Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

National mapping of hyperdiverse insect communities beyond named species

Jan Borgelt

jan.borgelt@ebc.uu.se

Uppsala University https://orcid.org/0000-0002-7372-0071

Francesca Verones

NTNU https://orcid.org/0000-0002-2908-328X

Tobias Andermann

Uppsala University https://orcid.org/0000-0002-0932-1623

Article

Keywords:

Posted Date: October 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11151855/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

National mapping of hyperdiverse insect communities beyond named species

Abstract

How biodiversity is distributed across landscapes is a fundamental question in ecology, yet spatial information remains sparse for many hyperdiverse groups. Here we combine standardized DNA metabarcoding from 196 Malaise traps with order-specific joint species distribution models to predict the distributions of 11,694 insect taxa across Sweden. Predicted richness showed a pronounced south-to-north decline, but this aggregate pattern concealed substantial differences among insect orders and individual taxa. More than half of the modelled assemblage comprised *dark taxa,* molecular clusters lacking species-level assignments. These taxa contributed 51.9% of mapped richness and closely reproduced the broad national richness geography of named species, despite being detected at somewhat fewer traps. At the individual-taxon level, however, dark taxa had modestly smaller predicted spatial distributions overall, with considerable variation among orders. Order-specific richness maps further showed that the overall insect richness pattern was not representative of all groups, with different orders showing distinct spatial concentrations across Sweden. By retaining taxon-level predictions across both named species and dark taxa, our framework provides a scalable route to mapping hyperdiverse assemblages at national scale, even when much of the assemblage remains taxonomically unresolved.

Introduction

Biodiversity is distributed unevenly in space, producing hotspots and coldspots that often differ among taxonomic groups and reflect interacting environmental, biological and anthropogenic processes[1–7]. Yet spatially explicit knowledge of these patterns remains profoundly uneven across taxa, regions and scales[8–10]. At the same time, governments and international organizations increasingly require accurate and reproducible spatial biodiversity information[11–15], both to establish robust baselines[16–18] and to evaluate progress toward conservation targets[19–21]. For instance, maps showing where individual species occur, how communities are distributed, or where species richness is highest can guide biodiversity surveys, conservation and restoration priorities, and efforts to maintain habitat connectivity[22–25]. They also support environmental assessments, national reporting, and biodiversity-inclusive land-use planning[26,27]. For most species groups, however, such information remains coarse, geographically limited or restricted to aggregate diversity measures[28,29].

Biodiversity can be sampled at specific places and times, but resource, personnel and time constraints make it impossible to survey entire landscapes continuously[27]. Species distribution models (SDMs) help overcome this limitation by linking species observations to environmental conditions and using these relationships to predict where species are likely to occur across the wider landscape[30,31]. The quality of these predictions, however, depends on the observations used to build the models. Opportunistic databases provide vast numbers of occurrence records but are affected by uneven sampling effort, inconsistent methods, accessibility bias, and taxonomic and geographic gaps[32–34]. Standardized biodiversity inventories can reduce many of these problems by surveying the same target groups using consistent methods across multiple

sites. Yet conventional inventories are labor-intensive, depend on specialist taxonomic expertise, and are difficult to scale across large areas. Biodiversity mapping therefore faces a persistent trade-off between taxonomic breadth, sampling consistency and spatial coverage[35].

These constraints are especially severe for hyperdiverse taxa[36,37]. Such groups contain very large numbers of species, many of them rare, cryptic or poorly known, making comprehensive surveys difficult with conventional approaches[36,38]. Insects are a prime example. They comprise a major fraction of terrestrial animal diversity and support key ecological functions and services, including pollination, decomposition, nutrient cycling, pest regulation and food-web dynamics[39–42], yet they remain underrepresented in national and international biodiversity assessments[43,44]. Many insect species also respond to local variation in vegetation structure, moisture, microclimate, land use and habitat configuration[45–47]. Consequently, insect biodiversity remains poorly represented spatially, limiting efforts to identify areas warranting closer attention, as well as targeted field surveys[10,39].

Recent advances in biodiversity monitoring, earth observation and computational capabilities are fundamentally changing what is possible[48–52]. DNA metabarcoding of bulk samples can identify thousands of taxa at once from standardized collections, greatly expanding the taxonomic breadth and throughput of biodiversity inventories[53–55]. Metabarcoding also recovers large numbers of distinct DNA sequence clusters that cannot be assigned to named species. Following established usage, we refer to such molecular clusters lacking species-level assignments as *dark taxa*[56–58], while recognizing that individual clusters need not correspond one-to-one with biological species. Standardized and clearly documented workflows can further enable geographically distributed sampling by volunteers and citizen scientists[59]. Meanwhile,

satellite observations, airborne LiDAR and digital environmental layers increasingly provide wall-to-wall information on climate, land use, vegetation structure, productivity and habitat heterogeneity at fine spatial resolution[60–63]. Growing computational capacities now allow these high-dimensional biological and environmental data streams to be analyzed together[64,65]. Together, these developments create an opportunity to move beyond maps of a few well-recorded taxa and instead capture entire species communities continuously across landscapes[27,66].

Joint species distribution models (jSDMs) provide the analytical framework needed to realize this opportunity[67,68]. Conventional species distribution models typically related each species' occurrence to environmental conditions separately, an approach that becomes increasingly cumbersome when applied consistently across assemblages containing thousands of taxa[66,67]. jSDMs instead model many taxa simultaneously and can account for residual associations among their occurrences after measured environmental and spatial variation has been considered[67–69]. This multivariate structure is well matched to DNA-based inventories, which provide observations of whole assemblages at each sampled site[69]. Importantly, jSDMs retain predictions for individual taxa, allowing community-level properties to be derived from the same underlying distributions. Species richness can therefore be reconstructed from taxon-specific predictions rather than modelled directly as an aggregate response, preserving the link between broad biodiversity patterns and the taxon-level composition from which they arise. To date, however, jSDMs have generally been applied to communities containing tens to hundreds of taxa, reflecting both computational constraints and limits in the taxonomic breadth of

available biodiversity data. Their potential for mapping hyperdiverse assemblages containing thousands of taxa across large areas therefore remains largely unexplored.

Here, we combine standardized metabarcoding inventories from Malaise-trap samples collected by the Swedish Insect Biome Atlas[53] with high-resolution environmental data and joint species distribution modelling to map insect diversity across Sweden. Using environmental predictors at 10-m spatial resolution, we generate taxon-level distribution predictions for more than 11,000 taxa across 12 taxonomic orders and derive richness and community-composition patterns from local landscapes to the national scale. We quantify how dark taxa contribute to these maps and test how aggregate richness patterns vary among insect orders and individual taxa. The resulting maps reveal a strong national diversity gradient, alongside clear differences in richness patterns among taxonomic orders and contrasting distributions among individual insect taxa. Dark taxa account for more than half of the mapped assemblage and contribute substantially to overall richness while generally displaying diversity patterns similar to those of named species. Our study demonstrates a scalable route from standardized metabarcoding-derived biodiversity inventories to national mapping of hyperdiverse assemblages even where taxonomic knowledge is incomplete, with potential application to other species-rich, poorly known groups.

Results

National mapping reveals pronounced insect-diversity gradients

For each of 12 insect orders with sufficient data, we evaluated 500 hyperparameter configurations using 5-fold cross-validation (CV). We first retained configurations whose mean test area under the receiver operating characteristic curve (AUC) fell within one standard error

of the best-performing configuration and then selected five models from this candidate set. Predictions from these five selected models were combined to map the distributions of 11,694 insect taxa across Sweden, including both named species and dark taxa. For each taxon, occurrence probabilities were converted to presence-absence predictions using taxon-specific F1 thresholds and summed to derive predicted insect richness. Predicted richness showed a pronounced national spatial gradient, ranging from 374 to 4,844 taxa per approximately 2-km mapping cell (Figure 1a). Richness was generally highest in southern Sweden and declined towards northern latitudes and higher elevations, with the lowest values occurring in the western mountains. The upper 5% of mapped richness values contained at least 3,263 predicted taxa per cell, whereas the lower 5% contained no more than 779.

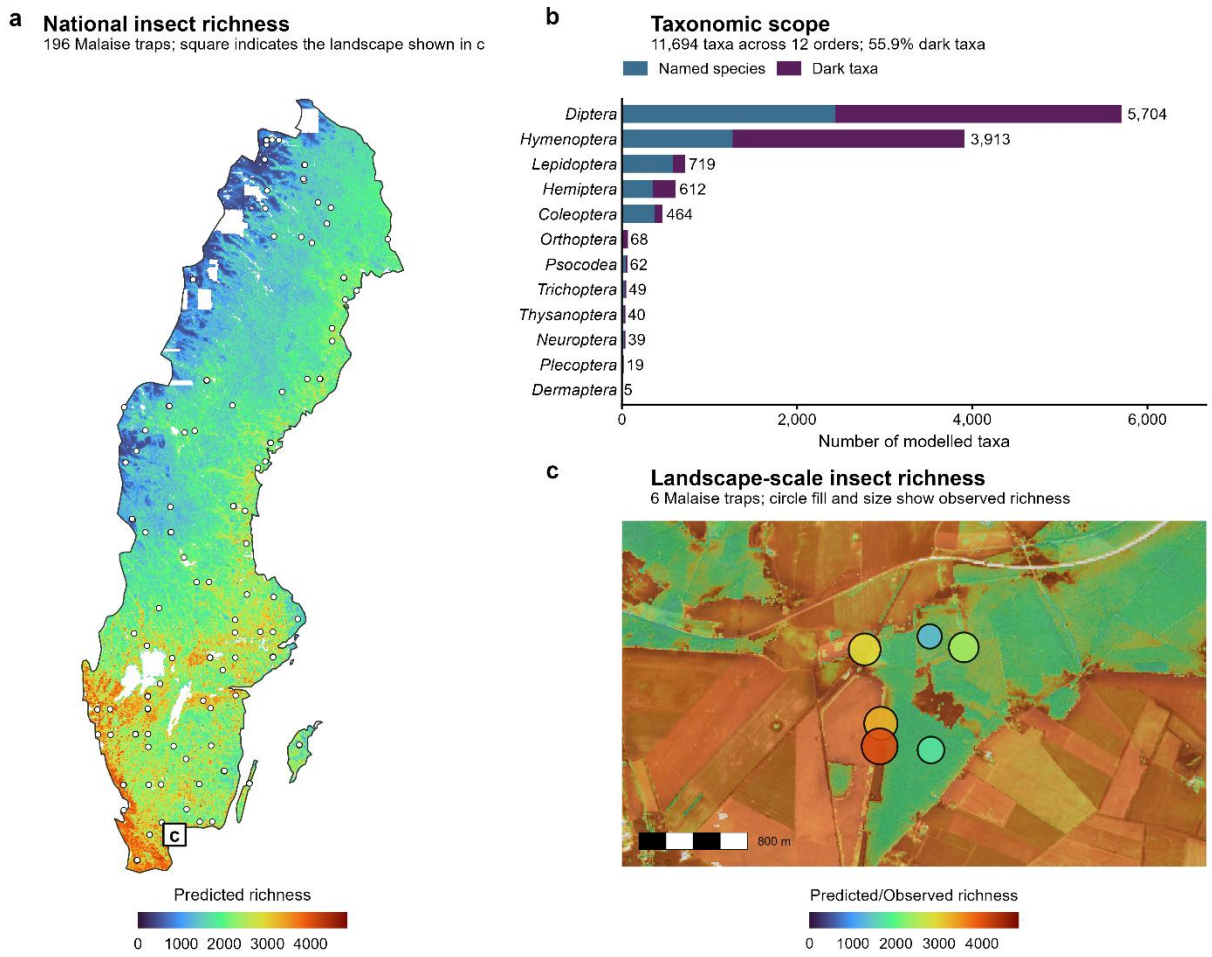


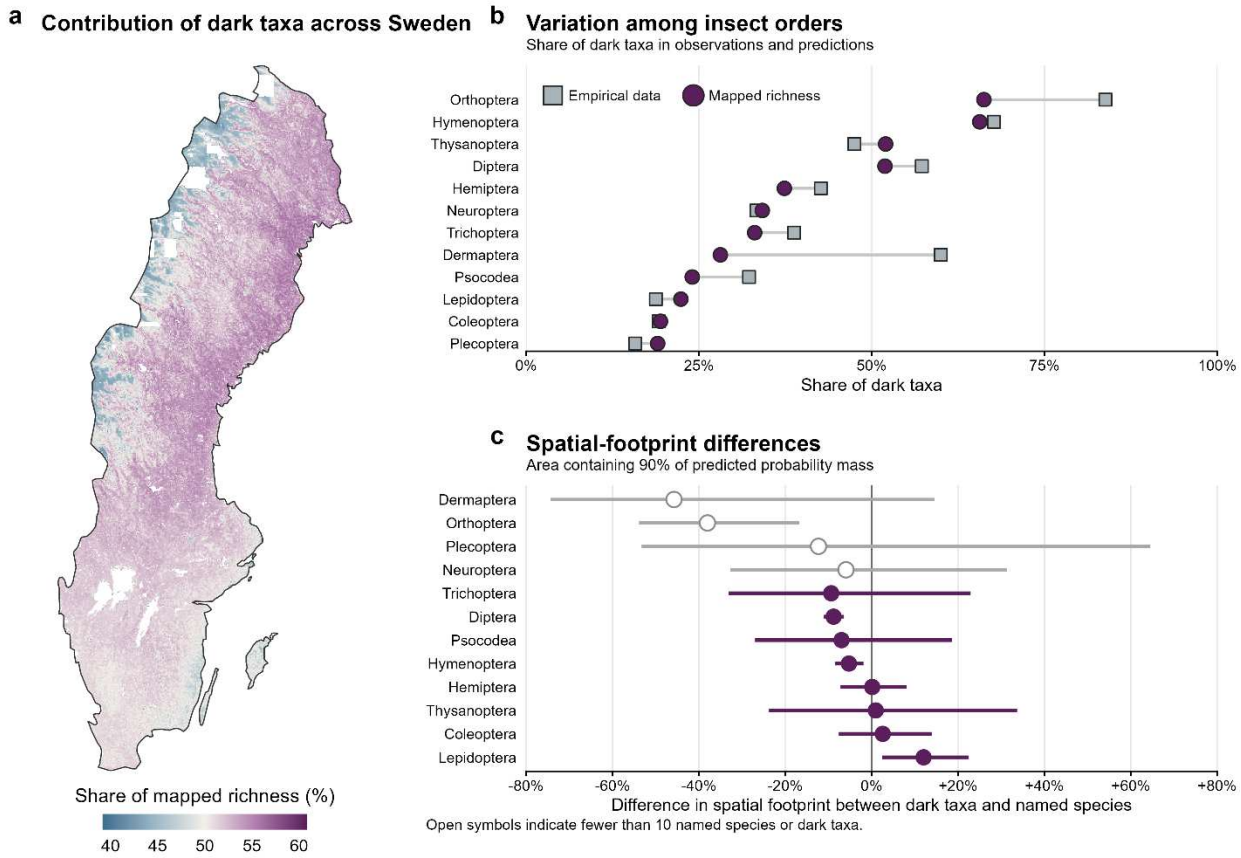
Figure 1: A national map of insect diversity across Sweden. a) Predicted total insect richness across Sweden based on joint species distribution models fitted separately for 12 insect orders. National predictions were generated for 2 million randomly sampled 10-m cells and averaged within approximately 2-km grid cells for visualization at the national scale. Open circles show sampling sites, and the box marks the landscape shown in c. b) Taxonomic scope of the mapped insect diversity across orders, showing the number of mapped insect taxa and the contributions of named species versus unnamed DNA sequence clusters, i.e., dark taxa. c) Predicted insect richness within an example landscape at the original 10-m prediction grain. Circle size and fill show observed richness at Malaise traps.

The broad national diversity gradient was accompanied by pronounced variation at local scales. Within the multi-trap landscape shown in Figure 1c, predicted richness varied markedly over short distances, revealing fine-scale spatial structure that was not apparent from the national-

scale map. This local variation was also reflected more generally in the correspondence between observed and predicted richness across sampling sites.

Across the 196 sampled traps, predicted richness closely tracked observed variation in richness across all 12 insect orders (Extended Data Figure 1). Spearman correlations between observed richness and the mean prediction across the five selected models ranged from 0.711 to 0.927 and exceeded 0.80 for nine of the 12 orders. Because these models were fitted using the full dataset, this comparison is most informative about relative spatial ranking at the sampled sites rather than independent predictive performance.

