Historic residential segregation impacts biodiversity data availability disparately across the tree of life

Dexter Henry Locke[1*], Melissa Chapman[2], Diego Ellis-Soto[3]

1 USDA Forest Service, Northern Research Station, Baltimore Field Station, Suite 350, 5523 Research Park Drive, Baltimore, MD21228, USA ORCID: 0000-0003-2704-9720

2 National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, USA ORCID: 0000-0002-1377-1539

3 Department of Ecology and Evolutionary Biology, Yale University, New Haven, USA ORCID: 0000-0003-4766-021X

*Dexter Henry Locke
Email: dexter.locke@usda.gov; dexter.locke@gmail.com

Author Contributions: DHL, MC and DE-S designed the study and performed the research. DHL performed the analysis with contributions from MC and DE-S. DE-S created figure 1. DHL led the writing, MC and DE-S contributed substantially to the writing with multiple rounds of edits.

Competing Interest Statement: None.
Classification: Paste the major and minor classification here. Dual classifications are permitted, but cannot be within the same major classification.

Keywords: residential segregation, biodiversity sampling, environmental justice, Global Biodiversity Information Facility, Home Owners' Loan Corporation

This PDF file includes:

Main Text
Figures 2

Abstract

Urbanization alters species ranges and nature's contributions to people, motivating urban conservation. Residential segregation policies have left an indelible impact on urban environments, greenspaces, and wildlife communities, creating socioeconomic heterogeneity and altering biota. However, the extent to which data sufficiently capture urban biodiversity patterns remains unclear, especially when considering historic segregation. We explore how biodiversity metrics (sampling density, estimated completeness of sampling, and expected species richness) vary by Home Owner's Loan Corporation (HOLC) grade across taxonomic groups, leveraging nearly 60 million amphibia, aves, fungi, insecta, mammalia, plantae, and reptilia observations collected between 2000 and 2020, for 145 Metropolitan Statistical Areas in the United States. After accounting for environmental conditions, we estimate significant differences in sampling density across HOLC grade for all taxonomic groups, with the lowest values found in areas previously redlined. Estimated completeness of biodiversity inventory was low (average ~42% across all taxa) and varied significantly by HOLC grade for birds, mammals, and plants. Expected richness only varied by HOLC grade for birds. Our findings highlight how differences in biodiversity sampling may not translate to differences in expected species richness patterns, and suggest that applying insights obtained from certain taxonomic groups and extrapolating to multiple others may not be appropriate. Urban wildlife communities are not well-documented despite the explosion of digital information, and what is documented is known to be biased along a housing segregation typology for some taxon. These findings add evidence to suggest long-lasting effects of legacies of segregation on the natural world.

Significance Statement

Historic race-based zoning policies like redlining in the United States are associated with present day health, income, and environmental inequities. We quantify how redlining across 195 cities in the United States is also related to key biodiversity metrics across a wide range of vertebrate and invertebrate taxa, plants and fungi. We show that while more biodiversity records are consistently collected in non-redlined neighborhoods, this did not translate to differences in estimated species richness across redlining grades. This work underpins how legacies of segregation and socioeconomic inequality may influence the distribution and availability of data on urban biodiversity, and how such biased biodiversity data in turn may influence our inference on species communities, their food webs, and ultimately, conservation decisions.

Main Text

Introduction

Global urbanization projections suggests a 55 to 111% increase in area, translating to a loss of 11-33 million hectares of habitat from 2015 to the year 2100 (1). A key features of the Anthropocene is the increasing rise of urban life and urban expansion, with approximately half of humans residing in cities, which is projected to grow to 2/3 by 2050 (2). The last decade has seen an increased appreciation on the importance of urban biodiversity for promoting physical and psychological well-being of city residents (3). Cities are the places where human experiences with biodiversity increasingly occurs for most humans (4), and where a growing proportion of wildlife face urban pressures. Urbanization therefore poses both opportunities and challenges for biodiversity conservation (5), particularly given disparate responses of species and taxa to urbanization (6). As Lambert and Schell describe, “it is not hyperbolic to suggest that cities are situated as the literal and figurative frontlines of biodiversity conservation” (7).

