Dietary abundance distributions: Dominance and diversity in vertebrate diets

Matthew C. Hutchinson¹,² | Andrew P. Dobson¹ | Robert M. Pringle¹

Abstract
Diet composition is among the most important yet least understood dimensions of animal ecology. Inspired by the study of species abundance distributions (SADs), we tested for generalities in the structure of vertebrate diets by characterising them as dietary abundance distributions (DADs). We compiled data on 1167 population-level diets, representing >500 species from six vertebrate classes, spanning all continents and oceans. DADs near-universally (92.5%) followed a hollow-curve shape, with scant support for other plausible rank-abundance-distribution shapes. This strong generality is inherently related to, yet incompletely explained by, the SADs of available food taxa. By quantifying dietary generalisation as the half-saturation point of the cumulative distribution of dietary abundance (sp50, minimum number of foods required to account for 50% of diet), we found that vertebrate populations are surprisingly specialised: in most populations, fewer than three foods accounted for at least half the diet. Variation in sp50 was strongly associated with consumer type, with carnivores being more specialised than herbivores or omnivores. Other methodological (sampling method and effort, taxonomic resolution), biological (body mass, frugivory) and biogeographic (latitude) factors influenced sp50 to varying degrees. Future challenges include identifying the mechanisms underpinning the hollow-curve DAD, its generality beyond vertebrates, and the biological determinants of dietary generalisation.

KEYWORDS
community ecology, diet selection, feeding ecology, food webs, generalist–specialist continuum, latitudinal gradient in niche breadth, macroecology, optimal foraging theory, species abundance distributions (SADs), trophic interaction networks

INTRODUCTION
Interactions between individual organisms constitute the web of life. Innumerable encounters between foraging consumers and the organisms they eat result in transfers of nutrients and energy; when integrated over space, time, and levels of biological organisation, these interactions define the diets of individuals, populations, species, and guilds. Dietary diversity, and the generalist–specialist spectrum that emerges from variation therein, play pivotal roles in ecology and evolution. Food-web structure and stability are influenced by the dietary diversity of the constituent species (Bascompte & Jordano, 2007; May, 1972). Overlap in diet composition modulates competition within and between species (Boag & Grant, 1984; Svanbäck & Bolnick, 2007; Tinker et al., 2008), and dietary flexibility affects the ability of populations to withstand environmental change (Ducatez et al., 2020; Reed & Tosh, 2019). Evolutionary differentiation and speciation have been linked to shifts in diet selection and specialisation (Grant & Grant, 2006; Riesch et al., 2012), and diet breadth can regulate the strength of
Despite its importance, dietary diversity is poorly characterised (Pringle & Hutchinson, 2020). There are various taxon-specific ideas about the determinants and dimensionality of dietary diversity (e.g. Freeland, 1991; Freeland & Janzen, 1974; Kartzinel & Pringle, 2020; Sih & Moore, 1990; Westoby, 1974) and general theories about why animals should forage in particular ways given energetic constraints (Stephens & Krebs, 1986). Despite this, many fundamental questions remain unanswered. Are there consistent patterns in dietary diversity or its drivers that hold across taxa and regions? Does dietary diversity differ in particular ways across trophic levels or environmental gradients? How sensitive are inferences about dietary diversity to the methods used to document diets? Synthetic approaches are necessary to answer these questions (Halpern et al., 2020).

Macroecological studies of community composition offer one model for a synthetic outlook on dietary diversity. Through the formalism of species-abundance distributions (SADs), investigators have probed relationships between biodiversity and factors such as latitude (Soininen et al., 2012; Ulrich et al., 2016), elevation (Escobar et al., 2007; Whittaker, 1960), and species’ traits (Cotgreave & Harvey, 1994). Conveniently, data describing the composition of animals’ diets are directly analogous to those describing ecological communities. Vertebrates are mobile animals that can feed on many co-occurring species; their diets are ‘communities’ that have passed through both environmental and esophageal filters. Thus, the macroecological frameworks and statistical methods used for free-living communities can equally be applied to dietary diversity. A hallmark of SADs is their adherence to a group of closely related ‘hollow-curve’ functions (McGill et al., 2007).

Community-assembly rules based on niche and neutral processes have been proposed to explain this consistency (Connolly et al., 2014; Enquist et al., 2019; McGill, 2010). Similarly consistent patterns in dietary ‘communities’ might suggest that similarly general rules govern animal foraging.

To explore this possibility, we examined the diets of 1167 globally distributed vertebrate populations (Figure 1) through the lens of dietary abundance distributions (DADs). Our unit of analysis is the population-level average diet: the mean representation of different food taxa eaten by a population of animals at a location, aggregated across individuals and foraging bouts; we define DADs as rank-abundance distributions that chart the relationship between each resource taxon’s rank in the population-level diet (most commonly eaten item = rank 1) and its relative abundance in that diet. By reducing diet composition from a heterogeneous set of resources into the vector of those resources’ contribution to the population-level diet, DADs bypass resource...
identities and the contingencies of local ecological context, which facilitates comparative assessment of diet structure across diverse taxonomic groups, geographies, and methods. We addressed three main questions: (1) Do vertebrate DADs follow a consistent shape? (2) What do DADs reveal about trophic specialisation in vertebrate populations? (3) How do key biological and biogeographical variables, along with methodological factors such as sampling effort, influence estimates of dietary specialisation? In addition, we conducted a preliminary evaluation of the relationship between SADs and DADs. Although DADs are inherently constrained by SADs, it does not necessarily follow that DADs mirror SADs, because most animals do not filter food randomly from the environment. We therefore explored the SAD-DAD relationship using case studies of consumers from well-studied systems.

**METHODS**

**Dataset assembly**

We identified studies using Google Scholar searches for the terms “feeding ecology” or “diet” and a haphazardly chosen vertebrate taxon (either a taxonomic family or a common name). We detail our approach, including caveats and potential biases, in Text S1. An overview of the dataset is in Figure 1 and Text S1. We recorded the following metadata for each diet to test a series of general hypotheses arising from the three main questions listed above.