At the taxon level, site-grouped CV indicated generally high discrimination across all 12 orders, including insects detected at relatively few traps (Extended Data Figure 2a,c). Across all 11,694 taxa, median cross-validated AUC, averaged across the five selected models, was 0.770 (interquartile range, 0.689-0.838). A total of 71.8% of insect taxa exceeded an AUC of 0.7 and 39.4% exceeded 0.8. Discrimination remained similar among taxa detected at no more than 25 traps and those detected at more than 100 traps, with median AUC values of 0.770 and 0.743, respectively.



142

143 *Figure 2: Dark taxa make a substantial contribution to mapped insect diversity. a) Spatial variation in the contribution of dark*
144 *taxa to mapped insect richness across Sweden, expressed as the percentage of mapped richness represented by dark taxa. b)*
145 *Variation among insect orders in the share of modelled taxa represented by dark taxa and their contribution to mapped richness.*
146 *Squares show the proportion of modelled taxa that were dark taxa and circles show the proportion of mapped richness*
147 *contributed by dark taxa; lines connect the two estimates within each order. c) Order-specific differences in spatial footprint*
148 *between dark taxa and named species, quantified as the area required to contain 90% of predicted probability mass. Points*
149 *show estimated percentage differences between dark taxa and named species and horizontal lines show 95% confidence*
150 *intervals; negative values indicate smaller spatial footprints for dark taxa. Open symbols indicate orders represented by fewer*
151 *than ten named species or dark taxa. Across all 12 orders, 11,694 taxa were mapped, comprising 5,160 named species and 6,534*
152 *dark taxa.*

Dark taxa substantially increase mapped richness but show similar national geography

Dark taxa comprised 6,534 (55.9%) of the 11,694 modelled taxa, compared with 5,160 named species (Figure 1b). Their representation differed markedly among insect orders, ranging from 15.8% of modelled taxa in Plecoptera and 19.2% in Coleoptera to 67.7% in Hymenoptera and 83.8% in Orthoptera (Figure 2b). Their contribution to mapped richness broadly followed these differences, although the two proportions were not identical among orders. Across Sweden, dark taxa contributed 51.9% of mapped insect richness, more than doubling mapped richness relative to named species alone. Most mapping cells were close to parity between named species and dark taxa, but values above 50% were concentrated across eastern parts of central and northern Sweden, whereas dark taxa generally contributed less than half of mapped richness in the western mountains (Figure 2a).

Dark taxa were detected at somewhat fewer traps than named species, with median detection frequencies of 24 and 30 traps, respectively (Extended Data Figure 2d). Despite this shift towards lower detection frequencies, discriminatory performance was almost identical overall, with median cross-validated AUC values of 0.772 for dark taxa and 0.768 for named species. Within orders, differences were generally modest and changed direction among groups, with neither named species nor dark taxa showing consistently higher discrimination (Extended Data Figure 2b). Thus, the lower detection frequency of dark taxa did not translate into systematically poorer model discrimination.

At the national scale, named species and dark taxa showed closely corresponding richness distributions, and reproduced the same dominant south-to-north richness gradient (Extended Data Figure 3a,b). When cells were ranked independently by richness within each component,

their percentile ranks were strongly correlated across the mapped cells (Spearman's ρ = 0.976; Extended Data Figure 3c). Thus, adding dark taxa changed absolute mapped richness far more than it changed the relative spatial ranking of places.

This agreement at the national level, however, masked differences in the distributions of individual taxa. Among the taxa with cross-validated AUC values of at least 0.7, dark taxa had modestly smaller predicted spatial footprints than named species. We quantified spatial footprint as the area containing 90% of a taxon's predicted probability mass. In mixed models accounting for differences among insect orders, this area was 6.2% smaller for dark taxa (95% confidence interval, 4.4-8.0%; Figure 2c). The overall difference remained negative under progressively stricter model performance filters (AUC ≥ 0.70, 0.75 and 0.80; Extended Data Figure 4), but varied among insect orders. Among well-represented orders, dark taxa had smaller spatial footprints in Diptera and Hymenoptera, whereas the relationship was reversed in Lepidoptera. Estimates for several smaller orders were much less precise because one of the two groups contained few taxa. Thus, strong concordance in overall richness patterns masked modest but taxonomically heterogeneous differences in the spatial extent and distribution of individual taxa.

Order-specific maps reveal shared and contrasting richness patterns

Order-specific richness maps revealed a broadly shared national diversity gradient while also exposing marked differences among individual insect groups (Figure 3a). Spatial correlation with overall insect richness was strongest for Diptera, Hymenoptera, Lepidoptera, Hemiptera and Coleoptera (Spearman's ρ = 0.911-0.953). In contrast, Trichoptera richness was effectively unrelated to the aggregate pattern (ρ = -0.007), whereas Plecoptera showed a strongly opposing

distribution (ρ = -0.764), with relatively high richness in northern and western Sweden. Accordingly, overlap among order-specific hotspots was limited. Using the upper 10% of richness values to define order-specific hotspots, only 10.1% of the mapped area was a hotspot for at least five orders, 4.1% for at least eight orders and 0.17% (744 km²) for at least ten of the 12 orders (Figure 3b). Diversity hotspots based on all insect taxa contained 80.7-83.0% of the order-specific hotspots for Diptera, Hemiptera and Hymenoptera, but only 26.3% of Trichoptera hotspots, 7.9% of Neuroptera hotspots and none of the Plecoptera hotspots (Figure 3c). The overall insect diversity map therefore captured the dominant multi-order pattern but did not encompass many high-richness areas for several other insect orders. These differences reflected contrasting geographical distributions among their constituent taxa, ranging from broad northern and southern patterns to geographically restricted distributions.

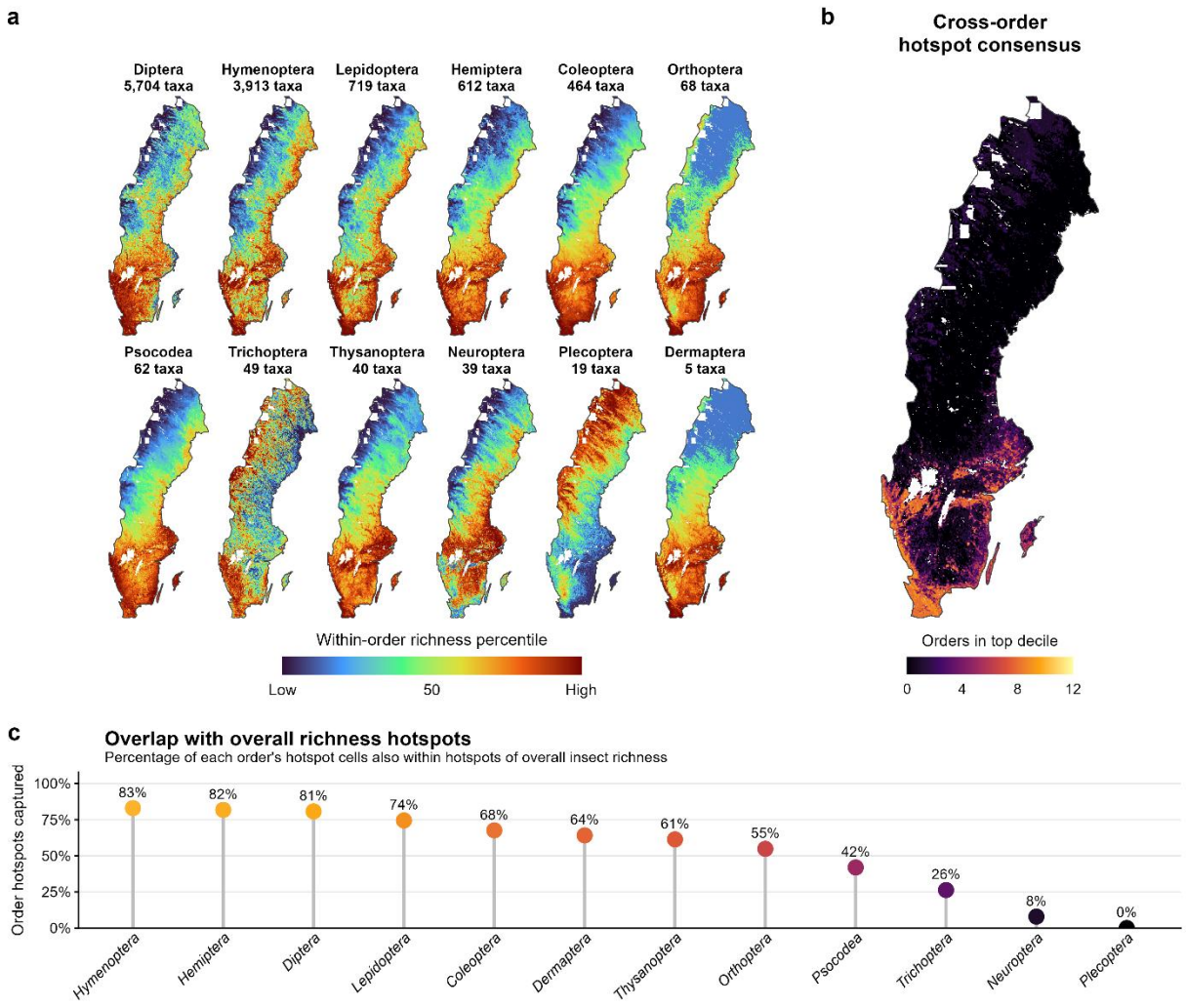
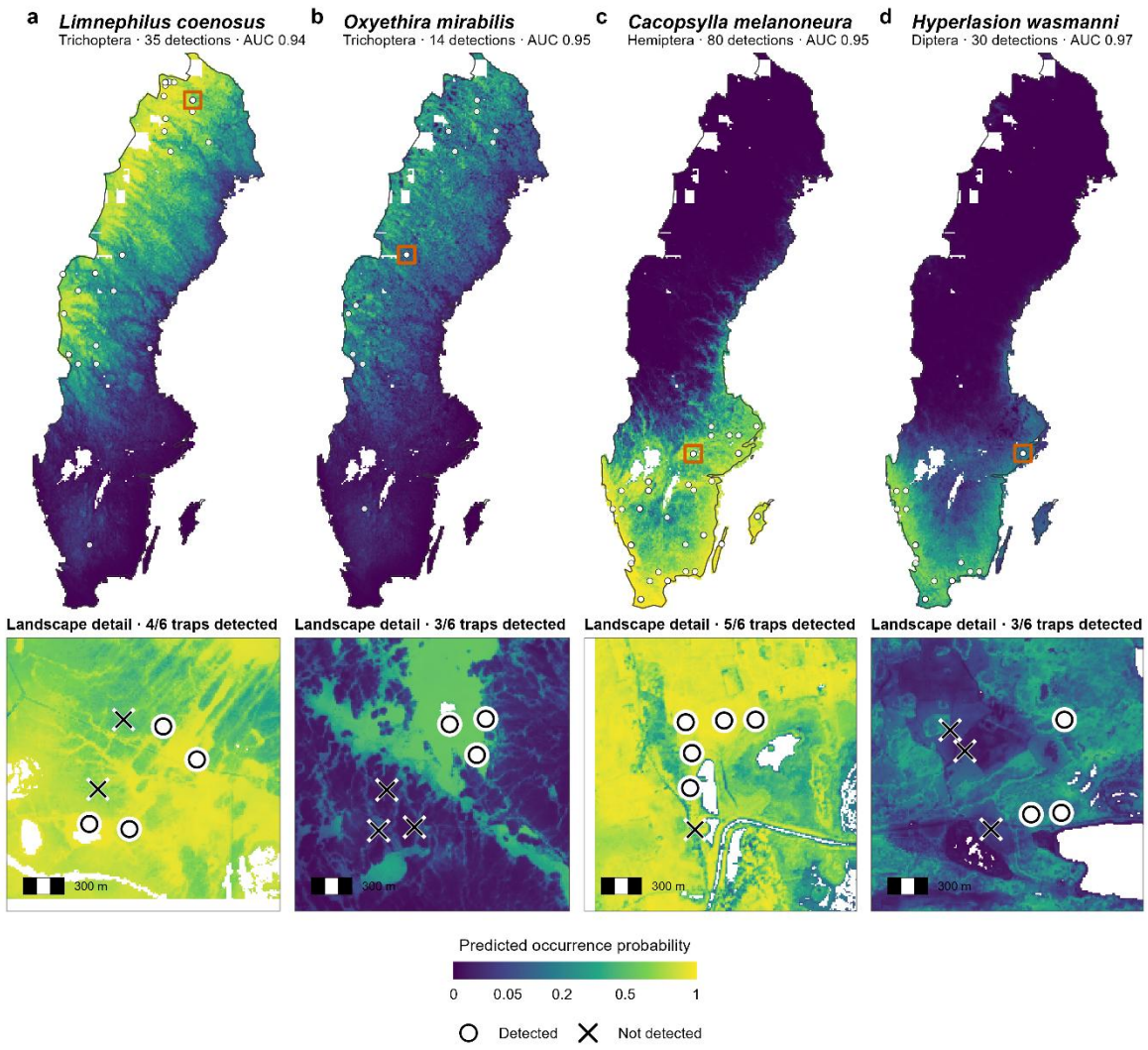


Figure 3: Order-specific richness patterns and hotspot overlap across Sweden. a) Predicted richness patterns for 12 insect orders, expressed as within-order richness percentiles to allow comparison among orders differing strongly in the number of modelled taxa. The number of modelled taxa is shown for each order. b) Cross-order hotspot consensus, showing the number of orders for which each approximately 2-km grid cell fell within the upper 10% of predicted richness values for that order. c) Overlap between order-specific richness hotspots and hotspots based on overall insect richness. Values show the percentage of hotspot cells for each order that were also classified as richness hotspots for all insect taxa combined. Hotspots were defined separately for each order, and for all insects, as cells at or above the corresponding 90th percentile.

Contrasting species distributions underpin national richness patterns

The national richness surface emerged from the combined predictions of thousands of individual taxa, whose distributions differed markedly in geographical position and spatial extent (Figure 4). Using probability-weighted area as a threshold-free measure of spatial extent, the two trichopterans illustrated contrasting northern distributions (Figure 4a,b). *Limnephilus coenosus* was predicted broadly across northern and western Sweden, with a probability-weighted centroid at 65.3° N and a probability-weighted area of 109,572 km², whereas *Oxyethira mirabilis* had a similarly northern centroid (65.1° N) but a substantially smaller probability-weighted area of 41,376 km². The two southern examples showed an analogous contrast (Figure 4c,d). *Cacopsylla melanoneura* was predicted widely across southern Sweden (centroid, 58.1° N; probability-weighted area, 91,293 km²), while *Hyperlasion wasmanni* was concentrated farther south and over a more restricted area (centroid, 57.5° N; probability-weighted area, 28,927 km²). All four taxa showed high discrimination, with mean cross-validated AUC values across the five selected models ranging from 0.943 to 0.974.



230

231 *Figure 4: Contrasting taxon-level distributions connect national and landscape-scale patterns of insect diversity. a-d) Predicted*
232 *occurrence probability at the national scale (upper row) and within selected multi-trap landscapes at 10-m resolution (lower*
233 *row) for Limnephilus coenosus (a), Oxyethira mirabilis (b), Cacopsylla melanoneura (c) and Hyperlasion wasmanni (d).*
234 *Predictions show the mean occurrence probability across the five selected joint species distribution models for each order. Taxa*
235 *were selected to represent contrasting northern, southern, broad and spatially restricted distribution patterns while retaining*
236 *informative local variation. White circles on the national maps indicate sampling locations at which the taxon was detected, and*
237 *orange squares locate the corresponding landscape panels. In the local panels, circles indicate detections and crosses indicate*
238 *non-detections at Malaise traps. AUC values are mean site-grouped cross-validated AUC values across the five selected models.*
239 *All panels use a common probability scale; the color scale is square-root transformed for visualization, while legend labels report*
240 *untransformed occurrence probabilities.*

The national differences were accompanied by fine-scale variation within the selected multi-trap landscapes. Mean predicted occurrence probabilities across the five selected models were higher at traps where each species was detected than at traps where it was not detected, with differences of 0.13-0.41 across the four species. The selected landscapes contained three to five detections among six neighboring traps. Together, these examples show how contrasting taxon distributions underpin national richness patterns while retaining fine-scale structure hidden by aggregate richness.

Discussion

By combining standardized DNA-based sampling with joint species distribution modelling, we translated observations from a geographically sparse national sampling network into national-scale distribution predictions for more than 11,000 insect taxa. The resulting maps revealed a dominant national richness gradient while retaining substantial variation among insect orders and individual taxa. More than half of the modelled assemblage consisted of dark taxa, which substantially increased mapped richness while closely matching the broad national richness patterns of named species. More generally, our study shows how standardized molecular monitoring and scalable community modelling can extend spatial biodiversity assessment to hyperdiverse groups for which both distributional and taxonomic knowledge remain incomplete.

