Received: 5 October 2022 | Revised: 28 April 2023 | Accepted: 25 May 2023

DOI: 10.1111/geb.13716

Global Ecology and Biogeography | A Journal of Macroecology | WILEY

DATA ARTICLE

Frugivoria: A trait database for birds and mammals exhibiting frugivory across contiguous Neotropical moist forests

Beth E. Gerstner[1,2] | Patrick Bills[3,4] | Phoebe L. Zarnetske[2,4,5]

[1]Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan, USA

[2]Ecology, Evolution and Behavior Program, Michigan State University, East Lansing, Michigan, USA

[3]Institute for Cyber-Enabled Research (ICER), Michigan State University, East Lansing, Michigan, USA

[4]Institute for Biodiversity, Ecology, Evolution, and Macrosystems (IBEEM), Michigan State University, East Lansing, Michigan, USA

[5]Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA

Correspondence
Beth E. Gerstner, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA.
Email: bgerstner90@gmail.com

Funding information
NASA FINESST, Grant/Award Number: Grant #80NSSC19K1332; National Science Foundation, Grant/Award Number: CC*Compute and #2200792

Handling Editor: Ines Martins

Abstract

Motivation: Biodiversity in many areas is rapidly declining because of global change. As such, there is an urgent need for new tools and strategies to help identify, monitor and conserve biodiversity hotspots. This is especially true for frugivores, species consuming fruit, because of their important role in seed dispersal and maintenance of forest structure and health. One way to identify these areas is by quantifying functional diversity, which measures the unique roles of species within a community and is valuable for conservation because of its relationship with ecosystem functioning. Unfortunately, the functional trait information required for these studies can be sparse for certain taxa and specific traits and difficult to harmonize across disparate data sources, especially in biodiversity hotspots. To help fill this need, we compiled Frugivoria, a trait database containing ecological, life-history, morphological and geographical traits for mammals and birds exhibiting frugivory. Frugivoria encompasses species in contiguous moist montane forests and adjacent moist lowland forests of Central and South America—the latter specifically focusing on the Andean states. Compared with existing trait databases, Frugivoria harmonizes existing trait databases, adds new traits, extends traits originally only available for mammals to birds also and fills gaps in trait categories from other databases. Furthermore, we create a cross-taxa subset of shared traits to aid in analysis of mammals and birds. In total, Frugivoria adds 8662 new trait values for mammals and 14,999 for birds and includes a total of 45,216 trait entries with only 11.37% being imputed. Frugivoria also contains an open workflow that harmonizes trait and taxonomic data from disparate sources and enables users to analyse traits in space. As such, this open-access database, which aligns with FAIR data principles, fills a major knowledge gap, enabling more comprehensive trait-based studies of species in this ecologically important region.

Main Types of Variable Contained: Ecological, life-history, morphological and geographical traits.

Spatial Location and Grain: Neotropical countries (Mexico, Guatemala, Costa Rica, Panama, El Salvador, Belize, Nicaragua, Ecuador, Colombia, Peru, Bolivia, Argentina, Venezuela and Chile) with contiguous montane regions.

Time Period and Grain: IUCN spatial data: obtained February 2023, spanning range maps collated from 1998 to 2022. IUCN species data: obtained June 2019–September 2022. Newly included traits: span 1924 to 2023.

Major Taxa and Level of Measurement: Classes Mammalia and Aves; 40,074 species-level traits; 5142 imputed traits for 1733 species (mammals: 582; birds: 1147) and 16 sub-species (mammals).

Software Format: .csv; R.

KEYWORDS

biodiversity, frugivore, functional diversity, IUCN, neotropics, traits

1 | INTRODUCTION

In a time of rapid global change and significant declines in biodiversity, there is an urgent need to identify, monitor and conserve biodiversity hotspots—regions of high endemic biodiversity with significant anthropogenic pressures (Myers et al., 2000). When identifying biodiversity hotspots, the level of biodiversity is often defined as endemic species richness (Myers et al., 2000). However, focusing on the number of endemic species assumes that each species plays an equal role in the environment and does not represent a holistic view of biodiversity (Devictor et al., 2010; Pollock et al., 2017). Functional traits of species are increasingly being considered in conservation with a focus on conserving functional diversity (i.e. the unique roles and functions of species in an ecosystem; Cadotte et al., 2011; Devictor et al., 2010; González-Maya et al., 2017; Gómez et al., 2021). Conserving areas with higher functional diversity—where species exhibit many functional roles—can help maintain ecosystem functioning (Cadotte et al., 2011; Cadotte & Tucker, 2018). Therefore, a functional trait approach to conservation is promising because species' responses to environmental change largely depend on their traits and incorporating functional diversity can help assess sensitivity to future changes (Newbold et al., 2014). When combined, functional diversity, taxonomic diversity and phylogenetic diversity provide an even more comprehensive assessment of an area's biodiversity, which is especially important for biodiversity hotspots such as in the Neotropics (Devictor et al., 2010; Pollock et al., 2017).

Neotropical forests contain numerous biodiversity hotspots yet lack sufficient data on species occurrences and traits (Collen et al., 2008), making it difficult to generate robust species habitat maps and forecasts of biodiversity change and species vulnerability. Despite tropical forests covering less than 5% of Earth's surface (Brandon, 2014), they contain almost half of the world's biodiversity (Dinerstein et al., 2017) and many of their species are considered data deficient (DD) by the International Union for the Conservation of Nature (IUCN; IUCN, 2022a). These DD species do not have enough information on population status, spatial distribution or threats, or combinations therein, to make an official threat status designation (IUCN, 2022a). For example, as of 2021, 15.6% of mammals, 0.15% of birds, 14.3% of reptiles, and 16.8% of amphibians are DD within the tropical forests of Central and South America; the percentages for this biodiversity hotspot are higher than the global forest percentages for both mammals and amphibians (mammal: 14.4%, bird: 0.47%, reptiles: 14.4%, amphibians: 16.5%) (IUCN, 2021). This data paucity is especially pressing in montane cloud forests—regions high in elevation that are frequented by cloud cover—because of their unique environment for many endemic species and vulnerability to climate change (Foster, 2001; Ponce-Reyes et al., 2012). Biodiversity in these regions is especially vulnerable to climate change because cloud forests exist in a narrow elevational band and are naturally fragmented (Foster, 2001). This pattern of fragmentation limits dispersal, and further, the lagging tree line shift (i.e. the upper elevational range of cloud forests) will make it difficult for species to shift their distributions to track changing climates (Fricke et al., 2022; Rehm & Feeley, 2015). Importantly, the distribution and community composition of forests in the tropics are largely determined by seed dispersal processes that are mediated by frugivores, fauna consuming fruit (Sales et al., 2021).

Seed dispersal is considered a key biotic interaction for maintaining biodiversity-ecosystem function, especially in tropical regions where almost 90% of woody plants rely on frugivores to disperse their seeds (Howe & Smallwood, 1982; Sales et al., 2021). For example, there have been multiple studies on the crucial role of frugivorous mammals and birds in dispersing seeds and how changes in their abundance and distribution influence seed dispersal services and the ability of plants to track changes in climate (Fricke et al., 2022; Sales et al., 2021). In addition to dispersing seeds for many plants, frugivores also contribute to nutrient cycling and are prey for carnivores such as felids and raptors (Farwig & Berens, 2012; Tóbon et al., 2004; Wilkie et al., 2011). Due to the strong interaction between frugivores and many fruiting plants, the trait composition of frugivore assemblages (e.g. gape size, body size) can likewise affect plant community composition (i.e. prevalence of plants with small or large seeds), and ecological resilience to perturbations (Sales et al., 2021). Comprehensive trait databases will help quantify the outsized roles that frugivorous species play in maintaining biodiversity and ecosystem functions for conservation and macroecological studies.

Over the last 20 years, there has been significant growth in the use of vouchered data from specimens housed in museum collections (Nelson & Ellis, 2019). Large quantities of data have been stored in these physical repositories for much longer; however,

access to these global institutions has not always been feasible (Nelson & Ellis, 2019). Recently, increased efforts and funding have made it possible to digitize these collections for scientific research purposes (Miralles et al., 2020; Nelson & Ellis, 2019). In particular, national and international funding over the last two decades has facilitated more open access data through institutional websites and databases (e.g. GBIF; GBIF, 2023, VertNet; Guralnick & Constable, 2010; Nelson & Ellis, 2019). This mobilization and digitization of once inaccessible data has opened a vast resource for studies in conservation, ecology and systematics, allowing researchers to perform biodiversity analyses without having to conduct costly field research or work in museums. This digitization has increased the openness and accessibility of trait data through online repositories and downloadable datasets (e.g. EltonTraits; Wilman et al., 2014, PanTHERIA; Jones et al., 2009; VertNet; Guralnick & Constable, 2010; COMBINE; Soria et al., 2021). However, existing databases only provide a subset of taxa and their traits. The lack of comprehensive trait databases is due to multiple reasons. First, compiling detailed ecological trait data for species rich clades is time-consuming. Second, actively maintaining databases is costly and time-consuming; changing taxonomies and newly discovered species require constant updating. Third, data gaps in existing databases can result from bias towards regions that are more easily accessed (Engemann et al., 2015), leading to the exclusion of more remote and rare species. As a result, tropical regions in particular have significant data gaps (Collen et al., 2008; Ferrier, 2002). Filling these data gaps is especially important in biodiversity hotspot regions which are often data poor and more sensitive to changes (Bellard et al., 2014).

To help fill the aforementioned data and knowledge gaps and aid in the use of species traits for use in conservation analyses (Cooke et al., 2019; González-Maya et al., 2017; González-Suárez & Revilla, 2013), we created a trait database, 'Frugivoria'. Frugivoria is a comprehensive trait database of species exhibiting frugivory in the Neotropics of Central America and the Andean states of South America within habitat designated 'Forest—Subtropical/Tropical Moist Montane' and/or 'Forest—Subtropical/Tropical Moist Lowland' by the IUCN. These data encompass Neotropical contiguous tropical moist montane forests and their adjacent lowland forests. Traits in Frugivoria encompass any morphological, phenological, physiological and behavioural characteristics of a species (Kissling et al., 2018), and we further extend this definition to key geographical traits relevant for species extinction risk assessments. Frugivoria harmonizes bird and mammal trait data and scientific names from existing databases and online species accounts (e.g. PanTHERIA (Jones et al., 2009), EltonTraits (Wilman et al., 2014), AnAge (Tacutu et al., 2018), Encyclopedia of Life (Parr et al., 2014), Cornell's Birds of the World (Billerman et al., 2022), University of Michigan's Animal Diversity Web (Myers et al., 2023), IUCN (IUCN, 2022a)), and adds 23,661 new records for 1732 species including information related to diet breadth, habitat breadth, habitat specialization, and adds new species and their associated traits as well as IUCN threat statuses. Furthermore, once only available to mammals, Frugivoria updates and expands select traits from the PanTHERIA dataset to birds. These traits include derived breadth traits (diet and habitat breadth) and range-based geographical traits (e.g. mean annual temperature, mean annual precipitation, human impacts across the range for multiple years), using more recent and relevant sources (e.g. CHELSA bioclimatic variables, WCS Human Footprint data) (Karger et al., 2017, 2018; Sanderson et al., 2022). The inclusion of these newly derived traits expands the utility of cross-taxa trait comparisons for mammals and birds.