1. **Sampling method**: ‘Faecal’, ‘stomach’ or ‘observation’. ‘Faecal’ included analyses of plant microhistology, prey parts, and (rarely) DNA. ‘Stomach’ included gastric lavage, gut contents, and regurgitated pellets. ‘Observation’ included focal watches or examinations of feeding sites. Sampling method clearly has potential to influence inferences about diet breadth. On the one hand, data collected post-ingestion and at successively later stages of digestion might yield steeper DADs (lower breadth) because food items may be harder to identify (Bowen & Iverson, 2012); on the other hand, observations might underestimate diet breadth because many feeding events go undetected (Wirta et al., 2014) and/or food items are difficult to identify in real time (Pringle & Hutchinson, 2020).

2. **Sample size**: Number of biological replicates. For faecal and stomach contents, this was the number of discrete samples. For observations, this was the number of bouts (if reported in days, we considered 1 day equal to one bout). If recorded only as the number of feeding observations (e.g. kills, bites), sample size was left as NA. This approach aimed to maximise comparability of sample sizes from different sampling methods. Given the influence of sample size on species richness (Gotelli & Colwell, 2001) and interaction diversity (Pringle & Hutchinson, 2020), we hypothesised that larger sample sizes would yield flatter DADs (more generalised diets).

3. **Intra-annual extent**: Temporal extent of a study within a year, a proxy for cross-seasonal representation. We considered a month to be 30 days (thus, 360 days year\(^{-1}\)). We ignored changes in year; a study from November to April, inclusive, would have an intra-annual extent of 210 days. We expected a greater intra-annual extent to flatten DADs, because seasonal variation in resource availability and foraging behaviour can diversify diets (McMeans et al., 2019).

4. **Inter-annual extent**: Number of years in which each study sampled, again ignoring changes in the calendar year; a November–April study would be 1 year; a study comprising two November–April sampling periods would be 2 year. When a study extended >360 days but sampling periods were inconsistent, inter-annual extent took non-integer values. For example a study spanning 14 consecutive months would be recorded as 14/12 = 1.17 year. Similar to intra-annual extent, we expected greater inter-annual extent to flatten DADs by encompassing year-to-year fluctuations in resource availability (Fiedler, 2002).

5. **Taxonomic resolution**: Proportion of diet resolved to species, genus, family, and above-family levels. We considered foods resolved to species only when a Latin binomial or equivalent common name was specified (e.g. Mus sp1 and Mus sp2 would be considered genus-level identifications but separate taxa). We computed resolution as: \(4 \times \text{species-level proportion} + (3 \times \text{genus-level proportion}) + (2 \times \text{family-level proportion}) + \text{above-family proportion}\), yielding a continuous index from 1 to 4, where 4 is a diet resolved entirely to species. We hypothesised that coarse resolution would yield steeper DADs because it collapses dietary diversity into fewer categories (Pringle & Hutchinson, 2020).

6. **Environment type**: Terrestrial, freshwater, or marine. These classifications were mostly unambiguous. In rare cases, they were determined by diet composition. For example, coastal birds were classified as terrestrial if most prey were land animals or marine if the reverse. We had no *a priori* expectation about the impact of environment type on DAD shape but included it as the basal ecological context of consumer–resource interactions.

7. **Consumer type**: Herbivorous, omnivorous, or carnivorous, based on proportion of diet derived from plant material. We used the following arbitrary thresholds: diets with ≥90% plant were herbivorous, diets with ≤10% plant were carnivorous, and anything in between was omnivorous. We expected herbivores to have flatter DADs and more generalised diets than carnivores because, in general, herbivores should be
less constrained by pre-ingestion factors such as capture and handling costs that promote specialisation (Clements et al., 2014; Wilson et al., 2018) but more restricted by post-ingestion detoxification (Freeland & Janzen, 1974) and nutrient extraction (Clauss et al., 2013). We expected omnivores to have the flattest DADs because eating both plants and animals should enable greater diet breadth.

8. **Frugivory presence**: Presence of fruit in diet. When studies distinguished pulp and seed consumption, or specified ripeness, we only recorded frugivory as ripe-pulp consumption. We expected fruit consumption to flatten DADs because fruits are often a low-cost dietary addition but do not provide a nutritionally balanced diet; they have low handling costs and high levels of non-structural carbohydrates and lipids but low protein content (Jordan, 2000), meaning that vertebrates should generally require additional foods to balance nutrient intake.

9. **Location**: Study locale. We used the `geocode` function in R package `ggmap` (v3.0.0, Kahle & Wickham, 2013) to estimate population latitude and longitude. If location was only reported to a geopolitical level (e.g., state, country), `geocode` returns the centroid of this location. Following the latitude-niche breadth hypothesis (MacArthur, 1972), we expected DADs to be steeper and diets more specialised in the tropics.

10. **Consumer body mass**: We used data for mammals from PANTHERIA (Jones et al., 2009) and Elton Traits (Wilman et al., 2014); birds from Elton Traits; amphibians from AmphIBIO (Oliveira et al., 2017); reptiles from Myhrvold et al. (2015); and fishes from FishBase (https://www.fishbase.org/), accessed via Encyclopedia of Life (https://eol.org/; June 2021). We found species-level values for 1117 diets; we used a genus-level mean for 36 more and a family-level mean for 14. The R (v3.6.3; R Core Team, 2020) script used to assign body-mass data is deposited alongside other code and data on Dryad [https://doi.org/10.5061/dryad.08kpr53v]. We expected larger-bodied species to have flatter DADs and more generalised diets due to the associations between body mass and home-range size, gape size, prey range and total biomass ingestion (Clauss et al., 2013; Gordon & Illius, 1988; McNab, 1963; Sinclair et al., 2003).