These challenges correspond to two long-recognized shortfalls in biodiversity knowledge: the Wallacean shortfall in knowledge of species distributions and the Linnean shortfall in taxonomic knowledge[29]. Large-scale Malaise-trap metabarcoding surveys have shown that thousands of

named species and taxonomically unresolved molecular units can be sampled systematically across national networks[53,70], while DNA-based community data combined with joint species distribution models have enabled fine-grained predictions of individual arthropod distributions at landscape scales[71]. More recently, the Swedish Insect Biome Atlas has supported spatially continuous, seasonal predictions of aggregate arthropod richness across Sweden[72]. Our approach connects these developments by scaling taxon-level distribution modelling to a hyperdiverse assemblage across an entire country and deriving aggregate richness from the underlying predictions. Importantly, progress in spatial mapping does not require the Linnean shortfall to be resolved first, taxonomically unresolved molecular clusters can remain represented in the analysis rather than being discarded. Retaining individual predictions also preserves information about the distributions that generate aggregate richness and enables subsequent analyses of particular taxonomic or ecological subsets.

The dominant national richness pattern was consistent with established latitudinal and climatic gradients reported for insect assemblages at continental and global scales[46,73,74]. Higher richness in southern Sweden is consistent with the region's warmer climate, greater productivity, longer growing season and greater habitat heterogeneity[75]. Yet the national gradient contained substantial fine-scale structure. Within the illustrated multi-trap sites, predicted richness and occurrence probabilities varied over short distances across heterogeneous environments (Figure 1c, Figure 4). This coexistence of broad geographical gradients and fine-scale variation is consistent with regional climate constraining the available species pool while local land cover and vegetation structure influence how taxa are distributed within landscapes[46,76,77].

Aggregate richness nevertheless obscured important differences among its constituent groups. Taxon-rich orders such as Diptera and Hymenoptera contributed strongly to the national pattern, whereas several less diverse orders followed markedly different geographies. Plecoptera showed a distribution strongly opposed to overall richness, and Trichoptera was largely unrelated to the aggregate pattern. Such contrasts are ecologically plausible because insect orders differ in climatic affinities, habitat requirements and dependence on aquatic or terrestrial environments[74,78,79]. Patterns for the smallest orders should be interpreted cautiously because their richness surfaces are necessarily more sensitive to the distributions of relatively few constituent taxa. Similar richness values can arise from assemblages composed of widespread, geographically restricted, northern or southern taxa. Richness maps are therefore most informative when interpreted alongside the taxon-level predictions from which they are derived[79–81]. These underlying distributions can support spatial screening for particular groups, including taxa of conservation concern, and help target locations where more detailed field assessment is warranted rather than substituting for such assessments.

The contribution of dark taxa further illustrates the value of retaining molecular diversity that lacks formal species names. Dark taxa constituted 55.9% of the modelled assemblage and contributed 51.9% of mapped richness, yet their national richness rankings were highly concordant with those of named species. Taxonomic incompleteness therefore did not translate into spatial invisibility: adding dark taxa changed the magnitude of mapped richness much more strongly than it changed its broad national geography. Dark taxa were also detected at somewhat fewer traps but showed similar cross-validated discrimination to named species, indicating that their unresolved taxonomic status did not itself prevent useful spatial modelling.

This does not imply that individual molecular clusters correspond one-to-one with undescribed biological species. Barcode clustering can partition intraspecific genetic variation among multiple molecular units, and incomplete reference libraries can leave otherwise known taxa unnamed[53,82–85]. Nevertheless, under standardized sequence processing, these clusters represent reproducible components of the sampled assemblage[53,83], and discarding them would remove a substantial component of the sampled assemblage and its spatial signal.

Similarity in aggregate richness geography did not, however, make named species and dark taxa spatially interchangeable. Dark taxa had modestly smaller spatial footprints overall, and this difference persisted as analyses were restricted to taxa with increasingly strong cross-validated discrimination. The direction and magnitude of the difference also varied among orders: dark taxa had smaller footprints in the species-rich Diptera and Hymenoptera, whereas the relationship was reversed in Lepidoptera. Such differences could reflect genuine geographic restriction, but several other processes could produce similar patterns, including variation in abundance, detectability, phenology or molecular recovery, and the partitioning of geographically structured intraspecific variation among multiple molecular clusters[84,86–89]. The present analysis cannot distinguish among these mechanisms, and the particularly uncertain estimates for orders represented by few named species or dark taxa should not be over-interpreted. The result nevertheless demonstrates that strong correspondence in aggregate richness can coexist with differences in the spatial distributions of the taxa that generate it.

The maps should be interpreted in light of the sampling design and observation process. National predictions were derived from 196 Malaise traps and therefore depend on how well this network represented Sweden's environmental and geographical variation[90,91]. Seasonal

activity, local abundance and imperfect detection influence whether a taxon is recorded, yet detection probability was not modelled explicitly[86,92] . The predictions should therefore be interpreted as distributions of taxa detectable under the Malaise-trap sampling protocol rather than as complete inventories of local insect assemblages. Likewise, the 10-m prediction grain reflects the spatial grain of the environmental predictors and model outputs, not independent validation of distributions at 10-m resolution. Site-grouped cross-validation indicated generally good discrimination, but AUC evaluates discrimination rather than calibration[93]. Uncertainty in taxon-level predictions, and in the thresholds used to convert them to presence-absence predictions, consequently propagates into mapped richness. Environmental extrapolation was generally limited (Extended Data Figure 5) but greater in parts of northern and northwestern Sweden, where predictions warrant additional caution[91,94]. Taxonomic uncertainty adds another layer because sequence clustering and incomplete reference libraries affect both the number and identity of modelled taxa[53,82]. Mapped values should consequently be viewed as model-based estimates of spatial biodiversity patterns rather than exact local species counts or confirmed occurrences at individual cells.

The framework is not specific to insects. Soil metabarcoding surveys already characterize exceptionally diverse fungal, microbial and soil-faunal communities at national to continental scales[95,96], and national-scale sampling has been used to map aggregate fungal diversity from more than 100,000 molecular units[97]. Water and airborne eDNA similarly provide broad multispecies observations of aquatic and terrestrial communities[98,99], while metabarcoding combined with satellite environmental data has enabled global prediction of plankton community types[100]. These systems differ substantially in sampling substrate and observation

process, but the general principle is transferable. Standardized molecular observations can be linked to remotely sensed, climatic and other predictors to infer spatial patterns beyond sampled locations. Where DNA is transported substantially before detection, as in water or air, spatial models will additionally need to represent source areas and transport processes rather than treating detections as strictly local observations[35,101].

Taken together, our results show that national biodiversity mapping need not be restricted to well-known taxa or to aggregate diversity metrics. Preserving taxon-level predictions allows broad richness patterns to remain connected to the individual distributions that generate them, while retaining dark taxa substantially expands the represented assemblage without requiring complete taxonomic resolution. Coupling standardized molecular monitoring with scalable spatial modelling could therefore provide a route towards country-scale mapping and assessment for hyperdiverse and poorly known groups. Such maps can complement repeated field inventories by identifying broad patterns, highlighting taxonomic or geographical departures from them and directing more detailed surveys and assessments. In this role, they could contribute spatial evidence to national biodiversity monitoring and reporting, including efforts associated with the Kunming-Montreal Global Biodiversity Framework.

Methods

Insect sampling and taxonomic processing

We used DNA metabarcoding data from insect communities sampled by the Insect Biome Atlas (IBA) project for Sweden[53]. Insects were collected throughout 2019 using 203 Malaise traps distributed across 100 sites representing common Swedish habitats, including forests, wetlands,

croplands, grasslands, alpine environments and urban areas. Traps were sampled at approximately biweekly intervals. The sampling design, laboratory processing, sequencing and initial bioinformatic workflow are described in detail by Miraldo et al.[53,102]. Seven traps contained missing values for one or more environmental predictors and were therefore excluded. The final modelling dataset comprised 196 traps from 98 sites.

We used the operational taxonomic units (OTUs) and associated taxonomic annotations provided by the IBA dataset[53]. Multiple OTUs assigned to the same named species were combined so that each named species was represented once in the modelling data. OTUs lacking a species-level assignment were instead retained as separate molecular units using their cluster identifiers. Following the terminology introduced in the main text, we refer to these unresolved molecular clusters as *dark taxa*. This procedure avoided treating multiple molecular clusters assigned to the same named species as separate taxa while retaining taxonomically unresolved molecular diversity rather than discarding it. Dark taxa were not assumed to correspond one-to-one with biological species.

For named species represented by multiple OTUs, read counts were summed across those OTUs. Reads were then summed across sampling intervals within each trap and modelling unit. Each taxon was subsequently represented as present when it had been detected in at least one sample from a trap and absent otherwise, producing a trap-by-taxon presence-absence matrix for the annual sampling period. Following current taxonomic convention, OTUs previously assigned to Psocoptera or Phthiraptera were grouped within Psocodea.

To ensure sufficient information for fitting and cross-validation, we retained taxa recorded in at least 10 of the 196 traps and absent from at least 10 traps. This ensured that each response class was represented at a minimum of approximately 5% of traps and reduced the occurrence of cross-validation folds containing only presences or only absences. We retained only insect orders containing at least five eligible taxa. The final analysis comprised 11,694 taxa from 12 orders: Coleoptera, Dermaptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Neuroptera, Orthoptera, Plecoptera, Psocodea, Thysanoptera and Trichoptera. Of these taxa, 5,160 were named species and 6,534 were dark taxa.

Environmental predictors

We selected 20 environmental predictors representing broad-scale climate together with fine-scale vegetation structure, productivity, moisture conditions, habitat configuration and anthropogenic influence. Predictors were selected to capture environmental gradients expected to structure insect distributions while being available consistently across Sweden.

Climate was represented using nine variables from CHELSA version 2.1[62]: mean annual temperature (BIO1), mean diurnal temperature range (BIO2), minimum temperature of the coldest month (BIO6), mean temperature of the warmest quarter (BIO10), mean temperature of the coldest quarter (BIO11), annual precipitation (BIO12), precipitation seasonality (BIO15), precipitation of the warmest quarter (BIO18) and precipitation of the coldest quarter (BIO19).

We obtained three Sweden-specific predictors from the Swedish Forest Agency and associated national geodata services: tree height derived from airborne LiDAR and aerial imagery (*Trädhöjd*

410 3.1; decimeters)[103], soil-moisture (*Markfuktighet* SLU 2.0; 0-100)[104], and peat depth (*Torvkarta*
411 1.0, decimeters)[105]. These layers were available at 10-m spatial resolution.

412 Distance to the nearest road was derived from Sweden's National Road Database (NVDB)[106].
413 Roads were classified using the functional road class (*funktionell vägklass*), which ranges from 0
414 for the most important roads in the national network to 9 for the least important. We retained
415 classes 0-6 and excluded classes 7-9 to focus on higher-importance roads. Retained road
416 features were rasterized to a 10-m grid and Euclidean distance to the nearest road cell was
417 calculated in meters.

418 We used the CLCplus Backbone 2018 land-cover dataset (10-m resolution)[107] to derive distance
419 to surface water and signed distance to forest edge. Water cells were converted to a binary
420 water/non-water raster and Euclidean distance to the nearest water cell was calculated. The
421 three tree-cover classes (needle-leaved, broadleaved deciduous and broadleaved evergreen)
422 were combined into a binary forest mask. Forest-edge cells were identified using a 3 × 3-cell
423 moving window, and Euclidean distance to the nearest edge was calculated. Distances were
424 assigned negative values within forest and positive values outside forest, thereby retaining both
425 distance from the nearest forest boundary and whether a cell occurred inside or outside forest.

426 To characterize surrounding canopy cover and urbanization, we used the Copernicus Land
427 Monitoring Service 10-m Tree Cover Density layer for 2019 and Imperviousness Density layer for
428 2018[108,109]. For each variable, we calculated the mean within a 21 × 21-cell moving window,
429 corresponding to a 210 × 210-m neighborhood centered on each 10-m cell. The resulting
430 predictors represent local average tree-cover density and impervious surface cover, respectively.

Distance to small woody features was derived from the 2018 Copernicus Small Woody Features product[110], which maps linear woody structures such as hedgerows and discrete woody patches at 5-m resolution. We calculated Euclidean distance to the nearest mapped woody feature and aggregated the resulting distance surface to 10-m resolution by averaging. We additionally used the Copernicus 2019 vegetation phenology and productivity products for total vegetation productivity[111] and vegetation-season length[112].

All spatial predictors were transformed to SWEREF 99 TM (EPSG:3006) and aligned to a common 10-m prediction grid. Continuous predictors with coarser native resolution were resampled using bilinear interpolation. The resulting 10-m grid therefore represents the common modelling and prediction grain and does not imply that all underlying source information was available at 10-m resolution.

Predictor standardization was performed on the national raster layers before values were extracted at the sampling locations. Each predictor raster was centered and scaled, and trap-level model covariates were subsequently extracted from these standardized national layers. National predictions therefore used exactly the same standardized predictor surfaces as model fitting. No-data values were converted to missing values before extraction, and traps containing missing predictor values were excluded as described above.

Joint species distribution models

We fitted separate joint species distribution models for each insect order using the R package sjSDM[69]. The response for each order consisted of the binary occurrence matrix of all eligible taxa across the 196 traps. Models used a binomial probit formulation and simultaneously

related the occurrences of all taxa within an order to environmental predictors while estimating residual covariance among taxa. Residual covariance represents association remaining after accounting for the measured environmental predictors and may arise from unmeasured environmental responses, biotic associations or other sources of dependence[67]. We therefore did not interpret covariance estimates as direct evidence of ecological interaction.

sjSDM estimates the multivariate likelihood using Monte Carlo integration and gradient-based optimization, enabling joint modelling of assemblages containing large numbers of taxa[69]. We represented nonlinear taxon-environment relationships using a feed-forward neural network as the environmental component of the model. We evaluated five hidden-layer architectures: 40-20-20-10, 40-40-20, 40-40-40, 80-20 and 80-80-40, where numbers indicate the number of neurons in successive hidden layers. Hidden layers used LeakyReLU activation with a negative slope of 0.01 and dropout of 0.10.

For each neural-network architecture, we evaluated 100 randomly sampled combinations of elastic-net regularization parameters, yielding 500 hyperparameter configurations per insect order. Separate penalties were applied to the environmental coefficients (λ_{coef}, α_{coef}) and the residual covariance component (λ_{cov}, α_{cov}). The α parameter controls the balance between L1 (lasso) and L2 (ridge) regularization, with $\alpha = 1$ corresponding to lasso and $\alpha = 0$ to ridge, whereas λ controls the overall strength of regularization. For both components, α was sampled independently from the discrete set {0, 0.1, …, 1.0}. For the residual covariance component, λ_{cov} was sampled from 20 values equally spaced on a log2 scale from 2^{-6} to 2^{-1}, whereas λ_{coef} was sampled from 20 values between 2^{-6} to 2^{-3}. Models were fitted for 150 iterations with a learning rate of 0.001 and 1,000 Monte Carlo samples per likelihood evaluation.

Cross-validation and model selection

We evaluated predictive performance using 5-fold site-grouped cross-validation. All traps belonging to the same sampling site were assigned to the same fold, preventing neighboring traps from the same site from being distributed between training and test data. Each model configuration was therefore fitted five times, with each site group held out once for evaluation.

The area under the receiver operating characteristic curve (AUC) was the primary metric used for model selection and subsequent summaries of taxon-level discrimination. We additionally calculated the area under the precision-recall curve (PR-AUC), Brier score and log loss as complementary threshold-independent performance metrics, with the latter two assessing the quality of predicted probabilities.

For each insect order and hyperparameter configuration, taxon-level test AUCs were summarized with equal weight across taxa within each fold, and mean performance and its standard error were subsequently calculated across the five folds. We identified the configuration with the highest mean test AUC and retained all configurations whose mean test AUC was no lower than the best-performing mean minus the standard error of that best configuration. Within this one-standard-error candidate set, configurations were ranked first by mean test AUC minus their own standard error, then by the summed regularization strength of the environmental and covariance components, with stronger regularization favored, and finally by mean test AUC. The five highest-ranking configurations were retained for each insect order.

The five selected configurations were subsequently fitted to the complete modelling dataset for national prediction. For taxon-level performance summaries, cross-validated AUC was averaged

across these five selected configurations for each taxon. These taxon-level means were used to summarize discrimination among orders, detection-frequency classes, named species and dark taxa.