Creating a reproducible workflow that aligns with FAIR (Findable, Accessible, Interoperable and Reusable) data principles (Wilkinson et al., 2016) is crucial for enabling transparent and efficient data sharing and reuse. We used these principles to generate Frugivoria, and therefore provide code for all aspects of database building, which can facilitate the creation of similar datasets for other taxa or different diet preferences. Furthermore, we provide code to explore and summarize the trait data and link associated traits with species occurrences. Many of the species within Frugivoria exist in other databases, particularly for mammals (e.g. EltonTraits and PanTHERIA), yet existing databases are not entirely overlapping, collectively they omit many frugivorous Neotropical species and traits, and certain geographical traits are outdated. In addition, there is no existing comprehensive trait database for frugivorous birds in this region. Together, Frugivoria and its open workflow in R fill essential data gaps for biodiversity conservation.

2 | METHODS

2.1 | Frugivoria overview

Frugivoria encompasses frugivorous species of the classes 'Mammalia' and 'Aves' in contiguous moist montane forests and adjacent moist lowland forests of Central and South America—the latter specifically focusing on the Andean states (i.e. Mexico, Guatemala, Costa Rica, Panama, El Salvador, Belize, Nicaragua, Ecuador, Colombia, Peru, Bolivia, Argentina, Venezuela and Chile).

The species list for this region was obtained through the IUCN API (IUCN, 2015) and subset to species with IUCN habitat designations of 'Forest—Subtropical/Tropical Moist Montane' and/or 'Forest—Subtropical/Tropical Moist Lowland' (Figure 1). We chose to use the IUCN as the basis for the database species list because their database includes the most current taxonomic information, has an open access inventory of all known species present per country that have been officially assessed and is the most appropriate species list for conservation applications.

2.1.1 | Existing databases

Databases composing a significant portion of Frugivoria include the IUCN (IUCN, 2022a), EltonTraits (Wilman et al., 2014) and PanTHERIA (Jones et al., 2009; Figure 2a), with supplemental

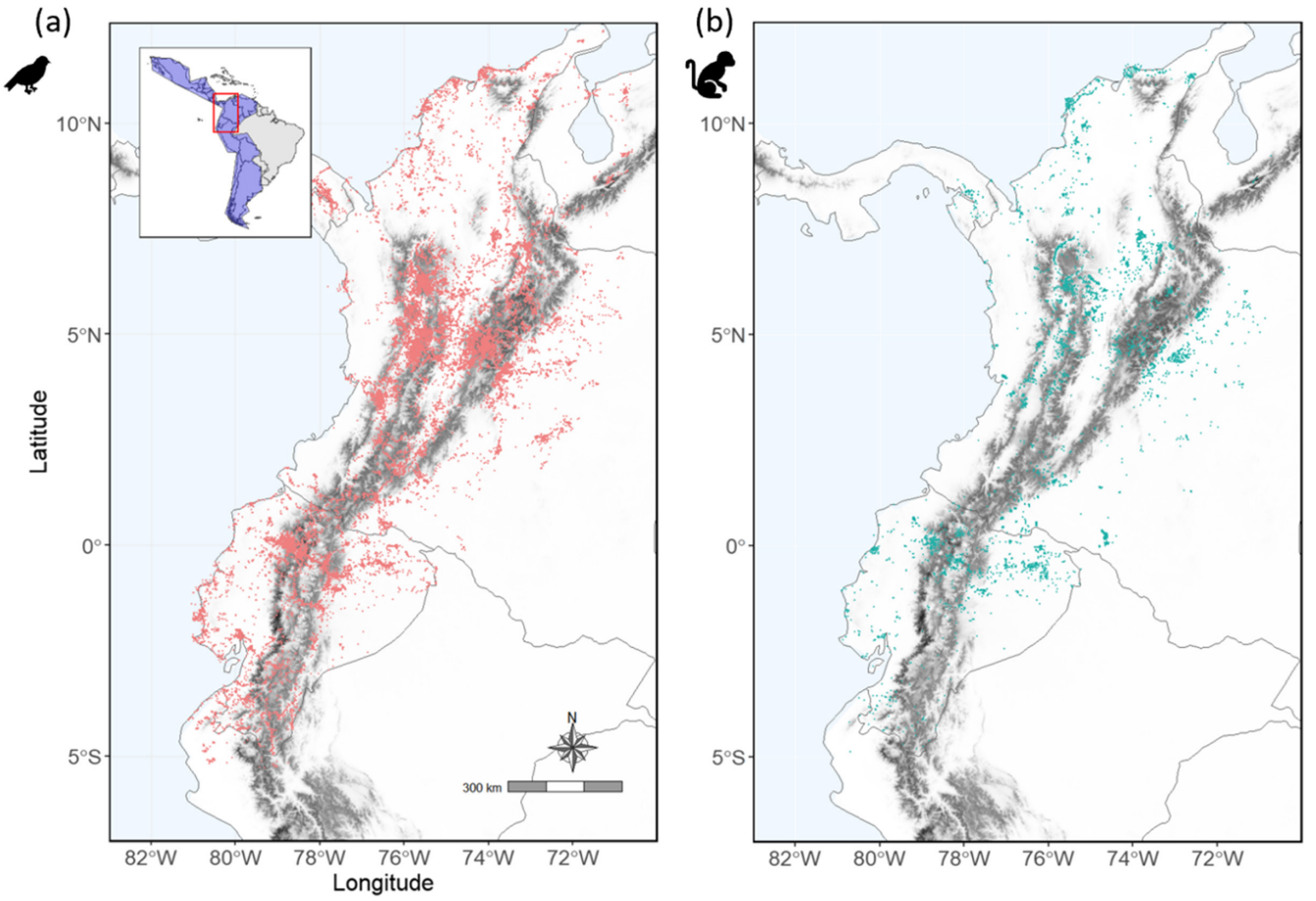

FIGURE 1 Inset map—Distribution of all Frugivoria data (purple) with a red box indicating the zoomed in region presented here. Panels show maps of occurrence records of frugivorous bird (panel a) and mammal (panel b) species in the greater Colombia and Ecuador region. Records were obtained through GBIF and pulled for differing time frames for viewing purposes (birds: 2021–2022; mammals: 2016–2022). Darker background areas show higher probability of cloud forest presence (moist montane forest; Wilson & Jetz, 2016).

information coming from the AnAge database (Tacutu et al., 2018). The IUCN Red List of Threatened Species, which formally assesses the conservation status of species, provides a large online database of formal assessments for many animals, including information on taxonomy, habitat, life history, and threats and serves as a powerful tool for biodiversity conservation and policy change (IUCN, 2022a). EltonTraits is a global database of activity patterns and feeding habits of mammals and birds (Wilman et al., 2014). PanTHERIA consists of morphological, life-history and geographical traits of mammals globally (Jones et al., 2009). In some instances, data permitting, longevity information for birds and mammals were obtained through the AnAge database when primary sources were unavailable (Tacutu et al., 2018). For a comprehensive list of all traits included in Frugivoria, their column names, definitions and sources, see Table 1.

Since the publication of all these data sources, new studies and genetic analyses have led to taxonomic revisions for many genera, which were standardized to IUCN taxonomic classifications in Frugivoria during the harmonization process (described below in Frugivoria Workflow part 2: Harmonizing).

2.2 | Frugivoria workflow

We outlined a workflow (Figure 2) for creating and replicating the Frugivoria database and to help with the organization of future databases with similar aims to this one. This workflow delineates each step of the database building process, including trait data sources, code for extracting and harmonizing the data, and highlighting the end products. Each step in the data workflow to construct Frugivoria was scripted with the R language (R Core Team, 2022) to maximize reproducibility. The data in Frugivoria are stored as comma separated data tables and R scripts, to convert raw data (Level 0) to harmonized data (Levels 1, 2) (Figure 2). The steps to generate Frugivoria are numbered below with reference to lettered sections within Figure 2 and specific scripts in the workflow.

2.2.1 | Subsetting

We downloaded static trait datasets for mammals and birds, EltonTraits (Wilman et al., 2014) and PanTHERIA (Jones et al., 2009).