Urban areas represent complex systems strongly shaped by social and economic factors that are often characterized by social inequity. Socioeconomic disparities are in turn associated with the spatial distribution of urban tree canopy cover, with higher income areas having more tree canopy, and minoritized communities having less (8–11). Tree canopy and urban green spaces provide crucial habitat for biodiversity, form the basis of more complex ecological communities, and shape urban food webs (11). Therefore the socioeconomic partitioning of urban spaces is expected to shape multiple facets of urban biodiversity (12–14), and even evolutionary processes and outcomes (12, 15–17).

Simultaneously, urban areas are increasingly places of extensive biodiversity data collection, primarily through participatory-science and education initiatives leveraging mobile phone data collection apps (18). For some species, particularly in urban environments, volunteer-collected data far exceed records museum collections (19). There are biodiversity data disparities within and across countries: higher income countries have more information (20, 21), and higher income areas within high income countries have the most (22). These data biases skew the our view of the natural world and mean that minoritized communities are often also data-poor (22–24), which may represent another form of environmental injustice.

Institutionalized racism is a major driver of social inequity, especially in cities (25). A spatial manifestation of institutionalized racism is housing segregation. One particularly well-known mapping of this housing segregation in cities across the United States was the Home Owners’ Loan Corporation (HOLC) in the mid to late 1930s, commonly known as Redlining. The Home Owners’ Loan Corporation, with input from local real estate actors, categorized neighborhoods based on a combination of housing stock (type, quality, age), favorable adjacent land uses such as parks and open space, proximity to transit, and the demographic and racial characteristics of the inhabitants. A-Graded areas were composed of U.S.-born White families living in new, single-family detached homes, were labeled “Best” and colored green on the maps. B-Graded or blue areas, labeled “Still Desirable” had older and/or denser housing stock. C-Graded or yellow areas, labeled “Definitely Declining” had more minoritized populations such as communities of color and/or immigrants. Finally, D-Graded or red areas, hence the name “redlining”, were labeled “Hazardous” and were composed of communities of color. It is important to note that the practices associated with redlining predate the maps. These practices are associated with covenants, codes, and restrictions (26); segregated newspaper advertisements for housing, among many others. These practices began in the early 1900s and continue today (27–30).

The urban ecology literature on redlining documents substantial disparities across neighborhood grades. Formerly A-Graded neighborhoods have more vegetation (31), more tree canopy (32–34), are cooler (35), and exhibit less noise pollution (36) than their formerly D-Graded counterparts. This means neighborhoods formerly comprised of US-born Whites in single family detached homes are more hospitable today – for both people and other species – than areas classified as “Hazardous”, marked red on maps by HOLC, and denied home loans, because they were occupied by poorer communities of color and immigrants living in denser, older housing. In Baltimore, MD, street trees are larger and more species diverse in A-Graded areas than their formerly D-Graded counterparts (37), so they produces more ecosystems services *and* are more resilient to urban forest pathogens. Redlined neighborhoods in California have higher pollution burdens, less vegetation, hotter temperatures, and more noise pollution than A-Graded areas (38).

In addition to vegetation and street tree diversity, bird biodiversity data and species composition knowledge is significantly greater in A than D areas, with differences persisting even after controlling for population density, vegetation greenness, and protected open space (24). Moreover, field-based biodiversity assessments further showed that bird communities in Los Angeles vary across HOLC grades (39). For example forest birds and migratory birds were ~24% and ~17% more abundant, respectively, in formerly A- and B-Graded areas than C and D areas, while non-migratory, introduced, and synanthropic dominated C-, and D-Graded areas (39). With discrepancies between volunteered bird data A- and D-Graded areas growing over time (24), there is pressing need to understand how housing segregation and urban biodiversity data relate to additional taxa. Early multi-taxon work using the citizen science platform iNaturalist in four Californian cities, shows that redlined areas have a lower number of insects, birds, and mammals species, and that species composition vary by HOLC grade (40).