### Describing DAD shape

We tested how well five generic, qualitatively different shapes for rank-abundance distributions approximated empirical DADs (Figure 2a–e): (i) a negative-exponential function representing few common and many rare foods (‘hollow curve’, Figure 2a); (ii) a logistic function representing a step-change in food rank abundance (‘staples and supplements’, Figure 2b); (iii) a negative linear function representing a constant decrease in rank abundance (‘linear hierarchy’, Figure 2c); (iv) a horizontal line indicating all resources are eaten in equal amounts (‘all equal’, Figure 2d); and (v) a concave-down parabolic function representing many common and few rare foods (‘weak hierarchy’, Figure 2e). We considered each of these shapes theoretically plausible, although we do not consider them to be diagnostic of or unique to any particular mechanism. For example, all-equal or weak-hierarchy shapes might arise if herbivore dietary diversity is limited by Species-specific, dose-dependent toxins (Freeland & Janzen, 1974); but the same mechanism could create hollow-curve, staples-and-supplements, or linear-hierarchy DADs if herbivores have greater tolerance/resistance to particular toxins. We fit the following functions to each DAD using `nlsLM` in `minpack.lm` (v1.2.1; Elzhov et al., 2016), listed in order of increasing complexity:

- **(All equal; Figure 2d)**
  \[
  y = a
  \]
  where \(a\) is the \(y\)-intercept (one free parameter);
- **(Linear hierarchy; Figure 2c)**
  \[
  y = a + bx
  \]
  where \(a\) is the \(y\)-intercept and \(b\) is the slope (two free parameters);
- **(Hollow curve; Figure 2a)**
  \[
  y = e^{(x-d)} + f
  \]
  where \(0 \leq c \leq 1, -\infty < f < \infty, \text{ and } -\infty < d < \infty\) (three free parameters);
- **(Weak hierarchy; Figure 2e)**
  \[
  y = - \left( \frac{(x-1)^g}{h} - i \right)
  \]
  where \(0 \leq h, 0 \leq i \leq 1 \text{ and } g \geq 2\) (three free parameters);
- **(Staples and supplements; Figure 2b)**
  \[
  y = \frac{j}{1 + e^{k(x-m)}} + n
  \]
  where \(0 \leq j \leq 1, k \geq 1, m \geq 4, \text{ and } 0 \leq n \leq 1\) (four free parameters).

In these formulae, \(y\) is the proportional contribution of a food to the diet and \(x\) is its dietary rank. We used bounds on parameter estimates (as reported) to restrict models to the desired shape; shapes could still converge at parameter bounds (e.g., ‘staples and supplements’ is a horizontal line when \(j = 0\), but in these cases, our evaluation criterion favours the simpler model. Because median dietary richness was 17 (IQR = 11–27), we compared shapes using the small-sample Akaike information criterion (AICc; Baldridge et al., 2016) and considered DADs to be best described by one shape when \(\Delta\text{AICc} > 2\) for the next-best model (Burnham & Anderson, 1998).
To avoid overfitting DADs with low richness, we did not analyse DADs with ≤4 food items (n = 37, leaving n = 1130). Similarly, because each additional free parameter requires an additional degree of freedom for fitting, the minimum richness of DADs fit to a given shape increased with model complexity. Thus, the staples-and-supplements’ shape with four free parameters, was only fit to DADs with ≥7 foods (n = 1095).

Axes of variation in DAD shape

We characterised dietary generalisation using the cumulative distribution of each DAD (the relationship between the rank order of foods, from commonest to rarest, and their cumulative proportional contribution to the diet). As an index, we calculated the half-saturation point of this distribution (sp50, the minimum number of foods required to account for 50% of diet; Figure 2a–e). Diets dominated by few foods have small sp50, indicating specialisation; diets with a balanced mix of foods have larger sp50, indicating generalisation. Because sp50 is likely to fall between integer food ranks, we calculated it from the cumulative DAD by identifying the ranks of the two foods whose cumulative contribution was immediately above and below 50% and solving the linear model between these two points for x (food rank) at y = 0.5. For example, if the top two foods account for
45% of the diet and the top three account for 52%, we fit a line \( (y = \text{intercept} + \text{slope} \times x) \) between (2, 0.45) and (3, 0.52), giving an intercept of 0.31 and slope of 0.07; solving for \( x \) at \( y = 0.5 \) gives (0.5–0.31)/0.07, thus \( sp50 = 2.71 \). While common biodiversity metrics were correlated with \( sp50 \) (Figure S1), we used \( sp50 \) because it is more intuitive than diversity and evenness, more holistic than dominance indices based on the few commonest species (Maurer & McGill, 2011), and sidesteps challenges inherent in quantifying richness and its derivatives (Chao & Jost, 2012; Gotelli & Colwell, 2001), which are even more pronounced for diets than for free-living communities (Pringle & Hutchinson, 2020; Wirta et al., 2014). More inclusive indices of dietary generalisation \( (sp75, sp90) \) were strongly correlated with \( sp50 \), suggesting that inferences about the determinants of dietary specialisation are robust to the proportion of diet included (Figure S2).

**Predictors of \( sp50 \)**

To identify correlates of \( sp50 \), we used AICc to evaluate support for 1024 candidate linear models, comprising all additive combinations of the 10 metadata variables described above and the 981 diets reporting all of those variables. We first \( \log_{10} \)-transformed \( sp50 \), body mass, sample size, and inter-annual extent to reduce skew and rescaled continuous variables to mean = 0, SD = 1. Two sets of continuous variables exhibited moderate correlations after transformations (taxonomic resolution and body mass, \( r = 0.51 \); sample size, intra-annual extent, and inter-annual extent, \( r = 0.28–0.42 \)) but collinearity was generally minor (mean \( |r| = 0.17 \) across all pairs of continuous variables) and always below thresholds considered problematic (e.g., \( |r| > 0.7 \); Dormann et al., 2013). We used glmulti (v1.0.8; Calcagno, 2020) to screen the candidate set and calculate relative variable importance (RVI). RVI is the sum of Akaike weights \( (w_i) \) for all models in which a given predictor occurs (Burnham & Anderson, 2002; Giam & Olden, 2016) and ranges from 0 (predictor found only in models with no chance of being the best in the candidate set) to 1 (predictor found in all models with any chance of being the best in the set; Burnham & Anderson, 2002; Giam & Olden, 2016). To test the sensitivity of our results to collinearity among the three sampling-effort variables, we repeated the analysis while replacing these variables with their first principal component as an integrative measure of sampling effort (\( PC_{\text{effort}} \); prcomp function with predictors scaled to unit variance); \( PC_{\text{effort}} \) accounted for 58% of variance in sampling effort (Figure S3).