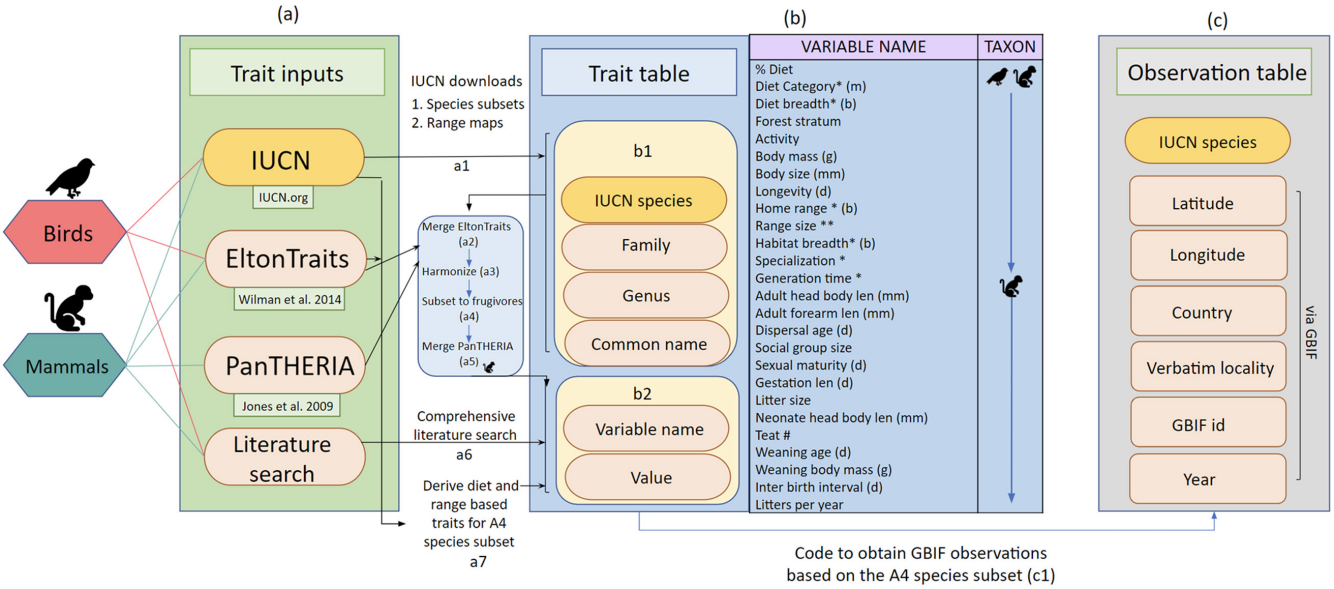

FIGURE 2 Workflow diagram of Frugivoria database generation for frugivorous birds and mammals in Central and South American moist forests. An IUCN-based species list was first downloaded for regions of interest (a1: 1) along with IUCN range maps for birds and mammals of the world (a1: 2). The species list was then subset to mammals and birds in 'Subtropical/Tropical Moist Montane' and/or 'Subtropical/Tropical Moist Lowland' habitat (a1; L0: 1_IUCN_species_list_subset.R). This IUCN species list was then merged with EltonTraits (Wilman et al., 2014) and species names were harmonized between databases (a2 & a3; L0: 2_external_trait_database_merge.R). Once these databases were merged, they were subset to only those species eating at least 10% fruit (a4; L0: 3_frugivore_subset.R). This final subset was then merged with the PanTHERIA dataset (Jones et al., 2009; a5; L0: 4_mammal_merge_pantheria.R). Traits found in the literature and other credible species accounts (a6), PanTHERIA and EltonTraits were combined for mammals, whereas only EltonTraits and traits from the literature and species accounts were combined for birds. We then derived traits such as diet breadth from the EltonTraits dataset, habitat breadth from the IUCN habitat designations and geographical traits based on the IUCN range maps (a1:2) for the a4 species subset (a7). Code to obtain observations per species is also provided and allows for spatial trait analyses (Panel c; c1; L2: downloading_gbif_records.R).

We used the R package '*rredlist*' (version 0.6.0) and the function 'rl_sp_country', which downloads IUCN Red List information over an API, to obtain IUCN species lists for each country of interest as defined above (IUCN, 2021). These lists were then subset to the classes 'Aves' and 'Mammalia' and further filtered to include only those that include an IUCN habitat designation of 'Forest—Subtropical/Tropical Moist Montane' and/or 'Forest—Subtropical/Tropical Moist Montane' (Figure 2a.1; L0: 1_IUCN_species_list_subset.R). We then merged this IUCN based species list with EltonTraits for mammals and birds (Figure 2a.2; L0: 2_external_trait_database_merge.R). After harmonizing and subsetting the dataset to include frugivorous species only (Figure 2a.3 & a4; L0: 3_frugivore_subset.R), PanTHERIA was merged with the mammal database (Figure 2a.5; L0:4_mammal_merge_pantheria.R). We then filled in new traits using information from existing literature and credible online sources (Figure 2a.6), and provided code to obtain occurrences from GBIF for all species within the final database (Figure 2c.1; L2_downloading_gbif_records.R).

2.2.2 | Harmonizing

Disparities in naming conventions between the IUCN, EltonTraits (as mentioned previously) and PanTHERIA led to some species not aligning correctly between databases and perpetuated the need for harmonization (Figure 2a.2; L0_ 2_external_trait_database_merge.R). To harmonize scientific names between databases, we first merged the final IUCN species list by scientific name with EltonTraits. We resolved conflicts in species name merges between databases using known synonyms of the IUCN species names that matched species names in the EltonTraits database. In general, any species name disparities were set aside for manual checking of synonyms using the IUCN website, existing literature and Avibase (for birds; Lepage et al., 2014), and later appended to the dataset. We generated a lookup table, hosted in the Environmental Data Initiative (EDI) repository, showing the corresponding species name in each database for all mismatched species (mammals: $n = 390$; birds: $n = 873$; lookup_table_all_mammals.csv; lookup_table_all_birds.csv), which should help facilitate the construction of databases for other species (not only frugivorous species) in this region. In Frugivoria, there were name disparities for 171 mammal and 195 bird species in the EltonTraits database. New species discovered since the publication of these databases (mammals: $n = 42$, birds: $n = 2$) and those species that had been split (taxonomically reclassified; mammals: $n = 132$, birds: $n = 182$) were assigned values of sister taxa, based on known phylogenies, or at the genus level, which was simple in EltonTraits since in some cases entire genera had matching trait values. For mammals, we then merged our final database with PanTHERIA (Jones et al., 2009; L0: 4_mammal_merge_pantheria.R) and resolved taxonomic naming issues in the same way as above, resolving the same number of species names.

TABLE 1 Frugivoria traits, their associated column name in the database, their definitions, units and sources.

Grouping feature	Trait	Column name	Description	Unit	Citation
Life history	Longevity	longevity†	Average lifespan; if only range is available this value indicates the maximum longevity.	years	Jones et al. (2009)
		min_longevity	Minimum longevity found if range was available or differing values from multiple sources.	years	
		max_longevity	Maximum longevity found if range was available or differing values from multiple sources.	years	
		max_longevity_p*	Maximum adult age measured.	months	
	Generation time	generation_time†	The average age of parents of the current cohort (i.e. newborn individuals); if only a range is available, this value indicates the maximum generation time.	years	IUCN (2022b)
		min_generation_time	The minimum age of parents of the current cohort (i.e. newborn individuals) if range was available or there were differing values from multiple sources.	years	IUCN (2022b)
		max_generation_time	The maximum age of parents of the current cohort (i.e. newborn individuals).	years	IUCN (2022b)
	Age at eye opening	age_at_eye_opening_d_p*	Age at which both eyes are fully open after birth; measure of central tendency.	days	Jones et al. (2009)
	Age at first birth	age_at_first_birth_d_p*	Age at which females give birth to their first litter (eutherians), or their young attach to teats (metatherians) or hatch out (monotremes); measure of central tendency.	days	Jones et al. (2009)
	Gestation length	gestation_len_d_p*	Length of time of non-inactive fetal growth; all measures of central tendency.	days	Jones et al. (2009)
	Interbirth interval	inter_birth_interval_d*	The length of time between successive births of the same female(s) after a successful or unspecified litter; all measures of central tendency.	days	Jones et al. (2009)
	Litter size	litter_size_p*	Number of offspring born per litter per female, either counted before birth, at birth or after birth; all measures of central tendency.		Jones et al. (2009)
	Litters per year	litters_per_year_p*	Number of litters per female per year; all measures of central tendency.		Jones et al. (2009)
	Sexual maturity age	sexual_maturity_age_d_p*	Age when individuals are first physically capable of reproducing, defined as physically sexually mature, age at first mating or unspecified (males and females), age at first oestrus or age at first pregnancy (females only), age at spermatogenesis or age at testes descent (males only), using captive, wild, provisioned, or unspecified populations; all measures of central tendency.	days	Jones et al. (2009)
	Weaning age	weaning_age_d_p*	Age when primary nutritional dependency on the mother ends and independent foraging begins to make a major contribution to the offspring's energy requirements; all measures of central tendency.	days	Jones et al. (2009)

(Continues)

TABLE 1 (Continued)

Grouping feature	Trait	Column name	Description	Unit	Citation
Ecology	% Composition diet	diet_inv_e, diet_vend_e, diet_vect_e, diet_vfish_e, diet_vunk_e, diet_scav_e diet_fruit_e, diet_nect_e, diet_seed_e, diet_plant_e	% Prevalence of foraging for: invertebrates, endotherms, ectotherms, fish, vertebrates-general or unknown, scavenging, fruit, nectar, seeds, plants.	%	Wilman et al. (2014)
	Diet category	diet_cat_e†, diet_cat†	Overall diet category (assigned for mammals using same method for birds as in Wilman et al. 2014). Assigned to the dominant diet categories based on the percent composition of diet for each species: plant and seeds, fruits and nectar, invertebrates, and vertebrates, fish, and carrion; omnivore: score of ≤50% in each category (Wilman et al. 2014).		Wilman et al. (2014)
		trophic_level_p*	Trophic level of each species; species were defined as (1) herbivore (not vertebrate and/or invertebrate), (2) omnivore (vertebrate and/or invertebrate plus any of the other categories) and (3) carnivore (vertebrate and/or invertebrate only); all measures of central tendency.		Jones et al. (2009)
	Diet breadth	diet_breadth_p*	Number of dietary categories eaten by each species; categories were defined as vertebrate, invertebrate, fruit, flowers/nectar/pollen, leaves/branches/bark, seeds, grass and roots/tubers; all measures of central tendency.		Jones et al. (2009)
		diet_breadth†	Diet breadth was calculated using the Shannon Diversity index and is based on EltonTraits % composition of diet columns. This metric takes into account the number of different types of food consumed, as well as the relative abundance of each type.		Wilman et al. (2014)
	Forest stratum	for_strat_value*†	Assigned forest strata value; marine, ground level, including aquatic foraging, scansorial, arboreal or aerial.		Wilman et al. (2014)
		for_strat_ground**†, for_strat_understory**†, for_strat_midheight**†, for_strat_canopy**†, forstrait_aerial**†	Prevalence of: foraging below on ground, understorey, mid-height, canopy and in flight.		Wilman et al. (2014)
		terrestriality_p*	Degree of terrestriality of each species; species were defined as fossorial and/or grounddwelling only and above ground dwelling; all measures of central tendency.		Jones et al. (2009)
	Activity patterns	activity_nocturnal_e† activity_crepuscular_e* activity_diurnal_e*	Presence of foraging activity at night, at dawn and dusk, and during the day.		Wilman et al. (2014)
		activity_cycle_p*	Activity cycle for wild populations: nocturnal only, nocturnal/crepuscular, cathemeral, crepuscular or diurnal/crepuscular and diurnal only. Measures of central tendency.		Jones et al. (2009)

TABLE 1 (Continued)