This paper contributes to the ongoing efforts address questions and test hypotheses about housing segregation, specifically how race-based housing policies multiple facets of urban biodiversity (12). Synthesizing across platforms, the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) includes data from iNaturalist, eBird, other popular taxon-specific apps, as well as from participant node organizations composed of scientific research entities like universities and museums. Building off prior research, we leverage 58,920,460 species observations from GBIF (41) across metropolitan areas in the United States to assess how the amount of biodiversity information (sampling density), knowledge of species pools (completeness), and expected species richness varies by HOLC grades, Urban Areas (UAs) and Metropolitan Statistical Areas (MSAs). Sampling density answers the question about whether or not there are biodiversity data disparities today related to historic residential segregation. Completeness and expected species richness result from species accumulation curve extrapolations. These measures provide related estimations of unobserved biodiversity, and therefore address the question of how present-day biodiversity data and biodiversity patterns relate to historic housing segregation. The aims are therefore twofold: A) understanding data disparities and bias, and B) spatial variation in urban biodiversity. The HOLC classification system categorized residential neighborhoods in the mid-1930s, meaning un-graded areas were not yet urbanized or were urbanized but non-residential land uses. Focusing only on graded areas excludes most of present-day American cities. By adding the non-graded UAs and MSAs we provide two reference sets to contextualize HOLC neighborhoods in their larger urban contexts. This research thus broadens the taxa under investigation (amphibia, aves, fungi, insecta, mammalia, plantae, and reptilia) and uses a larger and more comprehensive set of species observations across multiple cities than previous related efforts (40), while adding UA and MSA comparisons.

Results

Biodiversity information across HOLC grades, Urban Areas (UA) and Metropolitan Statistical Areas (MSA)

Formerly A-Graded areas had significantly greater sampling density than D-Graded areas for all taxa except fungi ($0.001 > p > 0.0001$, Figure S1). A-Graded areas had greater sampling density than either UA ($p < 0.0001$) or MSAs ($p < 0.0001$) for all nine taxonomic groups.

Completeness estimates from species accumulation curves represent how many species are thought to be present, if exhaustive sampling occurred. Completeness estimates were low and did not vary by HOLC grade for amphibians, fungi (species or family), insects (species or family), mammals, or reptiles ($p > 0.05$). For birds, A had greater completeness than B ($p < 0.05$), C, and D ($p < 0.001$) neighborhoods). Conversely, completeness was greater in D than A-neighborhoods for insects at the species level ($p < 0.01$) and among plants ($p < 0.001$). Completeness was greater in UAs and MSAs excluding previously HOLC-defined neighborhoods than A-Graded areas for all taxonomic groups ($p < 0.0001$). Expected richness did not vary by HOLC grade for taxonomic groups except for birds ($p < 0.001$) and plants ($p < 0.05$). Expected richness was always greatest among MSAs ($p < 0.0001$) and UAs ($0.001 > p > 0.0001$) than for HOLC-Graded areas.

Predictions of biodiversity information, biodiversity knowledge, and species richness across HOLC grades and urban areas

Model predictions show significant differences ($0.01 < p < 0.0001$) in sampling density between formerly A-Graded neighborhoods and formerly D-Graded areas for all nine taxonomic groups (Figure 2, top). The amount of model-predicted biodiversity data varied widely by taxonomic group. For example, amphibian and reptile sampling density, though significantly different across A and D areas, were orders of magnitude lower than bird sampling density regardless of HOLC grade.

Overall average model-predicted estimated completeness in formerly HOLC-defined neighborhoods was 41.7%, and lower for insects (mean estimated completeness = 24.3%), fungi (31.1%), and plants (25.4%) —the most species rich taxonomic groups examined here (Figure 2, middle) across all HOLC grades. Model predictions showed significant differences in estimated completeness by HOLC grades A to D for birds ($p < 0.0001$), mammals ($p < 0.05$), and plants ($p < 0.001$), while the other six taxonomic groups were HOLC-invariant ($p > 0.05$). Birds where the only taxonomic group with significant differences in expected species richness across HOLC grades (Figure 2, bottom, $p< 0.01$).

Discussion and Conclusions

In this study we quantified how the race-based, housing segregation policy called redlining relates to the amount of biodiversity information and the number of expected species across multiple taxonomic groups encompassing nearly every facet of the tree of life. The goals to were to both understand data collection biases and differences in urban biodiversity across varied neighborhoods. Despite prior research on redlining and biodiversity in small geographic regions (40) or taxonomic focus (24, 39), it remained unclear if observed data disparities reflected a general patterns across multiple taxa and cities experiencing a broader range of climates and socioeconomic conditions.

Sampling density was greater in formerly A-Graded neighborhoods than formerly D-Graded neighborhoods for all taxonomic groups examined. Moreover, sampling density is greater in HOLC neighborhoods than their encompassing urban areas and metropolitan regions, while the reverse was true for estimated completeness and expected richness. Few prior investigations have included non-graded comparisons (39), despite calls to do so (30). These patterns are unsurprising given differences in population density across these places, which reduce sampling density among the larger and less population dense spatial units, reflecting the amount of data in areas with higher populations. It remains unclear why people choose to record biodiversity data in formerly A-Graded areas compared to formerly D-Graded areas. One explanation is that there is more green space and tree canopy in A than D-areas (31–34), making these more attractive places to travel to and sample. Alternatively, those observing urban biodiversity already predominantly reside disproportionately in formerly A-Graded areas. The combination of GBIF and HOLC polygons alone does not let us arbitrate between these rival and complementary explanations.