We used ordinary least-squares regression to evaluate relationships between \( sp50 \) and each predictor. Each relationship was evaluated first for the full dataset, and then for herbivores, carnivores, and omnivores separately; we analysed these subsets because (a) consumer type proved to be the strongest predictor of \( sp50 \), (b) several methodological variables proved to be confounded across consumer types, and (c) there are fundamental differences in foraging ecology across trophic levels. Owing to the low representation of omnivorous diets \( (n = 47) \), results for this category are presented only in Tables S1–S8. In all cases, we \( \log_{10} \)-transformed \( sp50 \) to satisfy regression assumptions, excluded diets not reporting the focal predictor (Tables S1–S8), fit models with \( lm \), and inspected fitting with simulateResiduals \( (DHARMa; \) Hartig, 2020). For categorical predictors (consumer type, sampling method, frugivory presence, environment), we performed post hoc Tukey’s HSD tests using emmeans \( (\) Lenth et al., 2018\). For each subset, we adjusted \( p \)-values using the false-discovery-rate correction \( (Waite & Campbell, 2006) \) except for Tukey HSD tests for within-factor comparisons, which are automatically adjusted \( (\) Lenth et al., 2018\). To confirm that the high prevalence of even-toed ungulates \( (\) Artiodactyla \) in our dataset \( (\) Text S1, Figure 1\) did not bias results, we repeated analyses after randomly reducing artiodactyl representation down to the number of diets in the next most prevalent order \( (\) R seed = 1993\). We also separated diets by both consumer type and sampling method to test whether sampling method mediated the effect of taxonomic resolution on \( sp50 \). Last, we included \( PC_{\text{effort}} \) as a covariate in each model testing a biological predictor to control for effects of sampling effort.

**DADs and SADs case studies**

Because consumer DADs cannot be completely independent of resource SADs, we assessed the extent to which DADs diverged from SADs for six species with diverse foraging modes: reticulated giraffe \( (\) Giraffa camelopardalis \)\); lion \( (\) Panthera leo); scale-throated hermit hummingbird \( (\) Phaethornis eurynome); four-horned chameleon \( (\) Chamaeleo quadricornis); common kingfisher \( (\) Alcedo atthis); and Eurasian otter \( (\) Lutra lutra)\). For giraffes, the food-plant SAD included all tree species >1% relative abundance, which accounted for 82% of the population-level diet \( (\) Kartzinel et al., 2014; Kartzinel & Pringle, 2020\); diet was sampled from 2014 to 2016, whereas the SAD was sampled in 2012 \( (\) Kartzinel et al., 2014\), but tree relative abundance was largely stable over this interval \( (\) Alston et al., 2022\). For lions, chameleons, otters, hummingbirds, and kingfishers, resource SADs and diets were sampled contemporaneously. For lions, we used the relative abundance of prey and the proportion of kills of each species \( (\) Pienaar 1969\). For chameleons, availability was the relative density of invertebrate prey orders \( (\) Hofer et al., 2003\). For otters, relative abundances of fish in both diet and environment were based on biomass \( (\) Blanco-Garrido et al., 2008\). For hummingbirds, the SAD was relative floral abundance of hummingbird-pollinated plants \( (\) Vizentin-Bugoni et al.,
For kingfishers, SAD and DAD were based on relative abundances of fish in both diet and environment (Vilches et al., 2012).

We tested three predictions for each population. (1) DADs and SADs have indistinguishable rank-abundance distributions, evaluated with Kolmogorov–Smirnov tests (ks.test in R). (2) Dietary-abundance ranks are positively correlated with environmental-abundance ranks, evaluated using Spearman’s rank correlation. We resolved ties by randomly assigning consecutive ranks to equally abundant taxa; in these cases, we report the mean correlation coefficient and p-value across 10 iterations (seed = 1984). (3) Observed diet composition differs from unselective environmental sampling. We compared the observed diet-environment Jaccard dissimilarity to the same for 1000 random diets; we constructed random diets by sampling resource taxa with replacement (n = 100) from the environmental SAD, with weights determined by each taxon’s relative abundance and no constraint on dietary richness. We considered observed diets to differ significantly from unselective foraging if the Monte Carlo p-value [(number of random diets with dissimilarity ≥ observed) + 1]/(1000 + 1) < 0.05. For each hypothesis, we adjusted p-values across the set of SAD-DAD comparisons using the false discovery rate (Waite & Campbell, 2006).

RESULTS

DADs have a near-universal shape

For almost all DADs analysed (1084 of 1130), one of the five hypothesised shapes fit best (mean Akaike weight wi = 0.99). The overwhelming majority of these were hollow-curve, negative-exponential functions (92.5%; Figure 2a). The remainder matched the staples-and-supplements (3.6%; Figure 2b), linear-hierarchy (2.0%; Figure 2c), or all-equal shapes (1.8%; Figure 2d). The weak-hierarchy shape never fit best (Figure 2e). DADs best fit by all-equal and linear-hierarchy shapes generally comprised few foods (Figure S4); of the 895 DADs comprising ≥10 foods and best fit by one shape, 1 was linear-hierarchy, 35 (4%) were staples-and-supplements, and 859 (96%) were hollow curves. Forty-six DADs had no clear best fit (ΔAICc <2 for at least one model); for 33 of these, hollow curve and either linear-hierarchy or staples-and-supplements shapes were indistinguishable; for 10, all-equal and linear-hierarchy were indistinguishable; for one, all-equal and hollow curve were indistinguishable; and in two cases, three models were indistinguishable. Hollow-curve DADs did not differ in sample size or inter-annual extent from those best fit by other shapes (Kruskal–Wallis test; sample size, χ² = 1.27, p = 0.26; inter-annual extent, χ² = 0.18, p = 0.68) but had marginally greater intra-annual extent (χ² = 3.68, p = 0.05).