National prediction and richness mapping

National predictions were generated from the same standardized environmental raster layers used to extract the model-fitting covariates. Because exhaustive prediction for every taxon at every 10-m cell across Sweden was computationally prohibitive, we randomly sampled 2,000,000 10-m land cells from the national prediction domain and generated occurrence predictions at these locations. For national mapping and subsequent broad-scale spatial analyses, predictions were summarized onto an approximately 2-km grid.

For each of the five selected models within an order, taxon-specific occurrence probabilities were converted to presence-absence predictions using taxon-specific thresholds that maximized F1 on fitted predictions from the corresponding full-data model. Binary predictions were summed across taxa to obtain predicted richness within each order. Richness was calculated separately for each of the five selected models and then averaged across models. Order-level richness surfaces were then averaged across the five selected models, and richness was summed across the 12 orders to obtain total predicted insect richness.

Local landscape predictions were generated directly from the 10-m environmental layers and retained the original 10-m prediction grain. Consequently, 10 m denotes the spatial grain of the environmental model inputs and local prediction outputs rather than an independently validated ecological resolution.

Model evaluation at taxon and trap levels

Taxon-level discrimination was quantified from the site-grouped cross-validation results. For every taxon, AUC values were averaged across the five selected models for the respective insect order. Detection frequency was defined as the number of the 196 traps in which the taxon was recorded. These data were used to examine variation in discrimination among insect orders and across detection frequencies and to compare model performance between named species and dark taxa.

We additionally compared observed and fitted richness at the sampled traps. Taxon-level predictions from models fitted to the complete dataset were converted to presence-absence using the taxon-specific F1 thresholds and summed within each insect order. Richness predictions were averaged across the five selected models and compared with observed richness using Spearman's rank correlation. Because the models used in this comparison were fitted to the complete dataset, these correlations were interpreted as descriptive correspondence at sampled locations rather than independent measures of predictive performance.

Contribution and spatial patterns of dark taxa

We quantified the contribution of dark taxa both taxonomically and spatially. For each insect order, we calculated the proportion of modelled taxa classified as dark taxa and the proportion of mapped richness contributed by them. For each mapping cell, the contribution of dark taxa was calculated as dark-taxon richness divided by total richness and expressed as a percentage. National contribution was calculated from the corresponding sums across mapping cells.

To compare the national geography of named-species and dark-taxa richness independently of differences in absolute richness, cell values in each richness surface were converted separately to percentile ranks. Spatial correspondence was quantified using Spearman's rank correlation between the two sets of cell ranks.

Spatial footprints of named species and dark taxa

We compared individual prediction surfaces using a threshold-independent measure of spatial footprint. Occurrence probabilities were first averaged across the five selected models for each taxon. Spatial-footprint calculations were conducted on a common 2-km national grid.

For each taxon, probability mass in a cell was defined as predicted occurrence probability multiplied by represented cell area. Cells were ranked from highest to lowest predicted occurrence probability, and spatial footprint was defined as the minimum cumulative area required to contain 90% of total predicted probability mass. This metric therefore measures how geographically concentrated a taxon's predicted probability surface is and, unlike thresholded range size, does not depend on a taxon-specific binary occurrence threshold.

The primary analysis was restricted to taxa with cross-validated AUC ≥0.70. Spatial footprint was log-transformed and modelled as a function of taxonomic status (named species or dark taxon), with insect order included as a random intercept. The coefficient for dark taxa was back-transformed and expressed as a percentage difference relative to named species, with 95% confidence intervals.

Order-specific contrasts were estimated from a linear model of log spatial footprint including the interaction between taxonomic status and insect order, followed by estimated marginal

contrasts. Orders containing fewer than ten taxa in either group were retained in the analysis but flagged as low-sample-size comparisons. Robustness to the model-performance filter was evaluated by repeating the analysis after successively restricting taxa to cross-validated AUC ≥0.70, ≥0.75 and ≥0.80.

Order-specific richness patterns

Order-specific richness surfaces were aligned to a common approximately 2-km national grid. Spatial correspondence between the richness of each insect order and total insect richness was quantified using Spearman's rank correlation across mapped cells.

For each order, high-richness cells were defined independently as the upper 10% of that order's richness distribution. We then counted, for every national mapping cell, the number of insect orders for which it fell within this upper 10%. Overall insect-richness hotspots were likewise defined as the upper 10% of the total-richness distribution. For each insect order, we calculated the proportion of its high-richness cells that also occurred within the overall insect-richness hotspot.

Selection and analysis of illustrative taxon distributions

To illustrate the range of spatial patterns contained within the national richness surface, we selected four named taxa with high model discrimination that showed contrasting geographical distributions and spatial extents: *Limnephilus coenosus*, *Oxyethira mirabilis*, *Cacopsylla melanoneura* and *Hyperlasion wasmanni*. These taxa were selected deliberately as illustrative examples rather than as a statistically representative sample, with the aim of spanning northern

and southern distributions as well as relatively broad and restricted predicted ranges. Candidate taxa were also screened for the availability of informative local multi-trap landscapes.

For each focal taxon, occurrence probabilities were averaged across the five selected models. Geographical position was summarized by the probability-weighted centroid of the national prediction surface, and spatial extent by probability-weighted area, calculated as $\sum_i p_i A_i$, where $p_i$is predicted occurrence probability and $A_i$is represented cell area. Local 10-m maps were shown for selected multi-trap landscapes, and fitted occurrence probabilities were summarized separately for traps where the focal taxon was detected and not detected. These local comparisons were descriptive and were not treated as independent validation.

Environmental extrapolation

We assessed the degree to which national predictions extended beyond environmental conditions represented by the sampling network. For each predictor, we calculated the minimum, maximum and range across the modelling traps and compared each national prediction location with these training ranges.

For each prediction location, we first calculated the proportion of predictors lying below their training minimum or above their training maximum. For predictors outside their training range, the distance beyond the nearest boundary was divided by the predictor's observed training range. Values within the training range were assigned zero. The environmental extrapolation index was then calculated as the mean of these normalized distances across predictors. Thus, an index of zero indicates that all predictor values fell within the conditions observed at the modelling traps, whereas increasingly positive values indicate greater environmental departure.

Software and statistical analysis

Data processing, spatial analysis and statistical modelling were performed in R version 4.4.0[113]. Raster calculations were primarily implemented using *terra*[114], vector spatial operations using *sf*[115,116], and data manipulation using *dplyr*[117] and related *tidyverse* packages[118]. Joint species distribution models were fitted using *sjSDM*[69]. Mixed-effects analyses of taxon spatial footprints used *lme4*[119], and estimated marginal contrasts were obtained using *emmeans*[120].

Data availability

The underlying Swedish Insect Biome Atlas metabarcoding data are available as described by Miraldo et al.[53]. Environmental predictors are available from the public data sources cited above. Analysis-ready derived data and other outputs required to reproduce the reported statistical analyses and figures are deposited in Figshare and are available to editors and reviewers during peer review. These materials will be made publicly available upon publication.

Code availability

Code used in this study, together with the accompanying reproducibility materials, is deposited in Figshare and is available to editors and reviewers during peer review. The code will be made publicly available upon publication.

References

(1) Keck, F.; Peller, T.; Alther, R.; Barouillet, C.; Blackman, R.; Capo, E.; Chonova, T.; Couton, M.; Fehlinger, L.; Kirschner, D.; Knüsel, M.; Muneret, L.; Oester, R.; Tapolczai, K.; Zhang, H.; Altermatt, F. The Global Human Impact on Biodiversity. *Nat. 2025 6418062* **2025**, *641* (8062), 395–400. https://doi.org/10.1038/s41586-025-08752-2.

(2) Díaz, S.; Settele, J.; Brondízio, E. S.; Ngo, H. T.; Agard, J.; Arneth, A.; Balvanera, P.; Brauman, K. A.; Butchart, S. H. M. M.; Chan, K. M. A. A.; Lucas, A. G.; Ichii, K.; Liu, J.; Subramanian, S. M.; Midgley, G. F.; Miloslavich, P.; Molnár, Z.; Obura, D.; Pfaff, A.; Polasky, S.; Purvis, A.; Razzaque, J.; Reyers, B.; Chowdhury, R. R.; Shin, Y.-J. J.; Visseren-Hamakers, I.; Willis, K. J.; Zayas, C. N.; Garibaldi, L. A.; Ichii, K.; Liu, J.; Subramanian, S. M.; Midgley, G. F.; Miloslavich, P.; Molnár, Z.; Obura, D.; Pfaff, A.; Polasky, S.; Purvis, A.; Razzaque, J.; Reyers, B.; Chowdhury, R. R.; Shin, Y.-J. J.; Visseren-Hamakers, I.; Willis, K. J.; Zayas, C. N. Pervasive Human-Driven Decline of Life on Earth Points to the Need for Transformative Change. *Science (80-.).* **2019**, *366* (6471), eaax3100. https://doi.org/10.1126/science.aax3100.

(3) Mazor, T.; Doropoulos, C.; Schwarzmueller, F.; Gladish, D. W.; Kumaran, N.; Merkel, K.; Di Marco, M.; Gagic, V. Global Mismatch of Policy and Research on Drivers of Biodiversity Loss. *Nat. Ecol. Evol.* **2018**, *2* (7), 1071–1074. https://doi.org/10.1038/s41559-018-0563-x.

(4) Tilman, D.; Isbell, F.; Cowles, J. M. Biodiversity and Ecosystem Functioning. *Annu. Rev. Ecol. Evol. Syst.* **2014**, *45*, 471–493. https://doi.org/10.1146/annurev-ecolsys-120213-091917.

(5) Cardinale, B. J.; Duffy, J. E.; Gonzalez, A.; Hooper, D. U.; Perrings, C.; Venail, P.; Narwani, A.; MacE, G. M.; Tilman, D.; Wardle, D. A.; Kinzig, A. P.; Daily, G. C.; Loreau, M.; Grace, J. B.; Larigauderie, A.; Srivastava, D. S.; Naeem, S. Biodiversity Loss and Its Impact on Humanity. *Nature* **2012**, *486* (7401), 59–67. https://doi.org/10.1038/nature11148.

(6) Le Provost, G.; Schenk, N. V.; Penone, C.; Thiele, J.; Westphal, C.; Allan, E.; Ayasse, M.; Blüthgen, N.; Boeddinghaus, R. S.; Boesing, A. L.; Bolliger, R.; Busch, V.; Fischer, M.; Gossner, M. M.; Hölzel, N.; Jung, K.; Kandeler, E.; Klaus, V. H.; Kleinebecker, T.; Leimer, S.; Marhan, S.; Morris, K.; Müller, S.; Neff, F.; Neyret, M.; Oelmann, Y.; Perović, D. J.; Peter, S.; Prati, D.; Rillig, M. C.; Saiz, H.; Schäfer, D.; Scherer-Lorenzen, M.; Schloter, M.; Schöning, I.; Schrumpf, M.; Steckel, J.; Steffan-Dewenter, I.; Tschapka, M.; Vogt, J.; Weiner, C.; Weisser, W.; Wells, K.; Werner, M.; Wilcke, W.; Manning, P. The Supply of Multiple Ecosystem Services Requires Biodiversity across Spatial Scales. *Nat. Ecol. Evol. 2022 72* **2022**, *7* (2), 236–249. https://doi.org/10.1038/s41559-022-01918-5.

(7) Soliveres, S.; Van Der Plas, F.; Manning, P.; Prati, D.; Gossner, M. M.; Renner, S. C.; Alt, F.; Arndt, H.; Baumgartner, V.; Binkenstein, J.; Birkhofer, K.; Blaser, S.; Blüthgen, N.; Boch, S.; Böhm, S.; Börschig, C.; Buscot, F.; Diekötter, T.; Heinze, J.; Hölzel, N.; Jung, K.; Klaus, V. H.; Kleinebecker, T.; Klemmer, S.; Krauss, J.; Lange, M.; Morris, E. K.; Müller, J.; Oelmann, Y.; Overmann, J.; Pašalić, E.; Rillig, M. C.; Schaefer, H. M.; Schloter, M.; Schmitt, B.; Schöning, I.; Schrumpf, M.; Sikorski, J.; Socher, S. A.; Solly, E. F.; Sonnemann, I.; Sorkau, E.; Steckel, J.; Steffan-Dewenter, I.; Stempfhuber, B.; Tschapka, M.; Türke, M.; Venter, P. C.; Weiner, C. N.; Weisser, W. W.; Werner, M.; Westphal, C.; Wilcke, W.; Wolters, V.; Wubet, T.; Wurst, S.; Fischer, M.; Allan, E. Biodiversity at Multiple Trophic Levels Is Needed for Ecosystem Multifunctionality. *Nature* **2016**, *536* (7617), 456–459. https://doi.org/10.1038/nature19092.

(8) Daru, B. H.; Rodriguez, J. Mass Production of Unvouchered Records Fails to Represent Global Biodiversity Patterns. *Nat. Ecol. Evol. 2023 76* **2023**, *7* (6), 816–831.

https://doi.org/10.1038/s41559-023-02047-3.

(9) Isaac, N. J. B.; Pocock, M. J. O. Bias and Information in Biological Records. *Biol. J. Linn. Soc.* **2015**, *115* (3), 522–531. https://doi.org/10.1111/bij.12532.

(10) Troudet, J.; Grandcolas, P.; Blin, A.; Vignes-Lebbe, R.; Legendre, F. Taxonomic Bias in Biodiversity Data and Societal Preferences. *Sci. Rep.* **2017**, *7* (1), 1–14. https://doi.org/10.1038/s41598-017-09084-6.

(11) Gonzalez, A.; Vihervaara, P.; Balvanera, P.; Bates, A. E.; Bayraktarov, E.; Bellingham, P. J.; Bruder, A.; Campbell, J.; Catchen, M. D.; Cavender-Bares, J.; Chase, J.; Coops, N.; Costello, M. J.; Czúcz, B.; Delavaud, A.; Dornelas, M.; Dubois, G.; Duffy, E. J.; Eggermont, H.; Fernandez, M.; Fernandez, N.; Ferrier, S.; Geller, G. N.; Gill, M.; Gravel, D.; Guerra, C. A.; Guralnick, R.; Harfoot, M.; Hirsch, T.; Hoban, S.; Hughes, A. C.; Hugo, W.; Hunter, M. E.; Isbell, F.; Jetz, W.; Juergens, N.; Kissling, W. D.; Krug, C. B.; Kullberg, P.; Le Bras, Y.; Leung, B.; Londoño-Murcia, M. C.; Lord, J. M.; Loreau, M.; Luers, A.; Ma, K.; MacDonald, A. J.; Maes, J.; McGeoch, M.; Mihoub, J. B.; Millette, K. L.; Molnar, Z.; Montes, E.; Mori, A. S.; Muller-Karger, F. E.; Muraoka, H.; Nakaoka, M.; Navarro, L.; Newbold, T.; Niamir, A.; Obura, D.; O'Connor, M.; Paganini, M.; Pelletier, D.; Pereira, H.; Poisot, T.; Pollock, L. J.; Purvis, A.; Radulovici, A.; Rocchini, D.; Roeoesli, C.; Schaepman, M.; Schaepman-Strub, G.; Schmeller, D. S.; Schmiedel, U.; Schneider, F. D.; Shakya, M. M.; Skidmore, A.; Skowno, A. L.; Takeuchi, Y.; Tuanmu, M. N.; Turak, E.; Turner, W.; Urban, M. C.; Urbina-Cardona, N.; Valbuena, R.; Van de Putte, A.; van Havre, B.; Wingate, V. R.; Wright, E.; Torrelio, C. Z. A Global Biodiversity Observing System to Unite Monitoring and Guide Action. *Nat. Ecol.*

Evol. 2023 712 **2023**, *7* (12), 1947–1952. https://doi.org/10.1038/s41559-023-02171-0.

(12) Skidomore, A. K.; Pettorelli, N.; Coops, N. C.; Geller, G. N.; Hansen, M.; Lucas, R.; Mücher, C. A.; O'Connor, B.; Paganini, M.; Pereira, H. M.; Schaepman, M. E.; Turner, W.; Wang, T.; Wegmann, M. Agree on Biodiversity Metrics to Track from Space. *Nature* **2015**, *523*, 403–405.

(13) Proença, V.; Martin, L. J.; Pereira, H. M.; Fernandez, M.; McRae, L.; Belnap, J.; Böhm, M.; Brummitt, N.; García-Moreno, J.; Gregory, R. D.; Honrado, J. P.; Jürgens, N.; Opige, M.; Schmeller, D. S.; Tiago, P.; van Swaay, C. A. M. Global Biodiversity Monitoring: From Data Sources to Essential Biodiversity Variables. *Biol. Conserv.* **2017**, *213*, 256–263. https://doi.org/10.1016/j.biocon.2016.07.014.