While sampling density differed across HOLC grades for all taxonomic groups, differences in regression-adjusted estimated completeness of biodiversity inventory were only found in birds, mammals, and plants. Differences in expected species richness across HOLC grades was unique to birds. The birdwatching community may promote collecting and sharing data more than for other taxa, and mammal identification is relatively easier. Plants are immobile, very species rich and hard to identify, while there are few urban mammal species. Insect and fugus identification is more challenging, and reptiles and amphibians are relatively more rare, especially in urban areas. These attributes may explain taxon-specific findings. Future studies may consider quantifying species abundances or densities with co-located measurements across taxanomic groups. This may allow for answering questions about whether communities and wildlife food webs vary by race-based policies, as proposed by Schell and colleagues in 2020 (12).

More sampling density in A-grade in all taxa

Our findings that all taxonomic groups had higher sampling density in HOLC-A grade than in D-Graded areas, corroborate the relationships found among birds in prior empirical research (24) and supporting expectations (12). This evidence further suggests how formerly redlined areas have not only fewer environmental amenities today (31–33), greater pollution loads (38), but also less information across nearly every facet of biodiversity. These differences persisted even after accounting for human population density, vegetation productivity, protected and accessible open space, and water cover. Similar findings were observed in four Californian cities across 6 clades, using only iNaturalist data, effectively a subset of GBIF (40). The data disparities found in the larger and more comprehensive GBIF data used here, and across a wider range of taxonomic groups, are reflected within a subset of participatory science platforms, when examining a smaller subset of species in a specific geographic location.

Taxonomic groups differ in data availability and survey completeness

Completeness estimates of biodiversity data varied across taxa. Fungi, insects, and plants had the lowest estimated completeness, yet are the most species-rich taxonomic groups on earth. Of the observations analyzed here, 87.6% were birds, 7.37% plants, 3.16% insects, the remaining ~2% fungus, mammals, reptiles, and amphibians combined. To date, most urban ecology research has focused on birds and vascular plants (42), with invertebrates being among the least studies group (43). In addition, groups such as amphibians and reptiles remain even less-studied, despite being the vertebrate groups facing the highest rates of extinctions in the Anthropocene (44, 45). The taxonomic bias in urban ecology research remains a crucial knowledge gap, as identified by studies calling to include more taxonomic groups (46).Using estimated completeness, we show how the collective information on urban biodiversity differs across taxonomic groups. Specifically, we show higher survey completeness for birds, mammals, amphibians, and reptiles than plants, fungi and insects. Low levels of completeness in plants, fungi and insect likely do not accurately reflect species richness patterns, as these groups are species rich when compared to vertebrates and sampling density was relatively low – it is therefore challenging to disentangle these relationships. More comparative studies across multiple taxa, geographic areas, and over time in urban ecology might be considered a research priority (42).

We did not observe significant differences in expected species richness by HOLC grade in any taxa except for birds (Figure 2). For example, our models predicted similar expected species richness across HOLC grades for birds than for insects and plants, despite there being orders of magnitude more described insect and plant species across the United States than bird species. For example, there are ~1,150 bird species in the USA, but ~91,000 insect species and 16,670 vascular plant species (47–49). Our findings therefore may be reasonably indicative of the low sampling completeness among HOLC grades and the difficulty accurately identifying some species without molecular biology in plants, fungi and insects when compared to birds, mammals, reptiles and amphibians. Low sampling density, especially for species-rich groups, translates into low survey completeness and unrealistically low expected richness, severely limiting ecological inferences about actual community assemblages when using these types of data. Again, more extensive and targeted, local field, possibly with taxonomic experts, sampling may prove pivotal to better understand current urban biodiversity patterns.