Axes of variation in DADs

In addition to almost universally adhering to a hollow-curve shape, DADs in our dataset were generally steep (low sp50), indicating an unexpected degree of dietary specialisation in vertebrate populations. For the typical population, just two food taxa accounted for half the diet (sp50 mean = 2.3, median = 1.8, IQR = 1.1–2.9). Nearly a quarter of populations (23.5%) had sp50 < 1 (i.e. one food was >50% of the diet), and a further third (32.6%) had 1 ≤ sp50 < 2. By contrast, only 12.3% of populations had sp50 > 4, and merely 1.4% had sp50 > 8. When we limited analysis to diets that reached stringent thresholds of minimum sampling effort (≥20 samples and 6 months, n = 374), sp50 increased only marginally (mean = 2.7, median = 1.9, IQR = 1.2–3.5). Similarly, using a larger proportion of the diet to quantify specialisation (sp75, sp90) did not dramatically alter this conclusion; for most populations, just four foods accounted for 75% of the diet and just seven accounted for 90% of the diet (Figure S2). The largest sp50 in our dataset (22.6) was for forest elephants (Loxodonta cyclotis) in Ghana (Short, 1981).

The best-fitting model of sp50 (wi = 0.08) included consumer type, sampling method, sample size, intra-annual extent, and inter-annual extent; this model explained as much variance as the one including all variables (for both: adjusted R² = 0.11, multiple R² = 0.12). Eight models had ΔAICc values ≤2 from the best model (cumulative wi = 0.43); all of these included consumer type, sampling method, sample size, and intra-annual extent. Consumer type (RVI = 1.00) and three methodological factors—sampling method, intra-annual extent and sample size (RVI = 0.99, 0.99 and 0.86 respectively)—were the most important predictors of sp50 (Figure 3a). Overall, the five methodological factors had greater relative importance than the five biological ones (mean ± SEM RVI: methodological = 0.74 ± 0.14, biological = 0.43 ± 0.15; Figure 3a). Using PC1effort in lieu of sample size, intra-annual extent, and inter-annual extent did not qualitatively alter our conclusions (Figure S3).

While this analysis provides a broad view of the sources of variation in sp50, it is also limited by skewness in the dataset and correlations among predictors—especially those involving consumer type. For example, the two most important predictors of sp50, consumer type and sampling method, were largely confounded: 90% of 222 diets described by direct observation belonged to herbivores, whereas 71% of 382 stomach-contents analyses belonged to carnivores. Likewise, consumer types were not evenly distributed across latitudes or environments. Carnivorous diets had a mean absolute latitude of 36.6° (IQR = 22.4–51.0°), whereas herbivorous (mean = 28.4°, IQR = 12.6–40.5°) and omnivorous diets (mean = 29.8°, IQR = 21.0–41.0°) were more frequently tropical. Most strikingly, marine diets in our dataset were almost exclusively carnivorous (129 of 131); only green sea turtle (Chelonia mydas; Mortimer, 1981) and polar bear...
FIGURE 3  Major correlates of dietary generalisation. (a) RVI of five methodological and five biological predictors of sp50 (log10-transformed for analysis to increase normality and plotted as such in (b)-(e) for ease of visualisation). (b, c) Boxes show median and IQR, whiskers show 1.5 × IQR, points show outliers; horizontal lines show results of post hoc pairwise comparisons between each group (Tukey’s HSD; ***p < 0.0001, ▪0.05 < p < 0.10). (b) Herbivores and omnivores had more generalised diets (greater sp50) than carnivores. (c) Variation in sp50 among sampling methods was driven by herbivores and did not differ significantly within carnivores (Table S1). Number of diets in each category is shown below boxes. (d) Sample size and (e) intra-annual sampling extent were positively correlated with sp50 across all diets (black) and for herbivores (blue) and carnivores (pink) separately (Tables S2 and S3). Trendlines in (d, e) show statistically significant linear regressions (all p ≤ 0.04) with shaded 95% CI.
(Ursus maritimus; Gormezano & Rockwell, 2013) populations were classified as herbivore and omnivore respectively. In further exploring correlates of sp50, we analysed consumer types both separately and together to assess whether trends in dietary generalisation were consistent across trophic levels.

The high relative importance of consumer type reflected the fact that herbivores and omnivores had more generalised diets (higher sp50) than carnivores (ANOVA, $F_{2,1164} = 35.35, p << 0.0001, R^2 = 0.06$; Figure 3b); this discrepancy might arise from the low digestibility and secondary chemistry of plants and the wider resource pool available to omnivores, but it might to some extent also be a product of the association between consumer type and sampling method. Studies based on observation yielded more generalised diets than those from faecal or stomach-contents analysis (ANOVA, $F_{2,1161} = 23.15, p << 0.0001, R^2 = 0.04$); this overall pattern was driven by herbivores and was not apparent in carnivores (Figure 3c, Table S1). As predicted, diets assessed with larger sample sizes were more generalised overall (regression, $n = 1019: \beta = 0.04, p = 0.02$) and for herbivores and carnivores separately (Figure 3d, Table S2). Likewise, intra-annual sampling extent was positively associated with sp50 overall (regression, $n = 1128: \beta = 0.0003, p < 0.0001$) and for each consumer type (Figure 3e, Table S3), a trend that held when we analysed only studies with ≥100 samples (regression; $n = 303, R^2 = 0.10, \beta = 0.001, p < 0.0001$).