(14) Dornelas, M.; Chase, J. M.; Gotelli, N. J.; Magurran, A. E.; McGill, B. J.; Antão, L. H.; Blowes, S. A.; Daskalova, G. N.; Leung, B.; Martins, I. S.; Moyes, F.; Myers-Smith, I. H.; Thomas, C. D.; Vellend, M. Looking Back on Biodiversity Change: Lessons for the Road Ahead. *Philos. Trans. R. Soc. B* **2023**, *378* (1881). https://doi.org/10.1098/RSTB.2022.0199.

(15) Magurran, A. E.; Baillie, S. R.; Buckland, S. T.; Dick, J. M. P.; Elston, D. A.; Scott, E. M.; Smith, R. I.; Somerfield, P. J.; Watt, A. D. Long-Term Datasets in Biodiversity Research and Monitoring: Assessing Change in Ecological Communities through Time. *Trends Ecol. Evol.* **2010**, *25* (10), 574–582. https://doi.org/10.1016/J.TREE.2010.06.016/ATTACHMENT/643CF97B-C65C-4C36-A5B2-661A8E0EA6D6/MMC1.DOC.

(16) Chapman, M.; Goldstein, B. R.; Schell, C. J.; Brashares, J. S.; Carter, N. H.; Ellis-Soto, D.; Faxon, H. O.; Goldstein, J. E.; Halpern, B. S.; Longdon, J.; Norman, K. E. A.; O'Rourke, D.; Scoville, C.; Xu, L.; Boettiger, C. Biodiversity Monitoring for a Just Planetary Future. *Science (80-.).* **2024**, *383* (6678), 34–36. https://doi.org/10.1126/SCIENCE.ADH8874.

(17) Johnson, T. F.; Beckerman, A. P.; Childs, D. Z.; Webb, T. J.; Evans, K. L.; Griffiths, C. A.; Capdevila, P.; Clements, C. F.; Besson, M.; Gregory, R. D.; Thomas, G. H.; Delmas, E.; Freckleton, R. P. Revealing Uncertainty in the Status of Biodiversity Change. *Nat. 2024 6288009* **2024**, *628* (8009), 788–794. https://doi.org/10.1038/s41586-024-07236-z.

(18) Pereira, H. M.; Navarro, L. M.; Martins, I. S. Global Biodiversity Change: The Bad, the Good, and the Unknown. *Annu. Rev. Environ. Resour.* **2012**, *37* (1), 25–50. https://doi.org/10.1146/annurev-environ-042911-093511.

(19) Butchart, S. H. M.; Di Marco, M.; Watson, J. E. M. Formulating Smart Commitments on Biodiversity: Lessons from the Aichi Targets. *Conserv. Lett.* **2016**, *9* (6), 457–468. https://doi.org/10.1111/CONL.12278.

(20) Affinito, F.; Williams, J. M.; Campbell, J. E.; Londono, M. C.; Gonzalez, A. Progress in Developing and Operationalizing the Monitoring Framework of the Global Biodiversity Framework. *Nat. Ecol. Evol. 2024 812* **2024**, *8* (12), 2163–2171. https://doi.org/10.1038/s41559-024-02566-7.

(21) CBD. Kunming-Montreal Global Biodiversity Framework. **2022**.

(22) Strassburg, B. B. N.; Iribarrem, A.; Beyer, H. L.; Cordeiro, C. L.; Crouzeilles, R.; Jakovac, C.

C.; Braga Junqueira, A.; Lacerda, E.; Latawiec, A. E.; Balmford, A.; Brooks, T. M.; Butchart, S. H. M.; Chazdon, R. L.; Erb, K. H.; Brancalion, P.; Buchanan, G.; Cooper, D.; Díaz, S.; Donald, P. F.; Kapos, V.; Leclère, D.; Miles, L.; Obersteiner, M.; Plutzar, C.; Carlos, C. A.; Scarano, F. R.; Visconti, P. Global Priority Areas for Ecosystem Restoration. *Nature* **2020**, *586* (7831), 724–729. https://doi.org/10.1038/s41586-020-2784-9.

(23) Rodrigues, A. S. L.; Andelman, S. J.; Bakan, M. I.; Boitani, L.; Brooks, T. M.; Cowling, R. M.; Fishpool, L. D. C.; Da Fonseca, G. A. B.; Gaston, K. J.; Hoffmann, M.; Long, J. S.; Marquet, P. A.; Pilgrim, J. D.; Pressey, R. L.; Schipper, J.; Sechrest, W.; Stuart, S. H.; Underhill, L. G.; Waller, R. W.; Watts, M. E. J.; Yan, X. Effectiveness of the Global Protected Area Network in Representing Species Diversity. *Nat. 2004 4286983* **2004**, *428* (6983), 640–643. https://doi.org/10.1038/nature02422.

(24) Margules, C. R.; Pressey, R. L. Systematic Conservation Planning. *Nat. 2000 4056783* **2000**, *405* (6783), 243–253. https://doi.org/10.1038/35012251.

(25) Guisan, A.; Tingley, R.; Baumgartner, J. B.; Naujokaitis-Lewis, I.; Sutcliffe, P. R.; Tulloch, A. I. T.; Regan, T. J.; Brotons, L.; Mcdonald-Madden, E.; Mantyka-Pringle, C.; Martin, T. G.; Rhodes, J. R.; Maggini, R.; Setterfield, S. A.; Elith, J.; Schwartz, M. W.; Wintle, B. A.; Broennimann, O.; Austin, M.; Ferrier, S.; Kearney, M. R.; Possingham, H. P.; Buckley, Y. M. Predicting Species Distributions for Conservation Decisions. *Ecol. Lett.* **2013**, *16* (12), 1424–1435. https://doi.org/10.1111/ele.12189.

(26) Pereira, H. M.; Ferrier, S.; Walters, M.; Geller, G. N.; Jongman, R. H. G.; Scholes, R. J.; Bruford, M. W.; Brummitt, N.; Butchart, S. H. M.; Cardoso, A. C.; Coops, N. C.; Dulloo, E.;

Faith, D. P.; Freyhof, J.; Gregory, R. D.; Heip, C.; Höft, R.; Hurtt, G.; Jetz, W.; Karp, D. S.; McGeoch, M. A.; Obura, D.; Onoda, Y.; Pettorelli, N.; Reyers, B.; Sayre, R.; Scharlemann, J. P. W.; Stuart, S. N.; Turak, E.; Walpole, M.; Wegmann, M. Essential Biodiversity Variables. *Science (80-.).* **2013**, *339*, 277–278.

(27) Jetz, W.; McGeoch, M. A.; Guralnick, R.; Ferrier, S.; Beck, J.; Costello, M. J.; Fernandez, M.; Geller, G. N.; Keil, P.; Merow, C.; Meyer, C.; Muller-Karger, F. E.; Pereira, H. M.; Regan, E. C.; Schmeller, D. S.; Turak, E. Essential Biodiversity Variables for Mapping and Monitoring Species Populations. *Nat. Ecol. Evol. 2019 34* **2019**, *3* (4), 539–551. https://doi.org/10.1038/s41559-019-0826-1.

(28) Meyer, C.; Kreft, H.; Guralnick, R.; Jetz, W. Global Priorities for an Effective Information Basis of Biodiversity Distributions. *Nat. Commun.* **2015**, *6* (1), 1–8. https://doi.org/10.1038/ncomms9221.

(29) Hortal, J.; de Bello, F.; Diniz-Filho, J. A. F.; Lewinsohn, T. M.; Lobo, J. M.; Ladle, R. J. Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **2015**, *46* (1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400.

(30) Phillips, S. J.; Dudík, M.; Elith, J.; Graham, C. H.; Lehmann, A.; Leathwick, J.; Ferrier, S. Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data. *Ecol. Appl.* **2009**, *19* (1), 181–197. https://doi.org/10.1890/07-2153.1.

(31) Guisan, A.; Thuiller, W. Predicting Species Distribution: Offering More than Simple Habitat Models. *Ecol. Lett.* **2005**, *8* (9), 993–1009. https://doi.org/10.1111/J.1461-

0248.2005.00792.X.

(32) Bayraktarov, E.; Ehmke, G.; O'Connor, J.; Burns, E. L.; Nguyen, H. A.; McRae, L.; Possingham, H. P.; Lindenmayer, D. B. Do Big Unstructured Biodiversity Data Mean More Knowledge? *Front. Ecol. Evol.* **2019**, *7* (JAN), 1–5. https://doi.org/10.3389/fevo.2018.00239.

(33) Bowler, D. E.; Boyd, R. J.; Callaghan, C. T.; Robinson, R. A.; Isaac, N. J. B.; Pocock, M. J. O. Treating Gaps and Biases in Biodiversity Data as a Missing Data Problem. *Biol. Rev.* **2025**, *100* (1), 50–67. https://doi.org/10.1111/BRV.13127.

(34) Bowler, D. E.; Bhandari, N.; Repke, L.; Beuthner, C.; Callaghan, C. T.; Eichenberg, D.; Henle, K.; Klenke, R.; Richter, A.; Jansen, F.; Bruelheide, H.; Bonn, A. Decision-Making of Citizen Scientists When Recording Species Observations. *Sci. Reports 2022 121* **2022**, *12* (1), 11069-. https://doi.org/10.1038/s41598-022-15218-2.

(35) Sullivan, A. R.; Karlsson, E.; Svensson, D.; Brindefalk, B.; Villegas, J. A.; Mikko, A.; Bellieny, D.; Siddique, A. B.; Johansson, A. M.; Grahn, H.; Sundell, D.; Norman, A.; Esseen, P. A.; Sjödin, A.; Singh, N. J.; Brodin, T.; Forsman, M.; Stenberg, P. Airborne EDNA Captures Three Decades of Ecosystem Biodiversity. *Nat. Commun. 2025 161* **2025**, *16* (1), 11281-. https://doi.org/10.1038/s41467-025-67676-7.

(36) Srivathsan, A.; Ang, Y.; Heraty, J. M.; Hwang, W. S.; Jusoh, W. F. A.; Kutty, S. N.; Puniamoorthy, J.; Yeo, D.; Roslin, T.; Meier, R. Convergence of Dominance and Neglect in Flying Insect Diversity. *Nat. Ecol. Evol. 2023 77* **2023**, *7* (7), 1012–1021. https://doi.org/10.1038/s41559-023-02066-0.

(37) Rohr, J. R.; Mahan, C. G.; Kim, K. C. Developing a Monitoring Program for Invertebrates: Guidelines and a Case StudyDesarrollo de Un Programa de Monitoreo de Invertebrados: Directrices y Un Estudio de Caso. *Conserv. Biol.* **2007**, *21* (2), 422–433. https://doi.org/10.1111/J.1523-1739.2006.00578.X.

(38) Wilson, E. O. Biodiversity Research Requires More Boots on the Ground. *Nat. Ecol. Evol. 2017 111* **2017**, *1* (11), 1590–1591. https://doi.org/10.1038/s41559-017-0360-y.

(39) French, C. M.; Bertola, L. D.; Carnaval, A. C.; Economo, E. P.; Kass, J. M.; Lohman, D. J.; Marske, K. A.; Meier, R.; Overcast, I.; Rominger, A. J.; Staniczenko, P. P. A.; Hickerson, M. J. Global Determinants of Insect Mitochondrial Genetic Diversity. *Nat. Commun. 2023 141* **2023**, *14* (1), 5276-. https://doi.org/10.1038/s41467-023-40936-0.

(40) Wong, M. K. L.; Didham, R. A Global Dataset of Diel Activity Patterns in Insect Communities. *Sci. Data 2024 111* **2024**, *11* (1), 558-. https://doi.org/10.1038/s41597-024-03408-8.

(41) Stork, N. E. How Many Species of Insects and Other Terrestrial Arthropods Are There on Earth? *Ann. Rev. Entomol.* **2018**, *63*, 31–45. https://doi.org/10.1146/annurev-ento-020117-043348.

(42) Yang, L. H.; Gratton, C. Insects as Drivers of Ecosystem Processes. *Curr. Opin. Insect Sci.* **2014**, *2*, 26–32. https://doi.org/10.1016/J.COIS.2014.06.004.

(43) Eisenhauer, N.; Bonn, A.; A. Guerra, C. Recognizing the Quiet Extinction of Invertebrates. *Nat. Commun. 2019 101* **2019**, *10* (1), 50-. https://doi.org/10.1038/s41467-018-07916-1.

(44) Yang, C.; Dong, J.; Jenkins, C. N.; Zhang, X.; Li, Y.; Meng, Z.; Ma, K.; Zhao, L.; Garrett, R. D.; Ellis, E. C.; Xiao, X.; Zhang, G. Identifying Conservation Priorities in Global Biodiversity Hotspots to Protect Small-Ranged Vertebrates from Agricultural Pressure. *Commun. Earth Environ. 2025 71* **2025**, *7* (1), 78-. https://doi.org/10.1038/s43247-025-03099-y.

(45) Söderström, B.; Svensson, B.; Vessby, K.; Glimskär, A. Plants, Insects and Birds in Semi-Natural Pastures in Relation to Local Habitat and Landscape Factors. *Biodivers. Conserv.* **2001**, *10* (11), 1839–1863. https://doi.org/10.1023/A:1013153427422/METRICS.

(46) Uhler, J.; Redlich, S.; Zhang, J.; Hothorn, T.; Tobisch, C.; Ewald, J.; Thorn, S.; Seibold, S.; Mitesser, O.; Morinière, J.; Bozicevic, V.; Benjamin, C. S.; Englmeier, J.; Fricke, U.; Ganuza, C.; Haensel, M.; Riebl, R.; Rojas-Botero, S.; Rummler, T.; Uphus, L.; Schmidt, S.; Steffan-Dewenter, I.; Müller, J. Relationship of Insect Biomass and Richness with Land Use along a Climate Gradient. *Nat. Commun. 2021 121* **2021**, *12* (1), 5946-. https://doi.org/10.1038/s41467-021-26181-3.

(47) Outhwaite, C. L.; McCann, P.; Newbold, T. Agriculture and Climate Change Are Reshaping Insect Biodiversity Worldwide. *Nat. 2022 6057908* **2022**, *605* (7908), 97–102. https://doi.org/10.1038/s41586-022-04644-x.

(48) Kwok, R. Ecology's Remote-Sensing Revolution. *Nature* **2018**, *556* (7699), 137–138. https://doi.org/10.1038/D41586-018-03924-9.

(49) Tuia, D.; Kellenberger, B.; Beery, S.; Costelloe, B. R.; Zuffi, S.; Risse, B.; Mathis, A.; Mathis, M. W.; van Langevelde, F.; Burghardt, T.; Kays, R.; Klinck, H.; Wikelski, M.; Couzin, I. D.; van Horn, G.; Crofoot, M. C.; Stewart, C. V.; Berger-Wolf, T. Perspectives in Machine

Learning for Wildlife Conservation. *Nat. Commun. 2022 131* **2022**, *13* (1), 792-. https://doi.org/10.1038/s41467-022-27980-y.

(50) Pettorelli, N.; Williams, J.; Schulte to Bühne, H.; Crowson, M. Deep Learning and Satellite Remote Sensing for Biodiversity Monitoring and Conservation. *Remote Sens. Ecol. Conserv.* **2025**, *11* (2), 123–132. https://doi.org/10.1002/RSE2.415.

(51) Farley, S. S.; Dawson, A.; Goring, S. J.; Williams, J. W. Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *Bioscience* **2018**, *68* (8), 563–576. https://doi.org/10.1093/BIOSCI/BIY068.