Implications

Taken together, our results suggest against extrapolating results of data availability and biodiversity patterns from one taxonomic group to another, particularly when making inferences on invertebrates, plants or fungi based on vertebrate biodiversity patterns. Similarly, our results highlight how findings on sampling density, completeness and richness of birds are not representative of other taxonomic groups in urban environments when using primarily synthesized participatory science data. Biodiversity data from birds in particular may be distinct from other taxa in several ways: a) birds have significantly more observations than other taxa, b) the spatial distribution of their biodiversity records and expected species richness is matched by the rank-order of the HOLC's neighborhood ranking system, and c) birds are a highly mobile taxon. The rise of participatory science campaigns such as eBird and iNaturalist have led to a rapid and steady increase in the collection of such bird biodiversity data across the world, but participation and uptake is primarily by well-educated, white and affluent adults (50, 51). Future work could analyze the demographic profiles relatively small Census geographies like tracts or block groups in association with GBIF data to identify how present-day socioeconomic conditions relate to sampling density and urban biodiversity (22, 23). Concurrently, more research examining the socioeconomic and demographic composition at the individual observer level on who samples may reveal patterns and trends by taxonomic and social groups.

While the increasing use of crowdsourced, geolocated bird data in scientific studies and conservation decisions has led to policy change in urban environments (52), observed trends of bird biodiversity may not necessarily reflect other taxonomic groups of vertebrates, invertebrates and plants. In an era of ambitious global conservation, careful consideration for how data availability across space impacts ecological inference differently across taxonomic groups, and impacts downstream uses is warranted (21). Future work may provide more in-depth exploration into specific facets of biodiversity utilizing other biodiversity data repositories, such as the BIEN database for plant-specific analysis (53). Ultimately, more long term and locally collected field data is likely needed to understand if and how species communities and food webs are impacted by socioeconomic conditions within and across cities. Moreover, how those relationships themselves vary with race-based housing segregation remains less clear.

We are just beginning to understand how past and present practices of segregation and socioeconomic inequality have left (and are leaving) an indelible impact on the environment, urban wildlife communities, food webs, and their evolution (7, 12). Understanding the implications of these human dimensions could be critical for the equitable planning and execution of ambitious conservation and sustainability initiatives from local to national levels. Ecologists increasingly incorporate multiple aspects of human activities into biodiversity studies – from movement, to bi-products such as nightlights, roads and population density and land use change (54). Yet, socioeconomic disparities in biodiversity data are an often overlooked, but critical dimension to consider when leveraging these data for ecological insights or decision making (21). Redlining was just one of many housing segregation practices, similar research could include Urban Renewal project locations (https://dsl.richmond.edu/panorama/renewal/#view=-7726.48/-3679.22/11.13&viz=map&city=baltimoreMD&loc=13/39.2972/-76.5880).

This work provides strong evidence of differences in where we collect information of biodiversity across multiple taxonomic groups across large spatial extents, filling important knowledge gaps in urban ecology and environmental justice research. Future researchers may consider exploring how functional and phylogenetic diversity of these taxonomic groups differs across urban environments, providing a more ecologically-rich context on how species communities vary within and across urban areas. Future researchers may consider including more measurements on where segregationist policies shaped the built and social environments, which in turn effects the ecological contexts for other species.

Materials and Methods

Study Area

We obtained biodiversity information for 195 cities with existing digitized HOLC polygons at the time of our analysis. In order to include non-graded areas as a reference, two Census-defined units were used: urban areas (UA), and Metropolitan Statistical Areas (MSA). Urban areas are the smaller spatial unit among the two, and created by aggregating Census blocks that have 5,000 people or 2,000 housing units. MSA's are aggregations of counties with at least 50,000 people. UA and MSA boundaries were accessed via the `get_acs` function in the tidycensus package (55). Every MSA that contained digitized HOLC polygons that contained with GBIF data (n = 8,207) were included. The result was 145 MSAs, 147 UAs contained within 38 states, within 195 HOLC-defined cities. When calculating the sampling density, completeness, and expected richness, HOLC polygons were erased from their containing UAs and MSAs to avoid double-counting their biodiversity observations.

Biodiversity Data

Biodiversity observations came from the Global Biodiversity Information Facility (GBIF, https://www.gbif.org/), via the `gbif_remote` function in the gbifdb R package (56). GBIF synthesizes disparate sources of biodiversity data from repositories ranging from participatory science apps to museum collections. Observations were filtered to observations containing georeferenced records collected between 2000 and 2020, that were not fossil specimens or material. The total number of observations (n = 58,920,460) per taxon downloaded were amphibia (n = 131,585), aves (n = 51,590,588), fungi (n = 577,360), insecta (n = 1,864,414), mammalia (n = 224,351), plantae (n = 4,342,105), reptilia (n = 190,057). HOLC polygons were obtained from the University of Richmond's Mapping Inequality Project (57) via https://dsl.richmond.edu/panorama/redlining/static/fullDownload.geojson on December 8, 2022.