Predictors attributed lower importance in the model-selection analysis included the biotic variables of chief conceptual interest, among others. Inter-annual sampling extent was uncorrelated with sp50 at large (regression, $n = 1130: \beta = -0.003, p = 0.92$) and for herbivores, but positively associated with sp50 for carnivores (Figure 4a, Table S4). Body mass was positively correlated with sp50 overall (regression, $n = 1167: \beta = 0.02, p = 0.002$) despite no association within consumer types (Figure 4b, Table S5), likely due to herbivores’ higher average body mass and sp50 relative to carnivores. Latitude was negatively associated with sp50 (more generalised diets in the tropics; regression, $n = 1167: \beta = -0.01, p << 0.0001$), but this trend occurred only in herbivores (Figure 4c; Table S6). Diets containing fruit had higher sp50 (ANOVA, $F_{1,1165} = 38.06, p << 0.0001$), but only among herbivores (Figure 4d; Table S7). Taxonomic resolution was positively associated with sp50 (regression, $n = 1167: \beta = 0.04, p = 0.0002$), another trend driven by herbivores (Figure 4e; Table S8); further analysis indicated that sampling method mediated this effect, as herbivorous diets described from faeces exhibited the strongest association between resolution and sp50 (Table S9). Owing to the lack of aquatic herbivores and omnivores in our data, we could only analyse effects of environment type for carnivores, and it had no effect on sp50 (ANOVA, $F_{2,466} = 0.27, p = 0.77$; Figure 4f).

All associations were qualitatively unaffected when controlling for the prevalence of Artiodactyla in our dataset (Tables S1–S8). Correlations involving biotic predictors were unaffected by including PCI effort as a covariate, although the biological effect sizes were often similar to or smaller than those of sampling effort (Tables S5–S7).

**DISCUSSION**

Discovery of the consistent shape of SADs in the early 1990s led to a proliferation of research and recognition of the hollow-curve SAD as ‘one of ecology’s true universal laws’ (McGill et al., 2007). We characterised diets as DADs and found that they, too, overwhelmingly follow a hollow-curve shape with few common and many rare foods. Quantifying generalisation as sp50 suggests that vertebrate diets are surprisingly specialised: on average, fewer than three foods accounted for >50% of the diet. The inferred extent of generalisation obviously increases when we consider sp75 or sp90 instead, but those metrics were tightly correlated with sp50, and
even the most inclusive scenarios indicate specialisation: for >75% of populations, <12 foods accounted for 90% of the diet (Figure S2). Thus, vertebrate populations can have species-rich diets \((sp90 > 100\) for forest elephant in Ghana and gray woolly monkey in Brazil) but disproportionately eat just a few taxa, and most dietary richness occurs in the long DAD tails (e.g., Peter’s duiker in Gabon had \(sp50 = 1\) but \(sp90 = 33\)). Those tails may be important both nutritionally (e.g., for obtaining elemental nutrients) and ecologically (e.g., for differentiating niches), but the ubiquity of steep hollow-curve DADs is also important for how ecologists conceptualise and model foraging behaviour and food webs.

**Methodological and biological associations with the generalist–specialist continuum**

The strong influence of methodological factors on \(sp50\) is a caveat to our analysis of generalisation, and indeed to all synthetic analyses of data on species interactions (Pringle & Hutchinson, 2020). Biases arising from incomplete sampling have long been problematic in research on biodiversity and species interactions (Chao & Jost, 2012; Jordano, 2016; Paine, 1988; Preston, 1948). In our dataset, \(sp50\) was positively associated with sample size and intra-annual extent across consumer types, and the magnitude of these associations was often larger than the biological effects, reflecting the temporal and individual variability of diets and underscoring the importance of robust sampling and long-term studies. Nonetheless, general conclusions about \(sp50\) in vertebrates’ diets were qualitatively unchanged when we accounted for sampling effort. Because \(sp50\) characterises dominant food items, it is probably most affected when sample size is so low as not to capture intra- or inter-individual dietary specialisation. Seasonal dietary switching appeared to alter \(sp50\) even at large sample sizes, further emphasising the importance of temporal sampling extent. We found a modest effect of
taxonomic resolution on \( sp50 \) (Knowlton & Jackson, 1994), which manifested most strongly for faecal-sampled herbivore diets. We expect that the use of faecal DNA for diet analysis—which enables detection of morphologically cryptic species and other difficult-to-identify food items, together with large sample sizes collected across multiple seasons and years—will ultimately provide the most accurate characterisation of dietary diversity (Pringle & Hutchinson, 2020). There were too few such studies in our compilation (\( n = 16 \)) to test this prediction. As an illustrative anecdote, however, we note that \( sp50 \) for African buffalo (\( Syncerus caffer \)) in Kenya based on DNA metabarcoding of 92 samples collected in multiple seasons across 4 years (Kartzinel et al., 2019) was 7.8, compared with \( sp50 \approx 3 \) in our dataset.

Splitting analyses by consumer type revealed variability in the correlations between dietary generalisation and biological/biogeographic predictors. The latitude-niche-breadth hypothesis predicts narrower niches in the tropics (MacArthur, 1972; Vázquez & Stevens, 2004) but has received mixed support (Cirtwill et al., 2015; Granot & Belmaker, 2020; Moles & Ollerton, 2016; Schemske et al., 2009). We found that generalisation increased towards the tropics for herbivores—in contrast to previous work on insect herbivores (Forister et al., 2015) but consistent with higher plant diversity at lower latitudes (Francis & Currie, 2003)—and showed no trend in carnivores. Likewise, we found a positive effect of fruit consumption on \( sp50 \) among herbivores only, which is mildly surprising given that fruit consumption seems more likely to add new food taxa for carnivores and omnivores (whereas herbivores may eat other parts of the same plant).

The dataset-wide association between body mass and \( sp50 \) was an artefact of size differences between more generalised herbivores and less generalised carnivores, neither of which exhibited a trend. For carnivores, body mass strongly influences foraging (Williams & Martinez, 2000); its lack of association...
with \(sp50\) is therefore surprising and may indicate that predator-prey body-mass ratio, rather than predator size alone, regulates diet (Clements et al., 2014; Petchey et al., 2008; Shao et al., 2021). Our null result for herbivores—where mouth size, home-range size, and biomass ingestion increase with body mass (Clauss et al., 2013; Gordon & Illius, 1988; McNab, 1963)—suggests that these factors have little influence on the diversity of dominant foods (but may add rare taxa to the diet). A positive relationship between body size and vertebrate diet breadth is intuitive and often assumed; our results offer a counterpoint and highlight the need for further study.