Grouping feature	Trait	Column name	Description	Unit	Citation
	Habitat specialization	habitat_specialization†	Descriptions of species habitat mention cloud forest as an important or sole component.		
	Habitat category	habitat†	Degree of suitability, seasonality and importance of habitat type to species as designated by IUCN		IUCN (2022a)
		habitat_suitability_lowland			
		habitat_suitability_montane			
		habitat_season_lowland			
		habitat_season_montane			
		habitat_major_importance_lowland			
		habitat_major_importance_montane			
	Habitat breadth	habitat_breadth_p*	Number of habitat layers used by each species. Categories were defined as above ground dwelling, aquatic, fossorial and ground dwelling; all measures of central tendency.		Jones et al. (2009)
		habitat_breadth†	Number of habitat layers inhabited by each species according to the IUCN. The total number of different habitats inhabited by species within this dataset is 42. See the 'habitats_all_species.csv' for the full list of possible habitat types for each species.		
	Population statistics	population_density_n.km^2_p*	Number of individuals per square kilometre; all measures of central tendency.		Jones et al. (2009)
		population_grp_size_p*	Number of individuals, adults or definition unspecified in a group that spends the majority of their time in a 24-h cycle together; all measures of central tendency.		Jones et al. (2009)
		social_grp_size_p*	Number of individuals, adults or definition unspecified in a group that spends the majority of their time in a 24-h cycle together where there is some indication that these individuals form a social cohesive unit; all measures of central tendency.		Jones et al. (2009)
Morphology	Body mass	body_mass_g†	Average mass of animal.	grams	Wilman et al. (2014)
		adult_body_mass_g_p*	Mass of adult; measures of central tendency.	grams	Jones et al. (2009)
		basal_met_rate_mass_g_p*	Mass of individual(s) from which the basal metabolic rate was taken (see metabolic trait below).	grams	Jones et al. (2009)
		neonate_body_mass_g_p*	Mass of live or freshly killed specimens of infants at either a near-term embryonic stage, birth, immediately after birth or up to an age of 7 days after birth; all measures of central tendency.	grams	Jones et al. (2009)
		weaning_body_mass_g_p*	Mass of live or freshly killed specimens of weanlings; all measures of central tendency.	grams	Jones et al. (2009)

(Continues)

TABLE 1 (Continued)

Grouping feature	Trait	Column name	Description	Unit	Citation
	Body size	body_size_mm†	Length of animal (mammals: nose to base of tail; birds: beak to end of tail). This value indicates the mean body size unless a range of values are available. In the latter case, this value indicates the maximum body size.	mm	Ansell (1965)
		min_body_size_mm	Minimum body length (mammals: nose to base of tail; birds: beak to end of tai) if range was available or there were differing values from multiple sources.	mm	
		max_body_size_mm	Maximum body length (mammals: nose to base of tail; birds: beak to end of tail) if range was available or there were differing values from multiple sources.	mm	
		adult_forearm_len_mm_p*	Total length from elbow to wrist of adult; measures of central tendency.	mm	Jones et al. (2009)
		adult_head_body_len_mm_p*	Total length from tip of nose to anus or base of tail; measures of central tendency.	mm	Jones et al. (2009)
		weaning_head_body_len_mm_p*	Total length from tip of nose to anus or base of tail of live or freshly killed specimens of weanlings.	mm	Jones et al. (2009)
	Sexual dimorphism	sexual_dimorphism†	Body pattern differs between males and females.		
	Teat number	teat_number_p*	Total number of teats present; all measures of central tendency.		Jones et al. (2009)
Metabolic	Basal metabolic rate	basal_met_rate_mLO2hr_p*	Basal metabolic rate of adult; measures of central tendency.	$mL.O_2/hr$	Jones et al. (2009)
Geographical	Home range size	home_range_size†	Area traversed by an individual in its normal activities of food gathering, mating, and caring for young; if only range is available, this value indicates the maximum home range size.	km^2	Burt (1943)
		min_home_range	Minimum area traversed by an individual in its normal activities of food gathering, mating, and caring for young; if range was available or there were differing values from multiple sources.	km^2	
		max_home_range	Maximum area traversed by an individual in its normal activities of food gathering, mating and caring for young; if range was available or there were differing values from multiple sources.	km^2	
		home_range_km2_p*	Size of the area within which everyday activities of individuals or groups (of any type) are typically restricted; all measures of central tendency.	km^2	Jones et al. (2009)
		home_range_indiv_km2_p*	Size of the area within which everyday activities of individuals are typically restricted; all measures of central tendency (see Jones et al. 2009 for details).	km^2	Jones et al. (2009)

TABLE 1 (Continued)

Grouping feature	Trait	Column name	Description	Unit	Citation
	Range size	observed_range_size†	Range size calculated from IUCN range maps and projected using a global equal-area projection; derived only from parts of the range with the presence code 'Extant'. This is the range size of areas the species is known to occur. See Workflow section 4: Range Size.	km^2	IUCN (2022a)
		inferred_range_size†	Range size calculated from IUCN range maps and projected using a global equal-area projection; derived from parts of the range using presence codes 'Extant', 'Probably Extant', and 'Possibly Extant'. This range size incorporates inferred areas where the species may exist and represents potentially suitable habitat rather than just the known range of the species. See Workflow section 4: Range Size.	km^2	IUCN (2022a)
		GR_area_km2_p*	Geographical range size of species based on digital maps from Sechrest (2003). Range areas were calculated using ArcGIS; projected using a global equal-area projection.	km^2	Jones et al. (2009)
	Range extent	GR_max_lat_dd_p*	Extent of each species range (as of 2003) calculated using a global geographical projection (decimal degrees).	Decimal degrees	Jones et al. (2009)
		GR_min_lat_dd_p*			
		GR_mid_range_lat_dd_p*			
		GR_max_long_dd_p*			
		GR_min_long_dd_p*			
		GR_mid_range_long_dd_p*			
	Human population density	hu_pop_den_min_n.km2_p_OD*	Minimum, mean and 5th percentile human population density (persons per km^2) over the species range (as of 2003) using the Gridded Population of the World (GPW) (CIESIN and CIAT, 2005).	n/km^2	Jones et al. (2009)
		hu_pop_den_mean_n.km2_p_OD*			
		hu_pop_den_5p_n.km2_p_OD*			
		hu_pop_den_change_p_OD	Mean rate of increase in human population density over the species range (as of 2003) using the Gridded Population of the World (GPW) (CIESIN and CIAT, 2005) for 1990 and 1995 as: (1995–1990)/1990.		Jones et al. (2009)
	Human footprint	mean_human_fp_range_2010†, mean_human_fp_range_2020†	Average human footprint based on the WCS Human Footprint dataset for 2010 and 2020; human influence is mapped using the weighted sum of individual maps of population density, infrastructure such as roads, railways, factories, accessibility, use of electricity; https://wcshumanfootprint.org/.		Sanderson et al. (2022)
		percent_change_hf_2010_2020†	Percent change in average human footprint calculated between the years 2010 and 2020	%	
	Average precipitation	precip_mean_mm_p_OD*	Mean monthly precipitation (mm) of species range as of 2003.	mm	Jones et al. (2009)

(Continues)

TABLE 1 (Continued)

Grouping feature	Trait	Column name	Description	Unit	Citation
	Average precipitation	mean_CHELSA_bio12_1981.2010_V.2.1†	Average annual mean precipitation from the CHELSA bioclimatic dataset calculated over the IUCN range map for each species. Derived only from parts of the range with the presence code 'Extant' or 'Possibly Extant'.	mm	Karger et al. (2017, 2018)
	Average temperature	temp_mean_01degC_p_OD*	Mean monthly temperature (0.1°C) of species range as of 2003.	°C	Jones et al. (2009)
	Average temperature	mean_CHELSA_bio1_1981.2010_V.2.1†	Average annual mean temperature from the CHELSA bioclimatic dataset calculated over the IUCN range map for each species. Derived only from parts of the range with the presence code 'Extant' or 'Possibly Extant'.	°C	Karger et al. (2017, 2018)
	AET mean	AET_mean_mm_p_OD*	Mean monthly AET (Actual Evapotranspiration Rate) of species range (as of 2003) from 1920 to 1980 (mm). Calculated using the Global Resource Information Database of UNEP.	mm	Jones et al. (2009)
	PET mean	PET_mean_mm_p_OD*	Mean monthly PET (Potential Evapotranspiration Rate) over species range (as of 2003), from 1920 to 1980 (mm) calculated using the Global Resource Information Database of UNEP.	mm	Jones et al. (2009)

Note: Traits lacking a source are those specific to the Frugivoria dataset, or minimum and maximum values of the original trait definition. Some traits are only available for mammals (*) or only available for birds (**). Traits included in cross-taxa subsets are denoted with (†).

