scientific/weather/rainfall/2006/temps_rainfall_averages.pdf.