Parsing the independent effects of biological and methodological factors in such highly inclusive datasets is difficult. Parsing two-way and potential higher-order interactions among factors is even more difficult; we did not attempt to analyse interactions, because even with \(n >1000\) diets, there are empty sets at the intersection of multiple variables. The lack of marine herbivores in our data is one example. Another arises from interdependencies among consumer type, sampling method, latitude, and fruit consumption. These are exemplified by primates, which accounted for 4% of our data (45 of 1167 diets) but may have exerted considerable influence by having the most generalised diets among the well-represented taxonomic orders (mean \(sp50 = 5.4\); Figure S6). Ninety-eight percent of primates were frugivorous and 98% were studied observationally (vs. 13% and 16% for non-primates), and primate diversity is almost exclusively tropical (Graham et al., 2016; the IQR of primates in our dataset was 7.5°N to 4.6°S). The intersection of biology, biogeography, and study methodology in primates may thus influence the effects (or lack thereof) of body mass, fruit consumption, sampling method, and latitude on \(sp50\) in herbivores at large. Yet, attempting to unravel these contingencies by focusing on certain taxa would compromise taxonomic inclusivity and thwart our effort to identify broad generalities in diet structure.

Thus, as in other macroecological studies of trophic interactions (e.g. Hatton et al., 2015), deep mechanistic insights into the drivers of \(sp50\) are elusive and will require further analyses; we suggest zeroing in on particular consumer types or trophic guilds. Incorporating local ecological context—predation risk, competitor density, nutritional condition, individual specialisation, and especially resource availability—may account for much of the residual variance in our models. Situating DADs within their local ecological context could also help to address fundamental yet understudied questions in trophic ecology: What is the relationship between dietary generalisation and resource diversity (Ferretti et al., 2020; Matthews et al., 2019; Milana et al., 2016)? How does probability of interspecific encounter—the product of consumer and resource relative abundances (Jordano, 2016), such that abundant consumers are more likely to eat rare foods—shape diet and food-web structure?

**Universality of hollow-curve DADs**

We have shown that the hollow-curve DAD is a strong generality, but what does that mean ecologically? One obstacle to inference is the variety of factors that interactively constrain vertebrate foraging, including the relative availability and quality of foods, capture/handling times, gut capacity, prey/plant defences, consumer size and physiology, and predation risk. This diversity of influences is a longstanding barrier to general theories of animal foraging (e.g., Hughes, 1993) and makes the observed consistency in population-level DAD shape—the product of innumerable individual-level foraging decisions across hundreds of species—all the more intriguing. One possibility, which we take seriously given the ubiquity of hollow-curve abundance distributions in ecology and beyond, is that DAD shape is more of a statistical phenomenon than a biological one. In that case, foraging behaviour, food traits, and environmental factors may merely generate minor variation within fundamental mathematical constraints, as has been proposed for SADs (Diaz et al., 2021). Future work could examine the extent to which empirical DADs deviate from their statistical baselines (Diaz et al., 2021).

Another possibility is that DADs mirror SADs. Strict mirroring would imply that consumers indiscriminately filter food from their environments and therefore have DADs identical to resource SADs in both shape and rank abundance. Yet our case studies suggest otherwise (Figure 5), and the literature shows that (a) most animals (even filter feeders: DeMott, 1993; Sims & Quayle, 1998) select or avoid foods relative to availability (Ford & Ellis, 2006; Freeland & Janzen, 1974; Hayward & Kerley, 2008; Weinstein & Graham, 2017), and (b) that different species within guilds differ in diet composition (Kartzinel et al., 2015; Ross, 1986; Schoener, 1974), implying that their DADs cannot all mirror the shared SAD. Indeed, various factors modulate the relationship between resource availability and consumer choice, cause consumer choice to deviate from energy maximisation (e.g., partial preferences; McNamara & Houston, 1987), and make functional responses contingent on resource identity (Stouffer & Novak, 2021). Resource abundance influences DADs (almost by definition, as rare species are unlikely to be abundant in diets) and may tend to channel DADs towards hollow-curve shapes, but resource abundance alone is not a strictly sufficient explanation for DAD structure (Figure 5).

One hypothesis for the consistency in DAD shape that reconciles resource SADs with consumer selectivity is that foragers choose foods based on perceived ‘attractiveness’: some integrative measure of utility that is heavily swayed by availability but also incorporates attainability, nutritional value, and consumer-specific traits that influence the weighting of these attributes (e.g., morphology, predation vulnerability, condition).
Optimal foraging theory rests on a similar premise, in that food choice depends on energy yield relative to search costs (Schoener, 1987), but we envision a broader formulation of attractiveness and suspect that animals are more likely to satisfice than optimise (Carmel & Ben-Haim, 2005; Davidson & El Hady, 2019). Because few foods will satisfy all desiderata for any given consumer, few resources should be highly attractive, and the distribution of resource attractiveness should be right-skewed to varying degrees (e.g., more so in carnivores than herbivores given the lower sp50 of the former). We acknowledge that this conjecture is not fully satisfying: proposing a right-skewed distribution of a vaguely defined quantity to explain a right-skewed distribution of dietary abundance risks circularity. That said—and leaving aside exactly which parameters could suffice for an operational definition of resource attractiveness—animals obviously do perceive some foods as more attractive than others. Human foraging is a convenient reference point (e.g., grocery shopping: Furst et al., 1996), and economic consumer-choice theory has long grappled with how multiple variables influence product selection (Reisch & Zhao, 2017; Thaler, 1980). That literature validates the concept of resource attractiveness but also highlights the challenges of operationalising it.