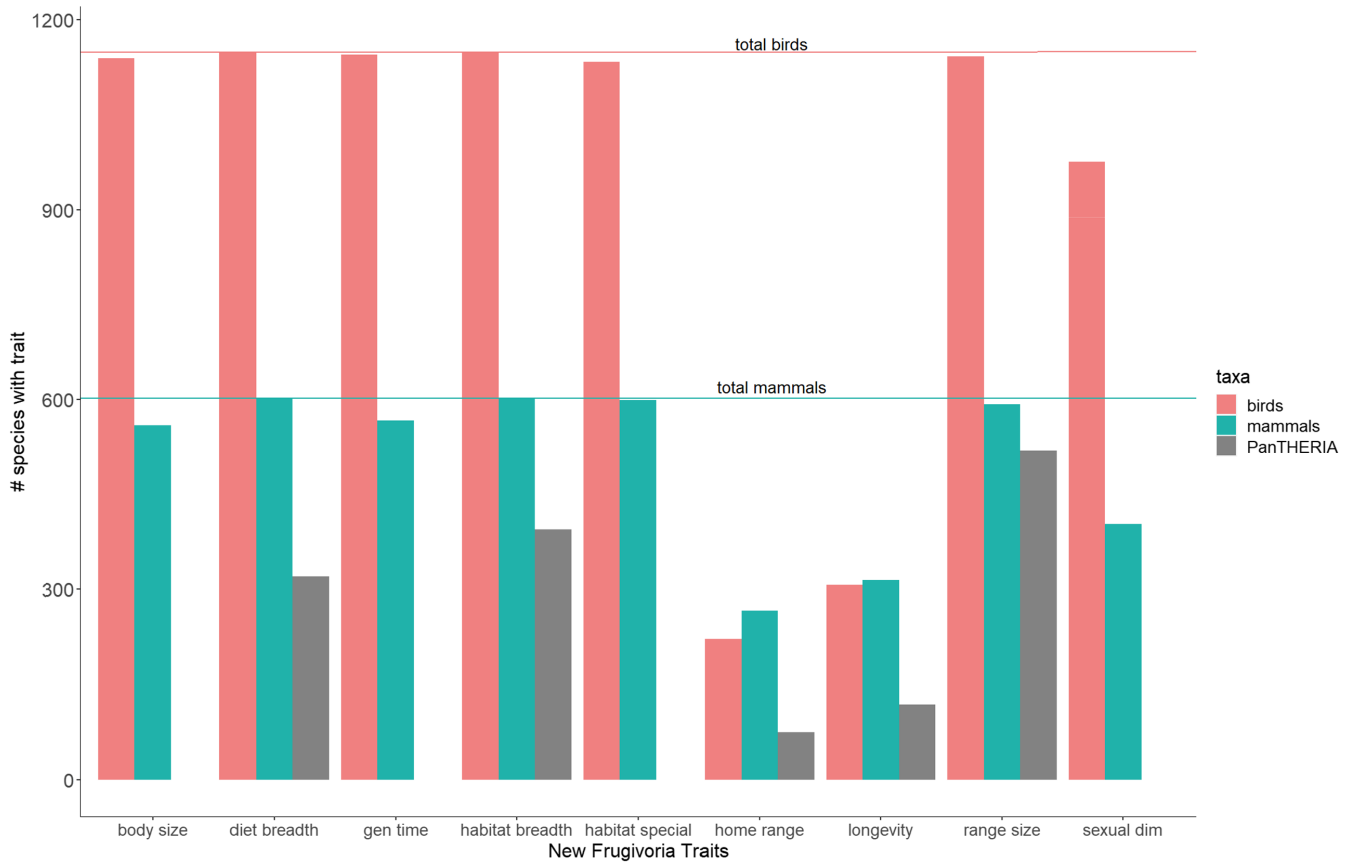

FIGURE 3 Newly added traits in Frugivoria for birds (red) and mammals (teal). Horizontal lines indicate the total number of taxa for each taxonomic group within the database (birds: $n = 1149$, mammals: $n = 602$). Certain traits for mammals overlapped with the PanTHERIA dataset (i.e. home range, longevity and range size); however, the methodology used for the trait calculations differs between Frugivoria and PanTHERIA. We have highlighted PanTHERIA traits in grey for comparison. Overall, for mammals, EltonTraits contributed 14.7% of traits, PanTHERIA 49.8% and 35.36% of traits are newly added traits. For birds, EltonTraits contributed 27.62% and newly added traits comprise 72.38% of the dataset.

We have included a 'taxonomic_disparity' column where a value of '1' indicates differences in taxonomic nomenclature between databases for some species. For each database merged in Frugivoria, we have retained the original database species name with a suffix '_e' (EltonTraits) and '_p' (PanTHERIA) as a unique column for ease of interpretation. These suffixes are also used for any trait originating from those respective databases.

2.2.3 | Frugivore selection

Once the IUCN species list and EltonTraits dataset were harmonized and successfully merged for 946 mammals and 2818 birds, we used EltonTraits 'diet_fruit' to select and retain species that had a frugivorous diet at or above 10%, leaving a total of 160 mammal genera (586 species and 16 subspecies) and 329 bird genera (1148 species) (L0: 3_frugivore_subset.R).

2.2.4 | Trait data

We then merged our frugivorous species subset and the PanTHERIA dataset for mammals (L0: 4_mammal_merge_pantheria.R). New traits and missing traits values that did not occur in existing openly accessible trait databases were then obtained by exhaustively identifying missing species and traits from the IUCN database, peer-reviewed literature, online reference material such as species accounts Encyclopedia of Life (Parr et al., 2014); University of Michigan's Animal Diversity Web (Myers et al., 2023); Cornell Lab of Ornithology's Birds of the World (Billerman et al., 2022), and reference books (Burton & Robert, 1974; Emmons & Feer, 1997; Eisenberg & Redford, 1999; Thorington et al., 2012; Fleagle, 2013 etc.; for a full list of materials see 'Frugivoria: Sources …' in the EDI hosted dataset). In total, Frugivoria contains 45,216 trait inputs with some traits having minimum and maximum values; 24,494 trait values for mammals (8662 newly added), and 20,722 for birds (14,999 newly added) (Figure 3). If multiple values were found for a given species trait, we gave preference to primary literature and the most recent data source. However, if we found multiple values from recent sources, we incorporated this information into a range of values with each source being cited in the source column for that trait but separated by a '|' symbol. For these traits, specific details about each source may be found in the trait's notes column. Relevant traits have a minimum and maximum column for instances where ranges of values were present or different values were available from alternate yet appropriate sources. Traits with no

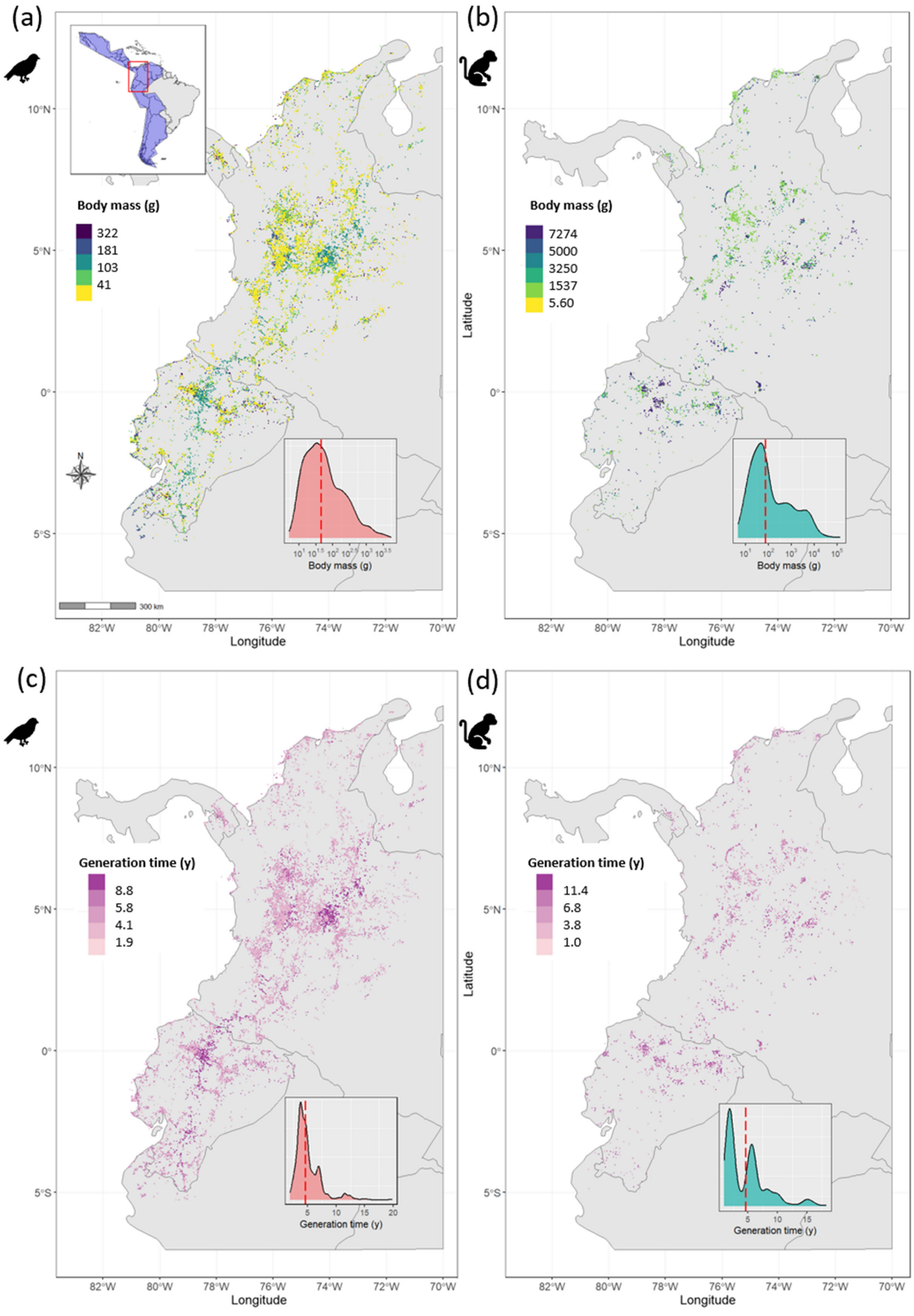

FIGURE 4 Distribution of mass and generation time for birds (a,c) and mammals (b,d) across the Northern Andes and Isthmus of Panama. Lighter colours indicate lower values. Birds and mammals differed in their minimums and maximums for each trait (birds a: 5.1–5525.0 g, c: 1.9–19.8 years; mammals b: 4.8–140,000.6 g, d: 0.94–18.3 years). Natural breaks were used to bin these data for plotting, with a few larger values included in the uppermost bin. The density curves in the lower right corner indicate the distribution of the mapped trait in the Frugivoria dataset. For mass (panels a & b), these densities were log-transformed for visualization purposes. Red dashed lines indicate the median value for each trait (birds a: 45.3 g, c: 4.6 years; mammals b: 71.0 g, d: 4.2 years).

available information for a given species were assigned to the genus level where possible and this was noted in the '*_level' column for the applicable trait. Including prior imputations in the original EltonTraits dataset and our further imputations to congenerics for reclassified and newly added species in both EltonTraits and PanTHERIA datasets, 5142 traits (11.37%) were imputed to either genus or family, with 9 phylogenetically imputed traits for mammals (previously imputed from EltonTraits). For the newly added traits, 6.92% were assigned to the level of genus or family (birds: 2.16% genus, 0.75% family; mammals: 9.02% genus, 4.83% family) if species-specific information was not available. To minimize observer error within new trait data assignments for species, all new trait data records in Frugivoria were independently entered by two technicians. Any discrepancies between the two technicians were noted and reviewed by the lead author and resolved by replacing the former value with that from the most accurate and reputable source.

Due to the disparate nature of methodology of data sources used to build the Frugivoria dataset, there are some overlapping traits between databases. Because of this, we retain them all here in their original form for ease of interpretation. For example, mass is measured as averages in EltonTraits (Wilman et al., 2014), and in PanTHERIA they are measures of central tendency (often using the median; Jones et al., 2009), thus the origin of these values is not the same and they have non-comparable units. This is true for traits shared between EltonTraits and PanTHERIA such as activity patterns and mass, as well as PanTHERIA and the data we have collected and derived for this dataset, such as body size, diet breadth, home range size, range size and longevity. For some PanTHERIA traits, we added an additional suffix '_OD', for range-based geographical traits we suspect may be outdated (e.g. mean population size, mean annual precipitation and temperature), but may be useful for some cross-time analyses or comparisons. For these traits, we have derived more recent versions and applied them across taxa, whereas the original PanTHERIA traits were only available for mammals. Similarly, diet and habitat breadth, both traits available in the PanTHERIA dataset, were recalculated and applied across taxa.