(52) Pringle, S.; Dallimer, M.; Goddard, M. A.; Le Goff, L. K.; Hart, E.; Langdale, S. J.; Fisher, J. C.; Abad, S. A.; Ancrenaz, M.; Angeoletto, F.; Auat Cheein, F.; Austen, G. E.; Bailey, J. J.; Baldock, K. C. R.; Banin, L. F.; Banks-Leite, C.; Barau, A. S.; Bashyal, R.; Bates, A. J.; Bicknell, J. E.; Bielby, J.; Bosilj, P.; Bush, E. R.; Butler, S. J.; Carpenter, D.; Clements, C. F.; Cully, A.; Davies, K. F.; Deere, N. J.; Dodd, M.; Drinkwater, R.; Driscoll, D. A.; Dutilleux, G.; Dyrmann, M.; Edwards, D. P.; Farhadinia, M. S.; Faruk, A.; Field, R.; Fletcher, R. J.; Foster, C. W.; Fox, R.; Francksen, R. M.; Franco, A. M. A.; Gainsbury, A. M.; Gardner, C. J.; Giorgi, I.; Griffiths, R. A.; Hamaza, S.; Hanheide, M.; Hayward, M. W.; Hedblom, M.; Helgason, T.; Heon, S. P.; Hughes, K. A.; Hunt, E. R.; Ingram, D. J.; Jackson-Mills, G.; Jowett, K.; Keitt, T. H.; Kloepper, L. N.; Kramer-Schadt, S.; Labisko, J.; Labrosse, F.; Lawson, J.; Lecomte, N.; de Lima, R. F.; Littlewood, N. A.; Marshall, H. H.; Masala, G. L.; Maskell, L. C.; Matechou, E.; Mazzolai, B.; McConnell, A.; Melbourne, B. A.; Miriyev, A.; Nana, E. D.; Ossola, A.; Papworth, S.; Parr, C. L.; Payo-Payo, A.; Perry, G.; Pettorelli, N.; Pillay, R.; Potts, S. G.;

Prendergast-Miller, M. T.; Qie, L.; Rolley-Parnell, P.; Rossiter, S. J.; Rowcliffe, M.; Rumble, H.; Sadler, J. P.; Sandom, C. J.; Sanyal, A.; Schrodt, F.; Sethi, S. S.; Shabrani, A.; Siddall, R.; Smith, S. C.; Snep, R. P. H.; Soulsbury, C. D.; Stanley, M. C.; Stephens, P. A.; Stephenson, P. J.; Struebig, M. J.; Studley, M.; Svátek, M.; Tang, G.; Taylor, N. K.; Umbers, K. D. L.; Ward, R. J.; White, P. J. C.; Whittingham, M. J.; Wich, S.; Williams, C. D.; Yakubu, I. B.; Yoh, N.; Zaidi, S. A. R.; Zmarz, A.; Zwerts, J. A.; Davies, Z. G. Opportunities and Challenges for Monitoring Terrestrial Biodiversity in the Robotics Age. *Nat. Ecol. Evol. 2025 96* **2025**, *9* (6), 1031–1042. https://doi.org/10.1038/s41559-025-02704-9.

(53) Miraldo, A.; Sundh, J.; Iwaszkiewicz-Eggebrecht, E.; Buczek, M.; Goodsell, R.; Johansson, H.; Fisher, B. L.; Raharinjanahary, D.; Rajoelison, E. T.; Ranaivo, C.; Randrianandrasana, C.; Rafanomezantsoa, J. J.; Manoharan, L.; Granqvist, E.; van Dijk, L. J. A.; Alberg, L.; Åhlén, D.; Aspebo, M.; Åström, S.; Bellviken, A.; Bergman, P. E.; Björklund, S.; Björkman, M. P.; Deng, J.; Desborough, L.; Dolff, E.; Eliasson, A.; Elmquist, H.; Emanuelsson, H.; Erixon, R.; Fahlen, L.; Frogner, C.; Fürst, P.; Grabs, A.; Grudd, H.; Guasconi, D.; Gunnarsson, M.; Häggqvist, S.; Hed, A.; Hörnström, E.; Jönsson, A.; Kanerot, S.; Karlsson, A.; Karlsson, D.; Klinth, M.; Kraft, T.; Lahti, R.; Larsson, M.; Lernefalk, H.; Lestander, Y.; Lindholm, L. T.; Lindholm, M.; Ljung, U.; Ljung, K.; Lundberg, J.; Lundin, E.; Malmenius, M.; Marquina, D.; Martinelli, J.; Mertz, L.; Nilsson, J.; Patchett, A.; Persson, N.; Persson, J.; Prus-Frankowska, M.; Regazzoni, E.; Rosander, K. G.; Rydgård, M.; Sandblom, C.; Skord, J.; Stålhandske, T.; Svensson, F.; Szpryngiel, S.; Tajani, K.; Tyboni, M.; Ugarph, C.; Vestermark, L.; Vilhelmsson, J.; Wahlgren, N.; Wass, A.; Wetterstrand, P.; Łukasik, P.; Tack, A. J. M.; Andersson, A. F.; Roslin, T.; Ronquist, F. Data of the Insect Biome Atlas: A Metabarcoding

Survey of the Terrestrial Arthropods of Sweden and Madagascar. *Sci. Data 2025 121* **2025**, *12* (1), 1–18. https://doi.org/10.1038/s41597-025-05151-0.

(54) Basset, Y.; Hajibabaei, M.; Wright, M. T. G.; Castillo, A. M.; Donoso, D. A.; Segar, S. T.; Souto-Vilarós, D.; Soliman, D. Y.; Roslin, T.; Smith, M. A.; Lamarre, G. P. A.; De León, L. F.; Decaëns, T.; Palacios-Vargas, J. G.; Castaño-Meneses, G.; Scheffrahn, R. H.; Rivera, M.; Perez, F.; Bobadilla, R.; Lopez, Y.; Ramirez Silva, J. A.; Cruz, M. M.; Galván, A. A.; Barrios, H. Comparison of Traditional and DNA Metabarcoding Samples for Monitoring Tropical Soil Arthropods (Formicidae, Collembola and Isoptera). *Sci. Reports 2022 121* **2022**, *12* (1), 1–16. https://doi.org/10.1038/s41598-022-14915-2.

(55) Hebert, P. D. N.; Cywinska, A.; Ball, S. L.; DeWaard, J. R. Biological Identifications through DNA Barcodes. *Proc. R. Soc. B Biol. Sci.* **2003**, *270* (1512), 313–321. https://doi.org/10.1098/RSPB.2002.2218/SUPPL_FILE/RSPB20032218S01.PDF.

(56) Page, R. D. M. DNA Barcoding and Taxonomy: Dark Taxa and Dark Texts. *Philos. Trans. R. Soc. B Biol. Sci.* **2016**, *371* (1702), 20150334. https://doi.org/10.1098/RSTB.2015.0334.

(57) Morinière, J.; Balke, M.; Doczkal, D.; Geiger, M. F.; Hardulak, L. A.; Haszprunar, G.; Hausmann, A.; Hendrich, L.; Regalado, L.; Rulik, B.; Schmidt, S.; Wägele, J. W.; Hebert, P. D. N. A DNA Barcode Library for 5,200 German Flies and Midges (Insecta: Diptera) and Its Implications for Metabarcoding-Based Biomonitoring. *Mol. Ecol. Resour.* **2019**, *19* (4), 900–928. https://doi.org/10.1111/1755-0998.13022.

(58) Chimeno, C.; Hausmann, A.; Schmidt, S.; Raupach, M. J.; Doczkal, D.; Baranov, V.; Hübner, J.; Höcherl, A.; Albrecht, R.; Jaschhof, M.; Haszprunar, G.; Hebert, P. D. N. Peering into

the Darkness: DNA Barcoding Reveals Surprisingly High Diversity of Unknown Species of Diptera (Insecta) in Germany. *Insects* **2022**, *13* (1). https://doi.org/10.3390/INSECTS13010082.

(59) Danielsen, F.; Ali, N.; Andrianandrasana, H. T.; Baquero, A.; Basilius, U.; de Araujo Lima Constantino, P.; Despot-Belmonte, K.; Frederiksen, P. O.; Isaac, M.; Jakobsen, P. K.; Klimmek, H.; Mapendembe, A.; Meng, H.; Schmidt-Vogt, D.; Sophat, S.; Sulyandziga, R.; Virnig, A. L. S.; Zhang, D.; Burgess, N. D. Involving Citizens in Monitoring the Kunming–Montreal Global Biodiversity Framework. *Nat. Sustain. 2024 712* **2024**, *7* (12), 1730–1739. https://doi.org/10.1038/s41893-024-01447-y.

(60) Hansen, M. C.; Potapov, P. V.; Moore, R.; Hancher, M.; Turubanova, S. A.; Tyukavina, A.; Thau, D.; Stehman, S. V.; Goetz, S. J.; Loveland, T. R.; Kommareddy, A.; Egorov, A.; Chini, L.; Justice, C. O.; Townshend, J. R. G. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science (80-.).* **2013**, *342* (6160), 850–853. https://doi.org/10.1126/science.1244693.

(61) Wulder, M. A.; Roy, D. P.; Radeloff, V. C.; Loveland, T. R.; Anderson, M. C.; Johnson, D. M.; Healey, S.; Zhu, Z.; Scambos, T. A.; Pahlevan, N.; Hansen, M.; Gorelick, N.; Crawford, C. J.; Masek, J. G.; Hermosilla, T.; White, J. C.; Belward, A. S.; Schaaf, C.; Woodcock, C. E.; Huntington, J. L.; Lymburner, L.; Hostert, P.; Gao, F.; Lyapustin, A.; Pekel, J. F.; Strobl, P.; Cook, B. D. Fifty Years of Landsat Science and Impacts. *Remote Sens. Environ.* **2022**, *280*, 113195. https://doi.org/10.1016/J.RSE.2022.113195.

(62) Karger, D. N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R. W.;

914 Zimmermann, N. E.; Linder, H. P.; Kessler, M. Climatologies at High Resolution for the
915 Earth's Land Surface Areas. *Sci. Data* **2017**, *4* (1), 170122.
916 https://doi.org/10.1038/sdata.2017.122.

917 (63) Mairota, P.; Cafarelli, B.; Labadessa, R.; Lovergine, F.; Tarantino, C.; Lucas, R. M.;
918 Nagendra, H.; Didham, R. K. Very High Resolution Earth Observation Features for
919 Monitoring Plant and Animal Community Structure across Multiple Spatial Scales in
920 Protected Areas. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *37*, 100–105.
921 https://doi.org/10.1016/J.JAG.2014.09.015.

922 (64) Bush, A.; Sollmann, R.; Wilting, A.; Bohmann, K.; Cole, B.; Balzter, H.; Martius, C.; Zlinszky,
923 A.; Calvignac-Spencer, S.; Cobbold, C. A.; Dawson, T. P.; Emerson, B. C.; Ferrier, S.;
924 Gilbert, M. T. P.; Herold, M.; Jones, L.; Leendertz, F. H.; Matthews, L.; Millington, J. D. A.;
925 Olson, J. R.; Ovaskainen, O.; Raffaelli, D.; Reeve, R.; Rödel, M. O.; Rodgers, T. W.; Snape,
926 S.; Visseren-Hamakers, I.; Vogler, A. P.; White, P. C. L.; Wooster, M. J.; Yu, D. W.
927 Connecting Earth Observation to High-Throughput Biodiversity Data. *Nat. Ecol. Evol.*
928 *2017 17* **2017**, *1* (7), 0176-. https://doi.org/10.1038/s41559-017-0176.

929 (65) Cavender-Bares, J.; Schneider, F. D.; Santos, M. J.; Armstrong, A.; Carnaval, A.; Dahlin, K.
930 M.; Fatoyinbo, L.; Hurtt, G. C.; Schimel, D.; Townsend, P. A.; Ustin, S. L.; Wang, Z.; Wilson,
931 A. M. Integrating Remote Sensing with Ecology and Evolution to Advance Biodiversity
932 Conservation. *Nat. Ecol. Evol. 2022 65* **2022**, *6* (5), 506–519.
933 https://doi.org/10.1038/s41559-022-01702-5.

934 (66) Brun, P.; Karger, D. N.; Zurell, D.; Descombes, P.; de Witte, L. C.; de Lutio, R.; Wegner, J.

D.; Zimmermann, N. E. Multispecies Deep Learning Using Citizen Science Data Produces More Informative Plant Community Models. *Nat. Commun. 2024 151* **2024**, *15* (1), 4421-. https://doi.org/10.1038/s41467-024-48559-9.

(67) Pollock, L. J.; Tingley, R.; Morris, W. K.; Golding, N.; O'Hara, R. B.; Parris, K. M.; Vesk, P. A.; Mccarthy, M. A. Understanding Co-Occurrence by Modelling Species Simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* **2014**, *5* (5), 397–406. https://doi.org/10.1111/2041-210X.12180.

(68) Ovaskainen, O.; Tikhonov, G.; Norberg, A.; Guillaume Blanchet, F.; Duan, L.; Dunson, D.; Roslin, T.; Abrego, N. How to Make More out of Community Data? A Conceptual Framework and Its Implementation as Models and Software. *Ecol. Lett.* **2017**, *20* (5), 561–576. https://doi.org/10.1111/ELE.12757.

(69) Pichler, M.; Hartig, F. A New Joint Species Distribution Model for Faster and More Accurate Inference of Species Associations from Big Community Data. *Methods Ecol. Evol.* **2021**, *12* (11), 2159–2173. https://doi.org/10.1111/2041-210X.13687.

(70) Buchner, D.; Sinclair, J. S.; Ayasse, M.; Beermann, A. J.; Buse, J.; Dziock, F.; Enss, J.; Frenzel, M.; Hörren, T.; Li, Y.; Monaghan, M. T.; Morkel, C.; Müller, J.; Pauls, S. U.; Richter, R.; Scharnweber, T.; Sorg, M.; Stoll, S.; Twietmeyer, S.; Weisser, W. W.; Wiggering, B.; Wilmking, M.; Zotz, G.; Gessner, M. O.; Haase, P.; Leese, F. Upscaling Biodiversity Monitoring: Metabarcoding Estimates 31,846 Insect Species from Malaise Traps across Germany. *Mol. Ecol. Resour.* **2025**, *25* (1), e14023. https://doi.org/10.1111/1755-0998.14023.

(71) Li, Y.; Devenish, C.; Tosa, M. I.; Luo, M.; Bell, D. M.; Lesmeister, D. B.; Greenfield, P.; Pichler, M.; Levi, T.; Yu, D. W. Combining Environmental DNA and Remote Sensing for Efficient, Fine-Scale Mapping of Arthropod Biodiversity. *Philos. Trans. R. Soc. B Biol. Sci.* **2024**, *379* (1904). https://doi.org/10.1098/RSTB.2023.0123/42864.

(72) Baggström, A.; Goodsell, R.; van Dijk, L.; Iwaszkiewicz-Eggebrecht, E.; Miraldo, A.; Tack, A. J. M.; Andermann, T. The Utility of Combining Deep Learning with Metabarcoding to Model Biodiversity Dynamics at a National Scale. *Ecol. Inform.* **2025**, *90*, 103318. https://doi.org/10.1016/J.ECOINF.2025.103318.

(73) Economo, E. P.; Narula, N.; Friedman, N. R.; Weiser, M. D.; Guénard, B. Macroecology and Macroevolution of the Latitudinal Diversity Gradient in Ants. *Nat. Commun. 2018 91* **2018**, *9* (1), 1778-. https://doi.org/10.1038/s41467-018-04218-4.

(74) Pinkert, S.; Farwig, N.; Kawahara, A. Y.; Jetz, W. Global Hotspots of Butterfly Diversity Are Threatened in a Warming World. *Nat. Ecol. Evol. 2025 95* **2025**, *9* (5), 789–800. https://doi.org/10.1038/s41559-025-02664-0.

(75) Hackett, T. D.; Sauve, A. M. C.; Maia, K. P.; Montoya, D.; Davies, N.; Archer, R.; Potts, S. G.; Tylianakis, J. M.; Vaughan, I. P.; Memmott, J. Multi-Habitat Landscapes Are More Diverse and Stable with Improved Function. *Nat. 2024 6338028* **2024**, *633* (8028), 114–119. https://doi.org/10.1038/s41586-024-07825-y.

(76) Schuldt, A.; Ebeling, A.; Kunz, M.; Staab, M.; Guimarães-Steinicke, C.; Bachmann, D.; Buchmann, N.; Durka, W.; Fichtner, A.; Fornoff, F.; Härdtle, W.; Hertzog, L. R.; Klein, A. M.; Roscher, C.; Schaller, J.; von Oheimb, G.; Weigelt, A.; Weisser, W.; Wirth, C.; Zhang,

J.; Bruelheide, H.; Eisenhauer, N. Multiple Plant Diversity Components Drive Consumer Communities across Ecosystems. *Nat. Commun. 2019 101* **2019**, *10* (1), 1460-. https://doi.org/10.1038/s41467-019-09448-8.

(77) Decker, O.; Uhler, J.; Redlich, S.; Chao, A.; Kortmann, M.; Steffan-Dewenter, I.; Tobisch, C.; Ewald, J.; Englmeier, J.; Fricke, U.; Ganuza, C.; Haensel, M.; Bozicevic, V.; Morinière, J.; Zhang, J.; Müller, J. Distance-Decay Reveals Contrasting Effects of Land-Use Types on Arthropod Community Homogenisation. *Nat. Commun. 2026 171* **2026**, *17* (1), 763-. https://doi.org/10.1038/s41467-025-67612-9.

(78) Shah, D. N.; Tonkin, J. D.; Haase, P.; Jähnig, S. C. Latitudinal Patterns and Large-Scale Environmental Determinants of Stream Insect Richness across Europe. *Limnologica* **2015**, *55*, 33–43. https://doi.org/10.1016/J.LIMNO.2015.11.001.