Scaling from individuals’ diets to those of populations and species suggests another set of predictions about DADs. High levels of individual specialisation increase population niche breadth (Bolnick et al., 2010; Van Valen, 1965) and should thus flatten population-level DADs and increase sp50 when individuals differ in their use of ‘core’ foods. Conversely, inter-individual variation in consumption of marginal foods should lengthen the tail of the population-level DAD but not appreciably affect sp50. Documenting how DAD shape changes across levels of organisation, from individuals to populations to species, may provide an integrative approach to dietary specialisation, enabling researchers to distinguish specialisation attributable to behaviour (as reflected in individual- or population-level DADs) from that arising from morpho-physiological traits (as reflected in species-level DADs). Ultimately, just as models of community assembly are needed to understand SADs, a holistic understanding of how animals select diets is needed to explain the ubiquity of hollow-curve DADs.

Prevalence of trophic specialisation in vertebrate consumers

The dominance of diets by just a few food items (sp50 < 3) suggests that vertebrate populations are more specialised than commonly assumed. Trophic specialisation in insects is widespread and often extreme (Forister et al., 2015; Smith et al., 2006, 2008), and our findings are consistent with the fact that food-webs generally comprise a few strong and many weak links (McCann et al., 1998; Paine, 1992). However, they run counter to the view that vertebrate diets are highly generalised. McNaughton (1978) described African buffalo and wildebeest as ‘un-selective herbivores’. Westoby (1978) sought to explain why sheep diet breadth was higher than predicted by optimal-foraging models. Some of the more comprehensive recent accounts of vertebrate dietary richness likewise convey the impression of extreme generalisation: 25 species of large herbivores in Kenya each ate between 78 and 142 plant species (Kartzinel & Pringle, 2020); leopards (Panthera pardus) in South Africa ate at least 50 species spanning four vertebrate classes (Balme et al., 2020); and bluefin tuna (Thunnus maccoyii) off Tasmania ate at least 92 prey taxa (Young et al., 1997).

Yet, in our dataset, this same set of species had mean sp50 < 3 (range = 0.6–16 across populations), indicating that the vast majority of dietary richness arises from the long tail of DADs. Thus, although many species are generalists in the sense of being able to eat a broad range of foods, populations tend to have narrow ‘core’ diets. This has implications for understanding foraging behaviour, niche relationships, and food-web structure. For instance, it suggests (a) that individual specialisation/variation (Bolnick et al., 2003) is typically not so strong as to preclude population-level convergence on a short list of core foods, and (b) that any decrease in the availability of those core foods may elicit strong behavioural (e.g., dietary switching) and/or demographic responses. At the scale of food webs, the prevalence of trophic specialisation suggests that binary interaction networks (Carscallen et al., 2012; Miranda et al., 2019) may severely misrepresent diet composition and be a poor basis for predictions about ecological dynamics, at least insofar as dietary proportions approximate interaction strengths (Vázquez et al., 2005; Wootton, 1997). The high specialisation of carnivores further suggests that the strongest food-web links and consumer-resource coupling should occur at high trophic levels (Scotti et al., 2009).

Future directions

The composition of our dataset highlights several avenues for further work. Reptiles, amphibians, and fishes were underrepresented; testing a wider range of taxa would help to establish the generality of these patterns beyond terrestrial endotherms. It would also be useful to know whether the patterns documented here hold for invertebrate consumers, which constitute the majority of nodes in most food webs. On the one hand, many invertebrates, ranging from cyamid crustaceans (Kaliszewska et al., 2005) to insect folivores (Forister et al., 2015) and parasitoids (Smith et al., 2006), are highly specialised on just one or a few closely related species. On the other hand, many flower-visitors
(Waser et al., 1996), grasshoppers (Joern, 1979), and suspension-feeders (Wells et al., 2021) can have highly generalised diets.

The foremost question that emerges from our study is the mechanistic basis of the hollow-curve DAD shape—the near-universality of which hints at the existence of fundamental rules governing diet assembly. DADs are not independent of resource SADs, nor are they identical. SADs presumably define a feasible set of DADs for a given consumer population, constraining the dietary abundance of resources and perhaps imposing a hollow-curve ‘shadow’ on DAD shape; the question then becomes to what extent observed DADs differ from this baseline, and why. Theoretical and statistical modelling will be useful in addressing these questions. Empirical work could expand on our case studies (Figure 5) to more definitively establish the degree to which resource relative abundance predicts dietary relative abundance. Contrasting the diets of sympatric guild members may be particularly useful in this regard, by controlling for resource availability and illuminating the role of competition and resource partitioning in causing DADs to deviate from SADs. Yet, a major barrier to empirical progress (especially in light of the methodological influences on DAD steepness) is the scarcity of high-resolution diet data matched with equally high-resolution data on resource relative abundance (Pringle & Hutchinson, 2020).

ACKNOWLEDGEMENTS
We acknowledge the hundreds of researchers whose work enabled this synthetic dataset. We thank Bernat Bramon Mora, Corina Tarnita, Cassie Stoddard, and Silvia de Monte for input throughout the development of this study. Support for this work was provided by the Department of Ecology and Evolutionary Biology and High Meadows Environmental Institute of Princeton University, Scholarships New Zealand’s William Georgetti Fellowship, the US National Science Foundation (DEB-1457697, IOS-1656527), the Carr Foundation, and the Cameron Schrier Foundation.

AUTHOR CONTRIBUTIONS
The authors jointly conceived the study. Matthew C. Hutchinson compiled data from the literature, conducted analyses and drafted the manuscript with input from Andrew P. Dobson and Robert M. Pringle. All authors agreed to manuscript submission.

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13948.

DATA AVAILABILITY STATEMENT
The entire database, associated metadata and data-processing scripts are deposited on Dryad [https://doi.org/10.5061/dryad.08kprr53v].

ORCID
Matthew C. Hutchinson https://orcid.org/0000-0002-2423-4026
Andrew P. Dobson https://orcid.org/0000-0002-9678-1694
Robert M. Pringle https://orcid.org/0000-0001-7487-5933

REFERENCES


Wells, C.D., Paulay, G., Nguyen, B.N. & Leray, M. (2021) DNA metabarcoding provides insights into the diverse diet of a dominant suspension feeder, the giant plumose anemone *Metridium farcimen*. Environmental DNA.


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.