Range-based traits

Range-based metrics were calculated based on species range maps (L0_spatial_traits.R; Figure 2a.7). We obtained range maps from the IUCN Spatial Dataset (Figure 2a.1; IUCN, 2022a). We then computed the range sizes in km^2 using the 'st_area' function built into the '*sf*' package in R. We used the geography transformation of the 'shape' column within the IUCN Spatial Dataset (IUCN, 2022a). We calculated these range sizes in two ways: (1) using presence code 'Extant' and (2) using presence codes 'Extant', 'Probably Extant' and 'Possibly Extant'. The latter incorporates inferred areas where the species may exist and therefore represents potentially suitable habitat rather than just the known range of the species. We chose to include all 'origin' (e.g. native, re-introduced, introduced, vagrant) and 'seasonality' (e.g. resident, breeding season, non-breeding season, passage) codes as these all constitute a part of the range where the species would be performing the function of seed dispersal. We then appended the resulting range sizes to the full database for those species with range information available ($n = 1737$; missing ranges for $n = 16$ species). It is important to note that IUCN range maps have the potential to either overestimate or underestimate the true range of species (Gaston & Fuller, 2009; Ramesh et al., 2017) and therefore these range size estimates should be interpreted with discretion. Also, these ranges differ from those found in the PanTHERIA dataset for mammals, with nearly two-thirds of species having smaller ranges in our dataset calculated using the most recent data from the IUCN. These differences in range size may be due to different methodologies in the estimation of species ranges (IUCN, 2022a; Sechrest, 2003), differences in nomenclature between datasets (e.g. a species has been recently split and is species level for the IUCN range value and the sister species value in PanTHERIA) and the effects of ongoing habitat loss and changes in climate since the publication of sources used for calculating range size in the PanTHERIA dataset (Sechrest, 2003).

We also used the IUCN range maps to derive average climate and human impacts over species ranges, which can be used to understand the species' climatic tolerances as well as quantify potential anthropogenic impacts across the range. Specifically, we calculated the average values of two CHELSA bioclimatic variables (Karger et al., 2017, 2018), mean annual air temperature (Bio 1) and mean annual precipitation (Bio 12) over the observed and inferred portions of the range (e.g. parts of the range designated 'extant', 'probably extant' and 'possibly extant'). To quantify anthropogenic impacts across the range, we used the WCS human footprint datasets for both 2010 and 2020 (Sanderson et al., 2022). This metric quantifies human impact by considering population density, infrastructure such as roads, railways, factories, accessibility and the use of electricity. We then calculated the percent change between years as an additional metric to indicate the degree to which these human impacts have escalated over time.

Breadth traits

In an effort to increase the number of cross-taxa traits, we recalculated diet and habitat breadth for mammals based on different

data than that of PanTHERIA and extended these traits to birds (L0_7_breadth_traits.R; Figure 2a.7). For diet breadth, a measure of diet diversity, we used the Shannon Diversity Index (Shannon & Weaver, 1949). Similar to Santini et al. (2018), we calculated the Shannon Diversity Index using the 10 food categories from EltonTraits (Wilman et al., 2014) representing the proportional makeup of the species' diet. The maximum potential diversity value was calculated using the natural log of the number of possible diet categories. Here, the maximum diet diversity value is 2.3 and indicates a strong generalist species that consumes each food category evenly, whereas a value of zero indicates a monotonous diet specializing in a single source of food.

We also calculated habitat breadth by summing the number of suitable habitat types for each species based on level 2 of the IUCN Habitat Classification scheme (IUCN, 2022a). This was done using the '*rredlist*' package in R and extracting the habitat types used by each species using the 'rl_habitat' function (IUCN, 2015). A full list of habitats suitable for each species is included in the published EDI dataset.

2.2.5 | Occurrence data

We provide code to extract GBIF (GBIF, 2023) records for species contained within this database (L2_downloading_gbif_records.R). This code is modified from that provided by the GBIF blog, which overcomes the issue of pulling large numbers of records through GBIF for multiple species at once (Waller & Grosjean, 2019). A free GBIF account is required to implement this code.

2.2.6 | Frugivoria datasets

In addition to providing the full Frugivoria dataset, we also provide a simplified subset of the full Frugivoria dataset (e.g. Frugivoria_mammal_database_simple.csv, Frugivoria_bird_database_simple.csv). This subset represents shared cross-taxa traits, simplifying comparisons among birds and mammals (code contained in L0_final_database_edits.R). These traits are generally well filled; however, we also include home range and longevity traits despite their missing values (filled for birds—longevity: 26.7%, home range: 19.3%; mammals—longevity: 52.1%, home range: 44.2%; Figure 3) because of their ecological relevance and their scarcity in the literature, which may make them of heightened interest to users. This simplified subset excludes repetitive traits (e.g. ranges of values, percent composition of diet, habitat suitability), and retains traits that encapsulate and synthesize this information (e.g. trait averages, diet category, diet breadth, habitat breadth).

3 | RESULTS AND DISCUSSION

Frugivoria and its workflow facilitate studies on frugivorous mammal and bird traits encompassing ecology, life-history, morphology and geographical occurrences for a region of great ecological importance. Although mammals and birds are among the taxonomic groups with higher degrees of sampling and research, there remain significant gaps in our knowledge of their traits in this region. The existing trait databases for birds and mammals only include a subset of high-level traits (EltonTraits, Wilman et al., 2014; Phylacine, Faurby et al., 2018) or are missing values for many species in biodiversity hotspots (PanTHERIA; Jones et al., 2009)—gaps which Frugivoria helps fill.

In total, we have added 8709 new traits for mammals (35.36% of all mammal traits) and 14,999 for birds (72.38% of all bird traits) (Figures 3 and 4) and of those, only 6.92% of traits were imputed. We added traits for 44 new species, and updated the taxonomy for 314 species. Of the new morphological, ecological and life-history traits that did not require explicit calculation (i.e. home range, longevity, generation time, sexual dimorphism and body size), 1285 come directly from the literature, which is 21.6% of new traits in those categories. The remaining 79.4% of these traits (4653) were collated across disparate online sources and datasets including species accounts requiring explicit interpretation from Encyclopedia of Life (Parr et al., 2014), Cornell's Birds of the World (Billerman et al., 2022), University of Michigan's Animal Diversity Web (Myers et al., 2023), IUCN (IUCN, 2022a) and AnAge database (Tacutu et al., 2018).

We also derive new traits such as diet breadth and habitat breadth, and calculate geographical range-based traits such as observed and inferred range size, climate-based traits and different aspects of human impact across the range, which all have the potential to be used to estimate extinction risk. Both the breadth traits and new geographical range-based traits update and expand the applicability of these traits across taxa, as these were traits once only available for mammals. Furthermore, we increased the completeness of traits in PanTHERIA (e.g. body mass, body size, range size, home range size and longevity for mammals)—in some cases more than doubling the traits available for mammals in PanTHERIA (Figure 3; home range and longevity). Not only does Frugivoria help fill existing data gaps and generate comparable cross-taxa traits for birds and mammals, but it also harmonizes existing databases into a single unified source for mammals and birds, making studies of these taxa much less cumbersome and time-consuming (Etard et al., 2020). This database (and its reproducible workflow for other taxa and regions of the world) is particularly beneficial for studying functional diversity in the Neotropics, where the sheer number of species and the complexity of their interactions can make it difficult to identify patterns of ecological importance.

The increased trait resolution (i.e. low levels of imputation and greater levels of filled traits for species) and spatial and taxonomic coverage in Frugivoria provides vital information to address fundamental and applied aspects of conservation biology. Specifically, Frugivoria comprises a unified and comprehensive source that can be used to understand community assembly (species coexistence; Zamudio et al., 2016), spatial patterns of biodiversity, trait distributions (Figure 4), and can help assess the vulnerability of species to

environmental changes (Pacifici et al., 2015) in moist montane and lowland regions of Central and South America. Some examples of more focused investigations that are now possible with Frugivoria include understanding how frugivore traits help mediate ecological processes in moist montane systems (Lim et al., 2020; Sekercioğlu et al., 2004); quantifying how traits vary across montane regions (Dehling et al., 2014; Santillán et al., 2019; Figure 4); and understanding how certain traits relate to species extinction risk (Bland et al., 2015; González-del-Pliego et al., 2019; Ripple et al., 2017). The latter topic is particularly relevant given the high endemicity in this region (Gradstein et al., 2008; Myers et al., 2000), the projected rapid changes in climate (IPCC, 2022) and land use (Armenteras et al., 2011; González-Maya et al., 2017; Powers & Jetz, 2019) and the anticipated shifts in ecological communities over the next century (Williams et al., 2007). Montane regions and cloud forests, in particular, are especially sensitive to environmental changes (Elsen & Tingley, 2015; Foster, 2001; Ponce-reyes et al., 2012; Toledo-Aceves et al., 2011); thus, species traits can aid in understanding and predicting shifts in biodiversity and associated ecosystem functions and services.

The IUCN categorizes species into risk categories to assess extinction risk and help prioritize species and habitat conservation (IUCN, 2022b). Despite the increase in and utility of trait-based approaches in assessing extinction risk and vulnerability (Foden et al., 2013; Kosman et al., 2019; Pacifici et al., 2015), setting spatial priorities for parks and reserves (Kukkala & Moilanen, 2017), and mapping diversity patterns (Cadotte & Tucker, 2018; Devictor et al., 2010), the IUCN does not explicitly incorporate trait-based approaches into the official IUCN Red List assessment process (IUCN, 2022b). Instead, the IUCN uses a population trend-based approach (changes in abundance or current and potential changes in distributions; Etard et al., 2020), relying on information often unavailable for many species. Of the species contained in Frugivoria, 12.9% are classified as threatened by the IUCN, with 10.1% of lowland species and 16.4% of montane species having threat categories of either vulnerable, endangered or critically endangered. The higher proportion of threatened montane species is not unexpected, since montane ecosystems are particularly vulnerable to threats such as climate change and fragmentation, and species in these regions often have restricted dispersal capabilities.

Of mammals specifically, 14.6% are classified as DD, because they lack one or both population and distributional (range size; EOO; IUCN, 2022a) information to make an official extinction risk assessment. Species such as this provide an opportunity to test trait-based approaches for assessing conservation statuses.