(79) Westgate, M. J.; Barton, P. S.; Lane, P. W.; Lindenmayer, D. B. Global Meta-Analysis Reveals Low Consistency of Biodiversity Congruence Relationships. *Nat. Commun. 2014 51* **2014**, *5* (1), 3899-. https://doi.org/10.1038/ncomms4899.

(80) Sánchez-Fernández, D.; Mañas-Jordá, S.; García-Meseguer, A. J.; Acosta, R.; Ariño, A.; Baquero, E.; Bartomeus, I.; de Biurrun, G.; Bonada, N.; Galicia, D.; García-Barros, E.; García-París, M.; García-Roselló, E.; Gil-Tapetado, D.; González, M.; Jurado-Angulo, P.; López-Rodríguez, M. J.; Martínez, J.; Millán, A.; Monserrat, V. J.; Morano, E.; Munguira, M. L.; Pallarés, S.; Pina, F.; Prieto, C. E.; Romo, H.; Sánchez-Campaña, C.; Sar, P.; de Figueroa, J. M. T.; Yela, J. L.; Lobo, J. M. Priority Areas for Arthropod Conservation in the Iberian Peninsula and Balearic Islands: Insights from a Multi-Taxon Distributional

Database. *Insect Conserv. Divers.* **2026**, *19* (4), 757–768. https://doi.org/10.1111/ICAD.70060.

(81) Fattorini, S.; Dennis, R. L. H.; Cook, L. M. Use of Cross-Taxon Congruence for Hotspot Identification at a Regional Scale. *PLoS One* **2012**, *7* (6), e40018. https://doi.org/10.1371/JOURNAL.PONE.0040018.

(82) Bonin, A.; Guerrieri, A.; Ficetola, G. F. Optimal Sequence Similarity Thresholds for Clustering of Molecular Operational Taxonomic Units in DNA Metabarcoding Studies. *Mol. Ecol. Resour.* **2023**, *23* (2), 368–381. https://doi.org/10.1111/1755-0998.13709.

(83) Sundh, J.; Granqvist, E.; Iwaszkiewicz-Eggebrecht, E.; Manoharan, L.; Van Dijk, L. J. A.; Goodsell, R.; Godeiro, N. N.; Bellini, B. C.; Orsholm, J.; Łukasik, P.; Miraldo, A.; Roslin, T.; Tack, A. J. M.; Andersson, A. F.; Ronquist, F. HAPP: High-Accuracy Pipeline for Processing Deep Metabarcoding Data. *PLOS Comput. Biol.* **2025**. https://doi.org/10.1371/journal.pcbi.1013558.

(84) Ratnasingham, S.; Hebert, P. D. N. A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. **2013**, *8* (7), e66213. https://doi.org/10.1371/JOURNAL.PONE.0066213.

(85) Iwaszkiewicz-Eggebrecht, E.; Goodsell, R. M.; Bengsson, B. Å.; Mutanen, M.; Klinth, M.; Van Dijk, L. J. A.; Łukasik, P.; Miraldo, A.; Andersson, A.; Tack, A. J. M.; Roslin, T.; Ronquist, F. High-Throughput Biodiversity Surveying Sheds New Light on the Brightest of Insect Taxa. **2025**, *292* (2046).

(86) Kirse, A.; Bourlat, S. J.; Langen, K.; Fonseca, V. G. Metabarcoding Malaise Traps and Soil EDNA Reveals Seasonal and Local Arthropod Diversity Shifts. *Sci. Reports 2021 111* **2021**, *11* (1), 10498-. https://doi.org/10.1038/s41598-021-89950-6.

(87) Krehenwinkel, H.; Wolf, M.; Lim, J. Y.; Rominger, A. J.; Simison, W. B.; Gillespie, R. G. Estimating and Mitigating Amplification Bias in Qualitative and Quantitative Arthropod Metabarcoding. *Sci. Reports 2017 71* **2017**, *7* (1), 17668-. https://doi.org/10.1038/s41598-017-17333-x.

(88) Skvarla, M. J.; Larson, J. L.; Fisher, J. R.; Dowling, A. P. G. A Review of Terrestrial and Canopy Malaise Traps. *Ann. Entomol. Soc. Am.* **2021**, *114* (1), 27–47. https://doi.org/10.1093/AESA/SAAA044.

(89) Ortiz, A. S.; Rubio, R. M.; Guerrero, J. J.; Garre, M. J.; Serrano, J.; Hebert, P. D. N.; Hausmann, A. Close Congruence between Barcode Index Numbers (Bins) and Species Boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. *Biodivers. Data J.* **2017**, *5* (1). https://doi.org/10.3897/BDJ.5.E19840.

(90) Moudrý, V.; Bazzichetto, M.; Remelgado, R.; Devillers, R.; Lenoir, J.; Mateo, R. G.; Lembrechts, J. J.; Sillero, N.; Lecours, V.; Cord, A. F.; Barták, V.; Balej, P.; Rocchini, D.; Torresani, M.; Arenas-Castro, S.; Man, M.; Prajzlerová, D.; Gdulová, K.; Prošek, J.; Marchetto, E.; Zarzo-Arias, A.; Gábor, L.; Leroy, F.; Martini, M.; Malavasi, M.; Cazzolla Gatti, R.; Wild, J.; Šímová, P. Optimising Occurrence Data in Species Distribution Models: Sample Size, Positional Uncertainty, and Sampling Bias Matter. *Ecography (Cop.).* **2024**, *2024* (12), e07294. https://doi.org/10.1111/ECOG.07294.

(91) Meyer, H.; Pebesma, E. Predicting into Unknown Space? Estimating the Area of Applicability of Spatial Prediction Models. *Methods Ecol. Evol.* **2021**, *12* (9), 1620–1633. https://doi.org/10.1111/2041-210X.13650.

(92) MacKenzie, D. I.; Nichols, J. D.; Lachman, G. B.; Droege, S.; Andrew Royle, J.; Langtimm, C. A. Estimating Site Occupancy Rates When Detection Probabilities Are Less than One. *Ecology* **2002**, *83* (8), 2248–2255.

(93) Phillips, S. J.; Elith, J. POC Plots: Calibrating Species Distribution Models with Presence-Only Data. *Ecology* **2010**, *91* (8), 2476–2484. https://doi.org/10.1890/09-0760.1.

(94) Mesgaran, M. B.; Cousens, R. D.; Webber, B. L. Here Be Dragons: A Tool for Quantifying Novelty Due to Covariate Range and Correlation Change When Projecting Species Distribution Models. *Divers. Distrib.* **2014**, *20* (10), 1147–1159. https://doi.org/10.1111/DDI.12209.

(95) George, P. B. L.; Lallias, D.; Creer, S.; Seaton, F. M.; Kenny, J. G.; Eccles, R. M.; Griffiths, R. I.; Lebron, I.; Emmett, B. A.; Robinson, D. A.; Jones, D. L. Divergent National-Scale Trends of Microbial and Animal Biodiversity Revealed across Diverse Temperate Soil Ecosystems. *Nat. Commun. 2019 101* **2019**, *10* (1), 1107-. https://doi.org/10.1038/s41467-019-09031-1.

(96) Labouyrie, M.; Ballabio, C.; Romero, F.; Panagos, P.; Jones, A.; Schmid, M. W.; Mikryukov, V.; Dulya, O.; Tedersoo, L.; Bahram, M.; Lugato, E.; van der Heijden, M. G. A.; Orgiazzi, A. Patterns in Soil Microbial Diversity across Europe. *Nat. Commun. 2023 141* **2023**, *14* (1), 3311-. https://doi.org/10.1038/s41467-023-37937-4.

(97) Djemiel, C.; Dequiedt, S.; Horrigue, W.; Bailly, A.; Lelièvre, M.; Tripied, J.; Guilland, C.; Perrin, S.; Comment, G.; Saby, N. P. A.; Jolivet, C.; Bispo, A.; Boulonne, L.; Pierart, A.; Wincker, P.; Cruaud, C.; Maron, P. A.; Terrat, S.; Ranjard, L. Unraveling Biogeographical Patterns and Environmental Drivers of Soil Fungal Diversity at the French National Scale. *SOIL* **2024**, *10* (1), 251–273. https://doi.org/10.5194/SOIL-10-251-2024.

(98) Carraro, L.; Mächler, E.; Wüthrich, R.; Altermatt, F. Environmental DNA Allows Upscaling Spatial Patterns of Biodiversity in Freshwater Ecosystems. *Nat. Commun. 2020 111* **2020**, *11* (1), 1–12. https://doi.org/10.1038/s41467-020-17337-8.

(99) Tournayre, O.; Littlefair, J. E.; Garrett, N. R.; Allerton, J. J.; Brown, A. S.; Cristescu, M. E.; Clare, E. L. First National Survey of Terrestrial Biodiversity Using Airborne EDNA. *Sci. Reports 2025 151* **2025**, *15* (1), 19247-. https://doi.org/10.1038/s41598-025-03650-z.

(100) Kaneko, H.; Endo, H.; Henry, N.; Berney, C.; Mahé, F.; Poulain, J.; Labadie, K.; Beluche, O.; Hourany, R. El; Chaffron, S.; Wincker, P.; Nakamura, R.; Karp-Boss, L.; Boss, E.; Bowler, C.; de Vargas, C.; Tomii, K.; Ogata, H.; Acinas, S. G.; Babin, M.; Bork, P.; Cochrane, G.; de Vargas, C.; Gorsky, G.; Guidi, L.; Grimsley, N.; Hingamp, P.; Iudicone, D.; Jaillon, O.; Kandels, S.; Karsenti, E.; Not, F.; Poulton, N.; Pesant, S.; Sardet, C.; Speich, S.; Stemmann, L.; Sullivan, M. B.; Sunagawa, S. Predicting Global Distributions of Eukaryotic Plankton Communities from Satellite Data. *ISME Commun. 2023 31* **2023**, *3* (1), 101-. https://doi.org/10.1038/s43705-023-00308-7.

(101) Carraro, L.; Blackman, R. C.; Altermatt, F. Modelling Environmental DNA Transport in Rivers Reveals Highly Resolved Spatio-Temporal Biodiversity Patterns. *Sci. Reports 2023*

131 **2023**, *13* (1), 8854-. https://doi.org/10.1038/s41598-023-35614-6.

(102) Miraldo, A.; Iwaszkiewicz-Eggebrecht, E.; Sundh, J.; Manoharan, L.; Granqvist, E.; Andersson, A.; Lukasik, P.; Roslin, T.; Tack, A. J. M.; Ronquist, F. Amplicon Sequence Variants from the Insect Biome Atlas Project. **2024**. https://doi.org/10.17044/scilifelab.25480681.v5.

(103) Skogsstyrelsen. Trädhöjd Laserdata Skog-Produktbeskrivning. *Accessed via Skogsstyrelsen ArcGIS REST ImageServer, Tradhojd_3_1.* **2025**.

(104) Ågren, A. M.; Larson, J.; Paul, S. S.; Laudon, H.; Lidberg, W. Use of Multiple LIDAR-Derived Digital Terrain Indices and Machine Learning for High-Resolution National-Scale Soil Moisture Mapping of the Swedish Forest Landscape. *Geoderma* **2021**, *404*, 115280. https://doi.org/10.1016/J.GEODERMA.2021.115280.

(105) Ågren, A. M.; Hasselquist, E. M.; Stendahl, J.; Nilsson, M. B.; Paul, S. S. Delineating the Distribution of Mineral and Peat Soils at the Landscape Scale in Northern Boreal Regions. *SOIL* **2022**, *8* (2), 733–749. https://doi.org/10.5194/SOIL-8-733-2022.

(106) Trafikverket. Dataproduktspecifikation – Funktionell Vägklass. *Version 9.0, 24 Sept. 2020. Trafikverket, Sweden.* **2020**.

(107) European Union's Copernicus Land Monitoring Service information. CLCplus Backbone 2018 (Raster 10 m), Europe, 3-Yearly. European Environment Agency. **2023**. https://doi.org/10.2909/cd534ebf-f553-42f0-9ac1-62c1dc36d32c.

(108) European Union's Copernicus Land Monitoring Service information. Tree Cover Density

2018–Present (Raster 10 m), Europe, Yearly. European Environment Agency. **2024**. https://doi.org/10.2909/e677441e-fb94-431c-b4f9-304f10e4dfd8.

(109) European Union's Copernicus Land Monitoring Service information. Imperviousness Density 2018 (Raster 10 m), Europe, 3-Yearly. European Environment Agency. **2020**. https://doi.org/10.2909/3bf542bd-eebd-4d73-b53c-a0243f2ed862.

(110) European Union's Copernicus Land Monitoring Service information. Small Woody Features 2018 (Raster 5 m), Europe, 3-Yearly. European Environment Agency. **2023**. https://doi.org/10.2909/a8e683b1-2f96-45c8-827f-580a79413018.

(111) European Union's Copernicus Land Monitoring Service information. Total Productivity 2017–Present (Raster 10 m), Europe, Yearly. European Environment Agency. **2021**. https://doi.org/10.2909/977e4bb8-407f-48ec-b4c4-403bca5a6a3b.

(112) European Union's Copernicus Land Monitoring Service information. Season Length 2017–Present (Raster 10 m), Europe, Yearly. European Environment Agency. **2021**. https://doi.org/10.2909/430c008c-7298-473e-a4b0-7c0a287446a6.

(113) R Core Team. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput. Vienna, Austria.* **2024**.

(114) Hijmans, R. J. Terra: Spatial Data Analysis. 2025. https://cran.r-project.org/package=terra.

(115) Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10* (1), 439. https://doi.org/10.32614/RJ-2018-009.

(116) Pebesma, E.; Bivand, R. *Spatial Data Science: With Applications in R*; Chapman and Hall/CRC, 2023. https://doi.org/10.1201/9780429459016.

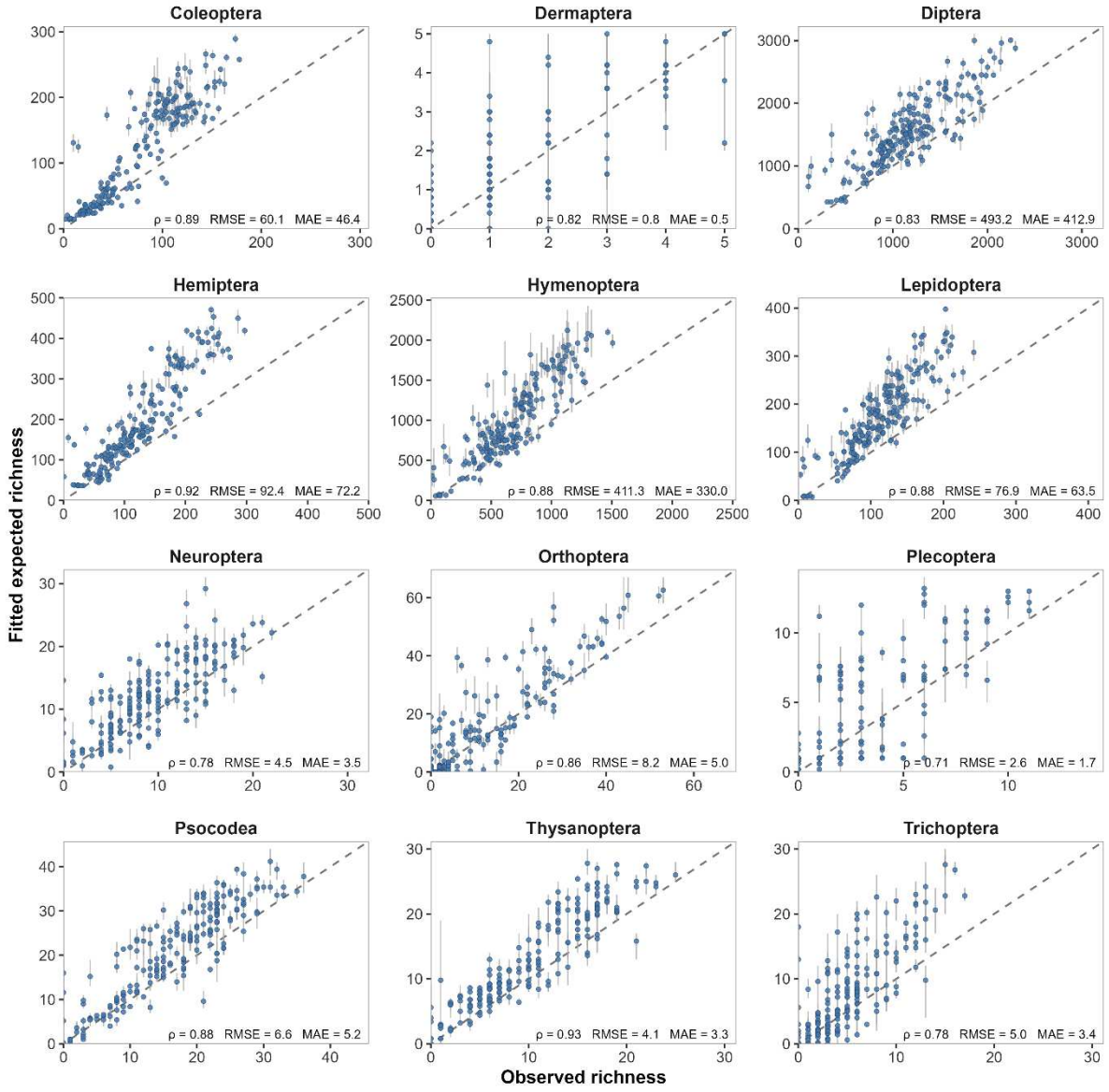
(117) Wickham, H.; Francois, R.; Henry, L.; Muller, K.; Vaughan, D. Dplyr: A Grammar of Data Manipulation. 2026. https://cran.r-project.org/package=dplyr.

(118) Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L. D.; François, R.; Grolemun, G.; Haye, A.; Henr, L.; Heste, J.; Kuh, M.; Pederse, T. L.; Mille, E.; Bach, S. M.; Müll, K.; Oo, , Jeroen; Robins, , David; Seid, , Dana Paige; Spi, , Vitalie; Takahas, , Kohske; Vaugh, , Davis; Wil, , Claus; W, , Kara; Yutani, , Hiroaki. Welcome to the {tidyverse}. *J. Open Source Softw.* **2019**, *4* (43), 1686. https://doi.org/10.21105/joss.01686.

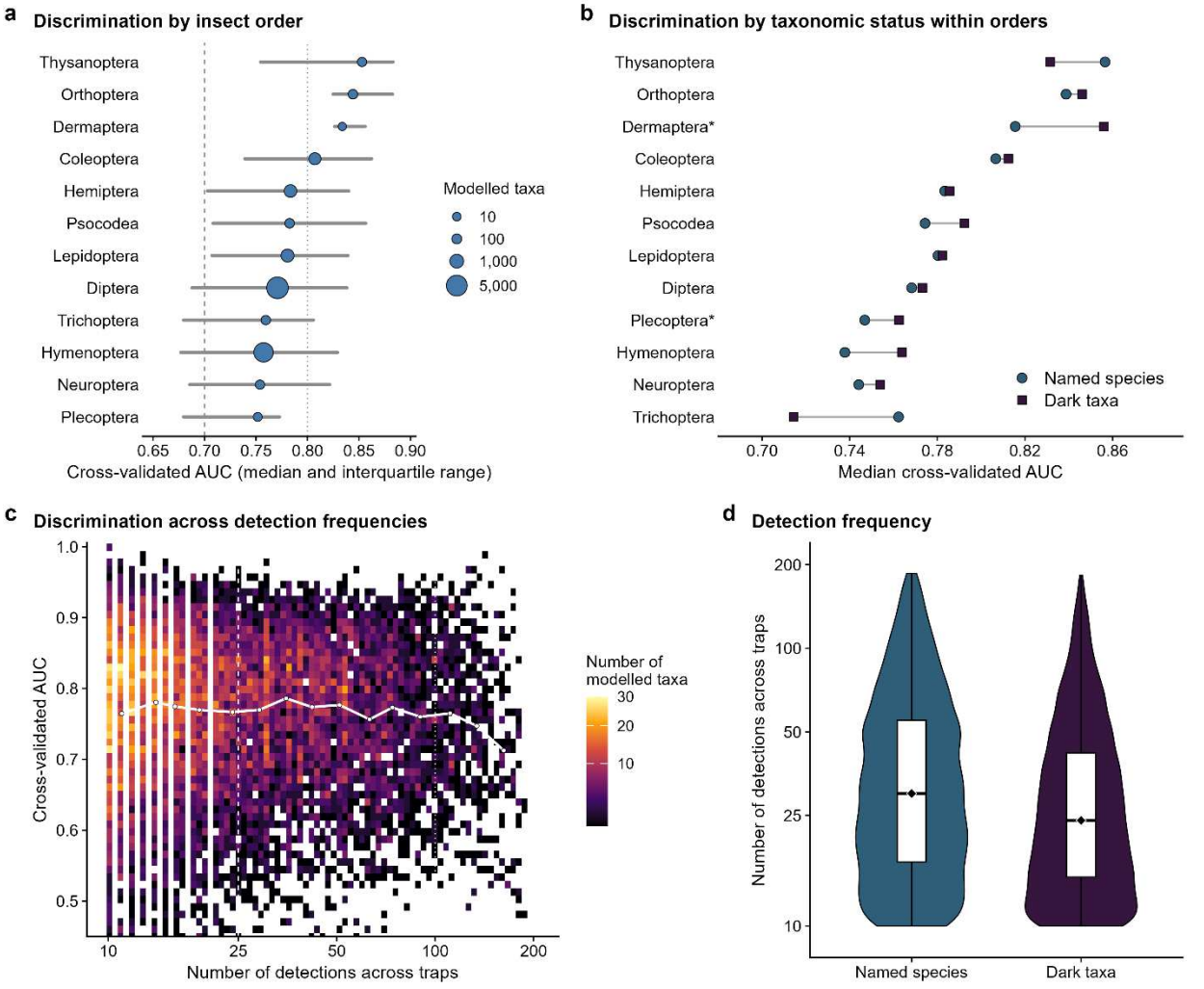
(119) Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using Lme4. *J. Stat. Softw.* **2015**, *67* (1). https://doi.org/10.18637/jss.v067.i01.

(120) Lenth, R. V. Emmeans: Estimated Marginal Means, Aka Least-Squares Means. 2024. https://cran.r-project.org/package=emmeans.

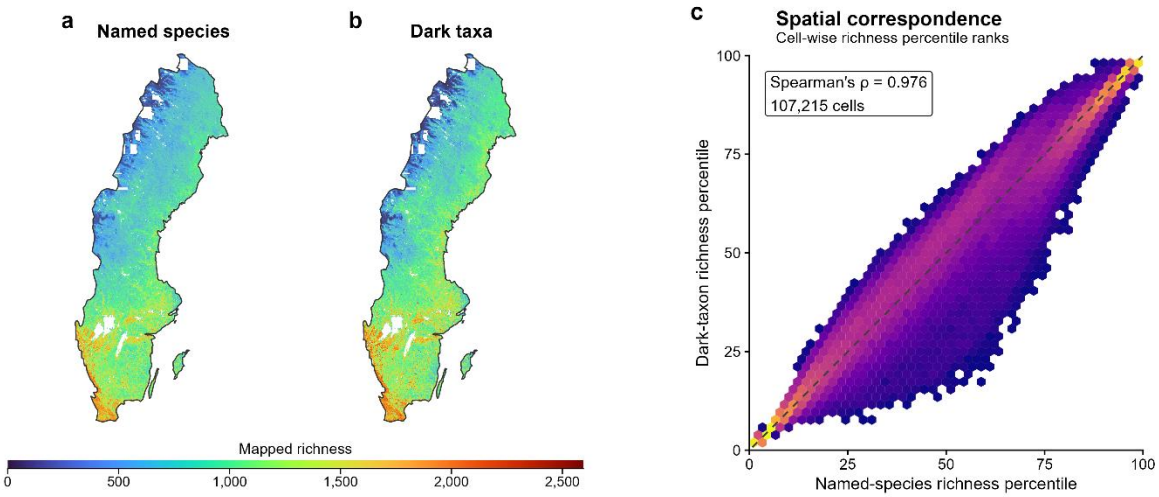
Extended Data



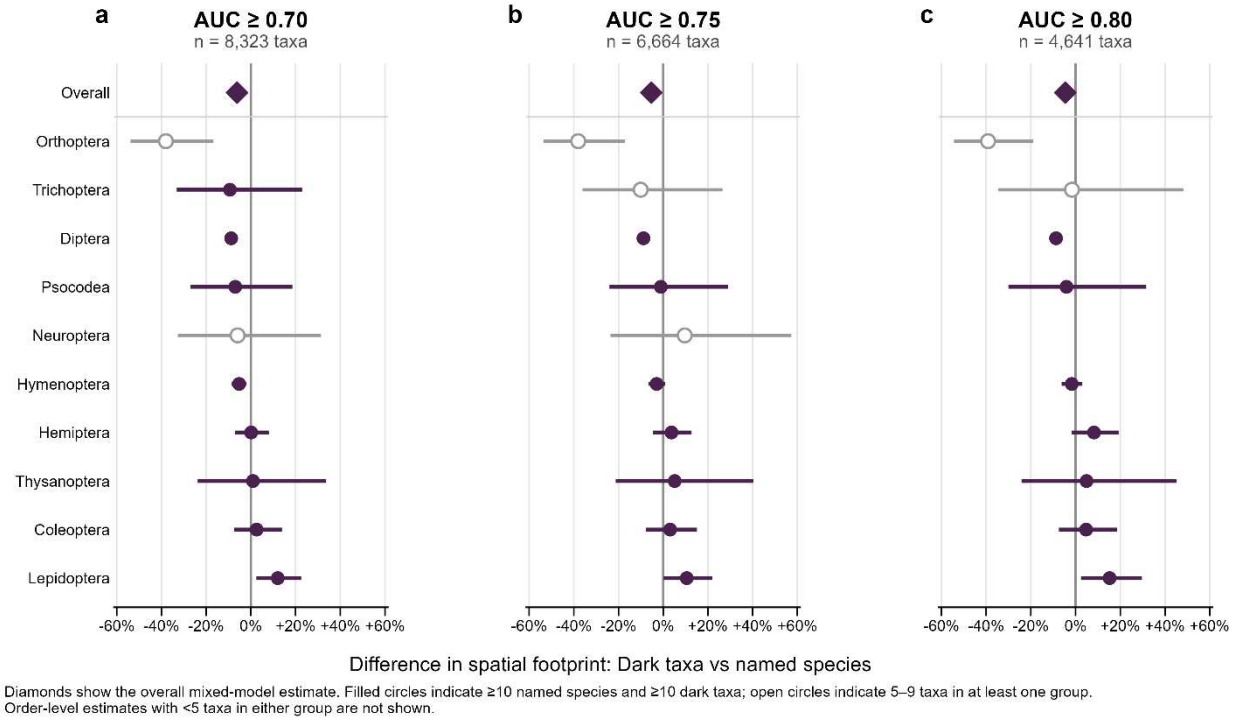
Extended Data Figure 1: Correspondence between observed and predicted insect richness across sampling sites. Observed richness at Malaise traps is shown against predicted richness for each of 12 insect orders. Predicted richness was obtained by converting taxon-level occurrence probabilities to presence-absence predictions using taxon-specific F1 thresholds and summing these predictions within each order. Points show the mean predicted richness across the five selected joint species distribution models, and vertical lines show the minimum-to-maximum range across these models. Dashed lines indicate the 1:1 relationship. Spearman's rank correlation coefficient (ρ), root mean squared error (RMSE) and mean absolute error (MAE) were calculated from the mean prediction across the five models. Because models were fitted using the full dataset, these comparisons describe correspondence at sampled sites rather than independent predictive performance.



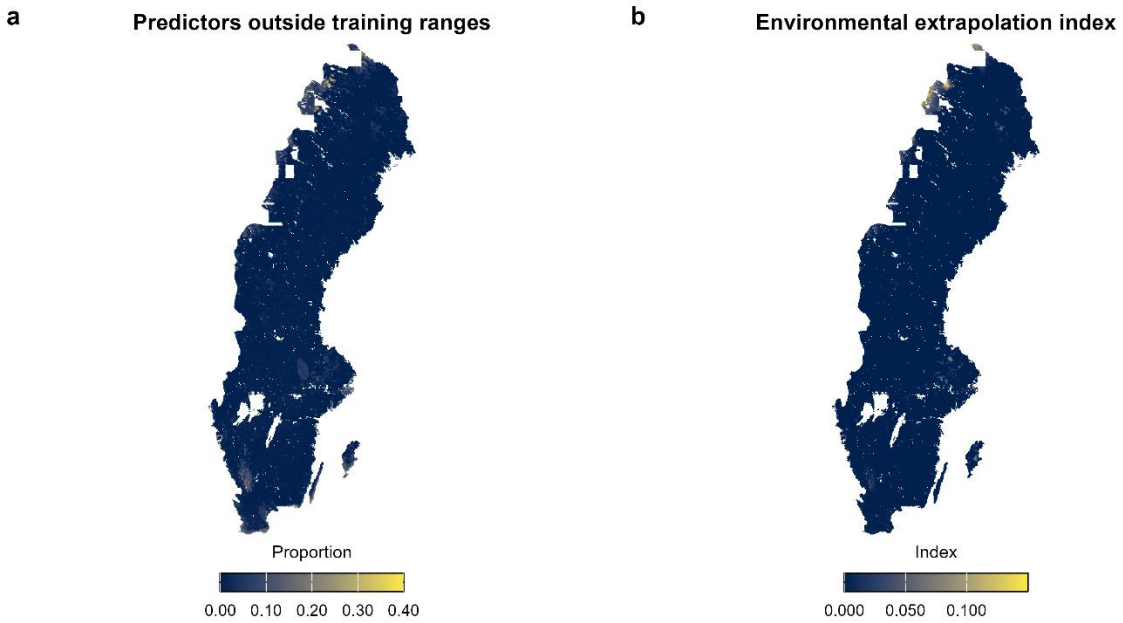
Extended Data Figure 2: Cross-validated model discrimination across insect orders, detection frequencies and taxonomic status. a) Median site-grouped cross-validated area under the receiver operating characteristic curve (AUC; points) and interquartile range (horizontal lines) for taxa in each of 12 insect orders. Point size indicates the number of modelled taxa; vertical reference lines mark AUC values of 0.7 and 0.8. b) Median cross-validated AUC for named species and dark taxa within each insect order. Dark taxa are defined here as molecular clusters lacking species-level taxonomic assignments. Lines connect the two taxonomic groups within orders. Asterisks indicate orders for which either the named-species or dark-taxon group contained fewer than ten taxa. c) Relationship between the number of trap detections and cross-validated AUC across all 11,694 modelled taxa. Color indicates the number of taxa in each two-dimensional bin, and the white line connects median AUC values for logarithmically spaced detection-frequency bins containing at least 50 taxa. Vertical reference lines mark 25 and 100 detections. d) Distribution of detection frequencies for 5,160 named species and 6,534 dark taxa. Violins show the distributions, white boxes show the interquartile ranges and diamonds indicate the medians; detection frequency is shown on a logarithmic scale.



Extended Data Figure 3: Spatial correspondence between mapped richness of named species and dark taxa. a,b) Mapped richness of named species (a) and dark taxa (b), defined here as molecular clusters lacking species-level taxonomic assignments, across Sweden. Both maps are shown on the same color scale. National predictions were generated for 2 million randomly sampled 10-m cells distributed across Sweden and averaged within approximately 2-km grid cells for national-scale visualization. c) Cell-wise correspondence between named-species and dark-taxon richness, expressed as percentile ranks across mapping cells. Hexagon color indicates the number of mapping cells, and the dashed line shows the 1:1 relationship.



Extended Data Figure 4: Robustness of dark-taxon spatial-footprint differences to model-performance filtering. a-c) Differences in predicted spatial footprint between dark taxa and named species after progressively restricting analyses to taxa with cross-validated AUC ≥ 0.70 (a), ≥ 0.75 (b) and ≥ 0.80 (c). Spatial footprint was quantified as the area required to contain 90% of a taxon's predicted probability mass. Diamonds show the overall dark-taxon vs. named-species difference estimated across insect orders, and circles show order-specific differences; horizontal lines indicate 95% confidence intervals. Negative values indicate smaller spatial footprints for dark taxa than for named species. Filled circles denote comparisons represented by at least ten named species and ten dark taxa, whereas open circles denote comparisons with 5-9 taxa in at least one group; order-level comparisons with fewer than five taxa in either group are not shown. The numbers above each panel give the total number of taxa retained at the corresponding AUC threshold.



Extended Data Figure 5: Environmental extrapolation in national insect predictions. a) Proportion of the 20 environmental predictors for which conditions in each mapped cell fell outside the range represented by the sampling network. b) Environmental extrapolation index, calculated as the mean normalized distance beyond the corresponding training range across the environmental predictors. Higher values indicate cells in which predictions rely more strongly on environmental extrapolation. At least one predictor fell outside its training range across 43.5% of mapped land, but only 1.5% of the area had at least 3 predictors outside their training ranges and 0.6% exceeded an extrapolation index of 0.05. Stronger extrapolation was concentrated locally, particularly in northern and northwestern Sweden.