Trait-based approaches to conservation offer an alternative approach to trend-based approaches because trait-based approaches rely on species' sensitivity to particular threats (Etard et al., 2020). If the response of species to a threat consistently aligns with certain traits (e.g. both narrow diet breadth, which indicates a specialized diet, and large-bodied and small-ranged species being potential indicators of extinction risk; Boyles & Storm, 2007; Harris & Pimm, 2008; Ripple et al., 2017), traits can be used as part of a proactive approach to generalize patterns or set rules for assigning extinction risk for species that do not have the sufficient population or geographical data (Cardillo & Meijaard, 2012) or can be used in addition to trend-based approaches. For instance, Bland et al. (2015) estimated the extinction risk of DD terrestrial mammals using predictive models based on life history, geographical ranges and information on threats, increasing the estimate of globally threatened terrestrial mammals by 5%. Frugivoria has the potential to be used for this purpose for frugivorous species in the montane Neotropics, as traits such as habitat specialization (strong reliance on cloud forest habitat), diet breadth (based on the Shannon Index and demonstrating the degree of diet specialization), and range size can indicate levels of risk to anthropogenic pressures such as habitat degradation and climate change (Bland et al., 2015). For example, range size can be used to infer sensitivity to environmental change because narrow-ranged species, which often have very little data, are more sensitive to anthropogenic disturbances and tend to have high extinction risk than those with broad ranges (Collen et al., 2016; Ripple et al., 2017). Furthermore, in Frugivoria, we include a geographical trait—the average human footprint—explicitly quantifying anthropogenic forces acting across the species range. As such, it would be straightforward to use Frugivoria to generate lists of potentially at-risk species in this region using these traits listed above. This dataset can be used in applied conservation to set targets for maintaining areas of high functional diversity and spatially prioritizing regions containing species with unique ecosystem roles.

The species currently in Frugivoria cover an important and highly diverse region and habitat type, yet it does not provide a complete picture of Central and South American frugivore biodiversity. Importantly, it excludes explicit incorporation of countries lacking contiguous mountain ranges, for example in Southeastern Brazil, the IUCN designated habitats 'subtropical/tropical dry forest' (e.g. seasonally dry inter-Andean valleys) and 'tropical high altitude' (e.g. páramo; regions above the timberline). Frugivoria has been compiled in an open and reproducible way that facilitates its future expansion to high-altitude montane and lowland regions of Central and South America—an essential undertaking for gaining a more complete picture of trait diversity for frugivorous species in the Neotropics. Though the species contained in Frugivoria are only a subset of the world's vast biodiversity, we hope the data and workflow facilitate further study in the Neotropics and beyond, as filling in more data gaps in hotspots like these continues to be a research priority.

ACKNOWLEDGEMENTS

We thank the following students for their assistance in data entry: Arpita Nayak, Minali Bhatt, Erik Ralston, Olivia Melville, Taylor Stehouwer, Aaron Guggenheimer, Ashley Ezzo, Chakata Hart, Caroline Blommel, Alison McCormick, Kate Kelly, Nellie Chalem and Emma Mushaka. Funding was provided by NASA FINESST Grant #80NSSC19K1332. National Science Foundation (CC*Compute: #2200792). This work was supported in part by computational

resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript. We certify that the submission is original work and is not under review at any other publication.

DATA AVAILABILITY STATEMENT

The database and associated metadata are openly available as .csv files through the Environmental Data. Initiative (EDI; https://doi.org/10.6073/pasta/168e95f04d4726d31d868bfe22d749a5) (Gerstner et al. 2023). We also provide the R scripts used for the database building workflow and analysis for this manuscript on EDI and GitHub (https://github.com/bioXgeo/neotropical_frugivores.git).

ORCID

Beth E. Gerstner https://orcid.org/0000-0001-6443-3487
Patrick Bills https://orcid.org/0000-0003-4235-255X
Phoebe L. Zarnetske https://orcid.org/0000-0001-6257-6951

REFERENCES

Ansell, W. F. H. (1965). Standardization of field data on mammals. *Zoologica Africana*, *1*, 97–113.

Armenteras, D., Rodríguez, N., Retana, J., & Morales, M. (2011). Understanding deforestation in montane and lowland forests of the Colombian Andes. *Regional Environmental Change*, *11*, 693–705.

Bellard, C., Leclerc, C., Leroy, B., Bakkenes, M., Veloz, S., Thuiller, W., & Courchamp, F. (2014). Vulnerability of biodiversity hotspots to global change. *Global Ecology and Biogeography*, *23*, 1376–1386.

Billerman, S. M., Keeney, B. K., Rodewald, P. G., & Schulenberg, T. S. (Eds.). (2022). *Birds of the world*. Cornell Laboratory of Ornithology. https://birdsoftheworld.org/bow/home

Bland, L. M., Collen, B., Orme, C. D. L., & Bielby, J. (2015). Predicting the conservation status of data-deficient species. *Conservation Biology*, *29*, 250–259.

Boyles, J. G., & Storm, J. J. (2007). The perils of picky eating: Dietary breadth is related to extinction risk in insectivorous bats. *PLoS One*, *2*, e672.

Brandon, K. (2014). *Ecosystem services from tropical forests: Review of current science*. Social Science Research Network.

Burt, W. H. (1943). Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy*, *24*, 346–352.

Burton, M., & Robert, B. (1974). *The Funk & Wagnalls wildlife encyclopedia*. Funk and Wagnalls.

Cadotte, M. W., Carscadden, K., & Mirotchnick, N. (2011). Beyond species: Functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology*, *48*, 1079–1087.

Cadotte, M. W., & Tucker, C. M. (2018). Difficult decisions: Strategies for conservation prioritization when taxonomic, phylogenetic and functional diversity are not spatially congruent. *Biological Conservation*, *225*, 128–133.

Cardillo, M., & Meijaard, E. (2012). Are comparative studies of extinction risk useful for conservation? *Trends in Ecology & Evolution*, *27*, 167–171.

Center for International Earth Science Information Network – CIESIN – Columbia University, and Centro Internacional de Agricultura Tropical – CIAT. (2005). *Gridded population of the world, version 3 (GPWv3): Population density grid*. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4XK8CG2

Collen, B., Dulvy, N. K., Gaston, K. J., Gärdenfors, U., Keith, D. A., Punt, A. E., Regan, H. M., Böhm, M., Hedges, S., Seddon, M., Butchart, S. H. M., Hilton-Taylor, C., Hoffmann, M., Bachman, S. P., & Akçakaya, H. R. (2016). Clarifying misconceptions of extinction risk assessment with the IUCN red list. *Biology Letters*, *12*, 20150843.

Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: Addressing disparity in global monitoring. *Tropical Conservation Science*, *1*, 75–88.

Cooke, R. S. C., Bates, A. E., & Eigenbrod, F. (2019). Global trade-offs of functional redundancy and functional dispersion for birds and mammals. *Global Ecology and Biogeography*, *28*, 484–495.

Dehling, D. M., Töpfer, T., Schaefer, H. M., Jordano, P., Böhning-Gaese, K., & Schleuning, M. (2014). Functional relationships beyond species richness patterns: Trait matching in plant–bird mutualisms across scales. *Global Ecology and Biogeography*, *23*, 1085–1093.

Devictor, V., Mouillot, D., Meynard, C., Jiguet, F., Thuiller, W., & Mouquet, N. (2010). Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: The need for integrative conservation strategies in a changing world. *Ecology Letters*, *13*, 1030–1040.

Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., Hansen, M., Locke, H., Ellis, E. C., Jones, B., Barber, C. V., Hayes, R., Kormos, C., Martin, V., Crist, E., … Saleem, M. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience*, *67*, 534–545.

Eisenberg, J. F., & Redford, K. H. (1999). *Mammals of the neotropics* (Vol. *3*). University of Chicago Press.

Elsen, P. R., & Tingley, M. W. (2015). Global mountain topography and the fate of montane species under climate change. *Nature Climate Change*, *5*, 772–776.

Emmons, L., & Feer, F. (1997). *Neotropical rainforest mammals: A field guide*. University of Chicago Press.

Engemann, K., Enquist, B. J., Sandel, B., Boyle, B., Jørgensen, P. M., Morueta-Holme, N., Peet, R. K., Violle, C., & Svenning, J.-C. (2015). Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecology and Evolution*, *5*, 807–820.

Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global Ecology and Biogeography*, *29*, 2143–2158.

Farwig, N., & Berens, D. G. (2012). Imagine a world without seed dispersers: A review of threats, consequences and future directions. *Basic and Applied Ecology*, *13*, 109–115.

Faurby, S., Davis, M., Pedersen, R. O., Schowanek, S. D., Antonelli, A., & Svenning, J.-C. (2018). PHYLACINE 1.2: The phylogenetic atlas of mammal macroecology. *Ecology*, *99*, 2626.

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, *51*, 331–363.

Fleagle, J. G. (2013). *Primate adaptation and evolution*. Academic Press.

Foden, W. B., Butchart, S. H. M., Stuart, S. N., Vié, J.-C., Akçakaya, H. R., Angulo, A., DeVantier, L. M., Gutsche, A., Turak, E., Cao, L., Donner, S. D., Katariya, V., Bernard, R., Holland, R. A., Hughes, A. F., O'Hanlon, S. E., Garnett, S. T., Şekercioğlu, Ç. H., & Mace, G. M. (2013). Identifying the World's Most climate change vulnerable species: A systematic trait-based assessment of all birds, amphibians and corals. *PLoS ONE*, *8*, e65427.

Foster, P. (2001). The potential negative impacts of global climate change on tropical montane cloud forests. *Earth-Science Reviews*, *55*, 73–106.

Fricke, E. C., Ordonez, A., Rogers, H. S., & Svenning, J.-C. (2022). The effects of defaunation on plants' capacity to track climate change. *Science*, *375*, 210–214.

Gaston, K. J., & Fuller, R. A. (2009). The sizes of species' geographic ranges. *Journal of Applied Ecology*, *46*, 1–9.

GBIF: The Global Biodiversity Information Facility. (2023). What is GBIF? https://www.gbif.org/what-is-gbif

Gómez, C., Tenorio, E. A., & Cadena, C. D. (2021). Change in avian functional fingerprints of a neotropical montane forest over 100 years as an indicator of ecosystem integrity. *Conservation Biology*, *35*, 1552–1563.

González-del-Pliego, P., Freckleton, R. P., Edwards, D. P., Koo, M. S., Scheffers, B. R., Pyron, R. A., & Jetz, W. (2019). Phylogenetic and trait-based prediction of extinction risk for data-deficient amphibians. *Current Biology*, *29*, 1557–1563.e3.

González-Maya, J. F., Martínez-Meyer, E., Medellín, R., & Ceballos, G. (2017). Distribution of mammal functional diversity in the neotropical realm: Influence of land-use and extinction risk. *PLoS One*, *12*, e0175931.

González-Suárez, M., & Revilla, E. (2013). Variability in life-history and ecological traits is a buffer against extinction in mammals. *Ecology Letters*, *16*, 242–251.

Gradstein, S. R., Homeier, J., & Gansert, D. (Eds.). (2008). *The tropical mountain forest: Patterns and processes in a biodiversity hotspot*. Göttingen University Press.

Guralnick, R., & Constable, H. (2010). VertNet: Creating a data-sharing community. *Bioscience*, *60*, 258–259.

Harris, G., & Pimm, S. L. (2008). Range size and extinction risk in Forest birds. *Conservation Biology*, *22*, 163–171.

Howe, H. F., & Smallwood, J. (1982). Ecology of seed dispersal. *Annual Review of Ecology and Systematics*, *13*, 201–228.

IPCC. (2022). In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate change 2022: Impacts, adaptation, and vulnerability. Contribution of working group II to the sixth assessment report of the intergovernmental panel on climate change* (3056). Cambridge University Press. https://doi.org/10.1017/9781009325844

IUCN. (2015). The IUCN red list of threatened species. Version 2022-2. https://www.iucnredlist.org

IUCN. (2021). The IUCN Red List of Threatened Species. Version 2021. https://www.iucnredlist.org

IUCN. (2022a). The IUCN red list of threatened species. Version 2022-2. https://www.iucnredlist.org

IUCN. (2022b). IUCN Standards and Petitions Committee. Guidelines for Using the IUCN Red List Categories and Criteria. Version 15.1. Prepared by the Standards and Petitions Committee. https://www.iucnredlist.org/documents/RedListGuidelines.pdf

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., … Purvis, A. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, *90*, 2648.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*, 170122.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2018). Data from: Climatologies at high resolution for the earth's land surface areas. *Dryad Digital Repository* https://doi.org/10.5061/dryad.kd1d4

Kissling, W. D., Walls, R., Bowser, A., Jones, M. O., Kattge, J., Agosti, D., Amengual, J., Basset, A., van Bodegom, P. M., Cornelissen, J. H. C., Denny, E. G., Deudero, S., Egloff, W., Elmendorf, S. C., Alonso García, E., Jones, K. D., Jones, O. R., Lavorel, S., Lear, D., … Guralnick, R. P. (2018). Towards global data products of essential biodiversity variables on species traits. *Nature Ecology & Evolution*, *2*, 1531–1540.

Kosman, E., Burgio, K. R., Presley, S. J., Willig, M. R., & Scheiner, S. M. (2019). Conservation prioritization based on trait-based metrics illustrated with global parrot distributions. *Diversity and Distributions*, *25*, 1156–1165.

Kukkala, A. S., & Moilanen, A. (2017). Ecosystem services and connectivity in spatial conservation prioritization. *Landscape Ecology*, *32*, 5–14.

Lepage, D., Vaidya, G., & Guralnick, R. (2014). Avibase – A database system for managing and organizing taxonomic concepts. *ZooKeys*, *420*, 117–135.

Lim, J. Y., Svenning, J.-C., Göldel, B., Faurby, S., & Kissling, W. D. (2020). Frugivore-fruit size relationships between palms and mammals reveal past and future defaunation impacts. *Nature Communications*, *11*, 4904.

Miralles, A., Bruy, T., Wolcott, K., Scherz, M. D., Begerow, D., Beszteri, B., Bonkowski, M., Felden, J., Gemeinholzer, B., Glaw, F., Gloeckner, F. O., Hawlitschek, O., Kostadinov, I., Nattkemper, T. W., Printzen, C., Renz, J., Rybalka, N., Stadler, M., Weibulat, T., … Vences, M. (2020). Repositories for taxonomic data: Where we are and what is missing. *Systematic Biology*, *69*, 1231–1253.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858.

Myers, P., Espinosa, R., Parr, C. S., Jones, T., Hammond, G. S., & Dewey, T. A. (2023). The Animal Diversity Web (online). https://animaldiversity.org

Nelson, G., & Ellis, S. (2019). The history and impact of digitization and digital data mobilization on biodiversity research. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*, 20170391.

Newbold, T., Scharlemann, J. P. W., Butchart, S. H. M., Şekercioğlu, Ç. H., Joppa, L., Alkemade, R., & Purves, D. W. (2014). Functional traits, land-use change and the structure of present and future bird communities in tropical forests. *Global Ecology and Biogeography*, *23*, 1073–1084.

Pacifici, M., Foden, W. B., Visconti, P., Watson, J. E. M., Butchart, S. H. M., Kovacs, K. M., Scheffers, B. R., Hole, D. G., Martin, T. G., Akçakaya, H. R., Corlett, R. T., Huntley, B., Bickford, D., Carr, J. A., Hoffmann, A. A., Midgley, G. F., Pearce-Kelly, P., Pearson, R. G., Williams, S. E., … Rondinini, C. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, *5*, 215–224.

Parr, C. S., Wilson, N., Leary, P., Schulz, K. S., Lans, K., Walley, L., Hammock, J. A., Goddard, A., Rice, J., Studer, M., Holmes, J. T. G., & Corrigan, R. J., Jr. (2014). The encyclopedia of life v2: Providing global access to knowledge about life on earth. *Biodiversity Data Journal*, *2*, e1079. https://doi.org/10.3897/BDJ.2.e1079

Pollock, L. J., Thuiller, W., & Jetz, W. (2017). Large conservation gains possible for global biodiversity facets. *Nature*, *546*, 141–144.

Ponce-reyes, R., Reynoso-rosales, V., Watson, J. E. M., Vanderwal, J., Fuller, R. A., Pressey, R. L., & Possingham, H. P. (2012). Vulnerability of cloud forest reserves in Mexico to climate change. *Nature Climate Change*, *2*, 448–452.

Powers, R. P., & Jetz, W. (2019). Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. *Nature Climate Change*, *9*, 323–329.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramesh, V., Gopalakrishna, T., Barve, S., & Melnick, D. J. (2017). IUCN greatly underestimates threat levels of endemic birds in the Western Ghats. *Biological Conservation*, *210*, 205–221.

Rehm, E. M., & Feeley, K. J. (2015). The inability of tropical cloud forest species to invade grasslands above treeline during climate change: Potential explanations and consequences. *Ecography*, *38*, 1167–1175.

Ripple, W. J., Wolf, C., Newsome, T. M., Hoffmann, M., Wirsing, A. J., & McCauley, D. J. (2017). Extinction risk is most acute for the world's largest and smallest vertebrates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 10678–10683.

Sales, L. P., Kissling, W. D., Galetti, M., Naimi, B., & Pires, M. M. (2021). Climate change reshapes the eco-evolutionary dynamics of a neotropical seed dispersal system. *Global Ecology and Biogeography*, *30*, 1129–1138.

Sanderson, E. W., Fisher, K., Robinson, N., Sampson, D., Duncan, A., & Royte, L. (2022). The march of the human footprint. *EcoEvoRxiv*.

Santillán, V., Quitián, M., Tinoco, B. A., Zárate, E., Schleuning, M., Böhning-Gaese, K., & Neuschulz, E. L. (2019). Different responses of taxonomic and functional bird diversity to forest fragmentation across an elevational gradient. *Oecologia*, *189*, 863–873.

Santini, L., González-Suárez, M., Russo, D., Gonzalez-Voyer, A., Hardenberg, A., & Ancillotto, L. (2018). One strategy does not fit all: determinants of urban adaptation in mammals. *Ecology Letters*, *22*(2), 365–376. https://doi.org/10.1111/ele.13199.

Sechrest, W. (2003). *Global diversity, endemism and conservation of mammals*. University of Virginia.

Sekercioğlu, C. H., Daily, G. C., & Ehrlich, P. R. (2004). Ecosystem consequences of bird declines. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 18042–18047.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Soria, C. D., Pacifici, M., Marco, M. D., Stephen, S. M., & Rondinini, C. (2021). COMBINE: A coalesced mammal database of intrinsic and extrinsic traits. *Ecology*, *102*, e03344.

Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D., Wang, J., Fraifeld, V. E., & de Magalhães, J. P. (2018). Human ageing genomic resources: New and updated databases. *Nucleic Acids Research*, *46*, D1083–D1090.

Thorington, R. W., Koprowski, J. L., Steele, M. A., & Whatton, J. F. (2012). *Squirrels of the world*. Johns Hopkins University Press.

Tóbon, C., Sevink, J., & Verstraten, J. M. (2004). Solute fluxes in throughfall and stemflow in four Forest ecosystems in Northwest Amazonia. *Biogeochemistry*, *70*, 1–25.

Toledo-Aceves, T., Meave, J. A., González-Espinosa, M., & Ramírez-Marcial, N. (2011). Tropical montane cloud forests: Current threats and opportunities for their conservation and sustainable management in Mexico. *Journal of Environmental Management*, *92*, 974–981.

Waller, J., & Grosjean, M. (2019). Downloading occurrences from a long list of species in R and python. *GBIF Data Blog*. https://data-blog.gbif.org/post/downloading-long-species-lists-on-gbif/

Wilkie, D. S., Bennett, E. L., Peres, C. A., & Cunningham, A. A. (2011). The empty forest revisited. *Annals of the new York Academy of Sciences*, *1223*, 120–128.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.

Williams, J. W., Jackson, S. T., & Kutzbach, J. E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 5738–5742.

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, *95*, 2027.

Wilson, A. M., & Jetz, W. (2016). Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions. *PLoS Biology*, *14*, e1002415.

Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 8041–8048.

BIOSKETCH

Beth Gerstner is a Ph.D. candidate at Michigan State University where her work focuses on ways to improve conservation assessments for data-poor species, leveraging species functional traits and high-resolution environmental data. She earned her MS in Biology at the City College of New York, where she generated distributional estimates for a newly discovered species, the Olinguito.

How to cite this article: Gerstner, B. E., Bills, P., & Zarnetske, P. L. (2023). Frugivoria: A trait database for birds and mammals exhibiting frugivory across contiguous Neotropical moist forests. *Global Ecology and Biogeography*, *32*, 1466–1484. https://doi.org/10.1111/geb.13716