The three dependent variables analyzed were sampling density, completeness, and expected richness. Sampling density was calculated as the number of observations records per square kilometer. Completeness (%) and expected species richness were calculated using species accumulation curves via the `KnowBPolygon` function in the KnowBR package (58). Completeness represents the percentage of all species estimated to be present given the observed GBIF observations within a spatial unit (HOLC polygon, Urban Area, or Metropolitan Statistical Area). Expected richness was calculated as by extrapolating species accumulation curves (58).

Covariates

In regression analyses, each of the dependent variables were modeled as a function of population density, vegetation cover, protected open space, and water cover. Prior research on birds and HOLC polygons has found significant relationship between human population density, NDVI, and open space with sampling density and percent estimated completeness (24). Additionally, it could be expected that places with more people could be more likely to have participatory science-collected biodiversity data since more potential observers are present. Population counts for HOLC polygons were interpolated using an area-weighted method, where the population was attributed by percent of polygon overlap (59) and year 2019 Census block groups accessed via the `get_acs` function in the tidycensus package (55). Normalized Difference Vegetation Index (NDVI) was computed using the mean of the average monthly MODIS (250m) data from 2015-2019. NDVI captures photosynthetically-active plants, and was included as a vegetation measure. The percent cover of protected open space was included because observers are likely to use parks and open space to collect data. We used a version of USGS' Parks and Protected Areas Database of the United States (PAD-US) that was augmented to included accessible and recreational lands (PAD-US-AR), which is a more accurate and comprehensive representation of open space (60). Spatial water data came from the U.S. Fish and Wildlife Service's National Wetlands Inventory (https://www.fws.gov/program/national-wetlands-inventory/download-state-wetlands-data).

Statistical Analyses

Two sets of statistical analyses were performed: 1) an unadjusted examination of each dependent variable for each taxonomic group by HOLC grade, UA, and MSA categories; and 2) regression analyses excluding UA- and MSA-observations but including continuous covariates. In both cases, sampling density and expected richness were log-transformed to approximate normal distributions. In the first set of analyses, each outcome in the A-Graded polygons was analyzed against the B-, C-, D-Graded, UA's and MSA's values in a series of 5 pair-wise Wilcoxon rank sum tests. Not all possible pairwise tests were performed, rather the endmember was compared against each other value; A serves as a reference and all other values referent. Figures S1-3 show the entire distributions.

Regression analysis incorporated all HOLC polygons, but omitted the UAs and MSAs. This is because UA and MSA represent large geographic areas with high levels of internal heterogeneity, making interpretations difficult. Within an MSA, the mean NDVI does not adequately represent the internal distribution which may have values of zero and one. A mean of 0.5 would not faithfully characterize the region in social or ecological terms. Instead, each of the three dependent variable was analyzed for each of the nine taxonomic groups with three different regression model specifications. The first specification was the outcome as a function of the HOLC grade alone. This linear model is a baseline, simple model. The second specification added a random intercept for unobserved variability associated with each MSA. The third and most complex model adds continuous covariates to the mixed model to control for population density (people per km^2), mean NDVI (a measure of vegetation greenness), protected accessible open space (%, from PAD-US-AR), and water cover (%, from the National Wetlands Inventory). The second and third specifications were fit with the lme4 package (61) in R. Per dependent variable and taxonomic group, the AIC minimization criteria was used to find the best fitting and parsimonious model among the three specifications. Model predictions were then derived with the \`ggpredict\` function and pairwise significance testing was applied using the \`hypothesis_test\` functions in the ggeffects package (62).

Data and code availability

Underlying raw data, the summarized analysis-ready data, and the R scripts for curating, compiling and conducting the final analyses will be freely available on an openly-accessible government data repository upon publication of this manuscript. This combination gives the broadest range of end users the most flexibility.

Acknowledgments

The findings and conclusions in this paper are those of the author(s) and should not be construed to represent any official USDA or US Government determination or policy. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Max Lambert and Morgan Grove for constructive comments on a previous version that helped strengthen the manuscript. D.E.S. acknowledges support from the Yale Institute for Biospheric Studies.

References

1. Li G, et al. (2022) Global impacts of future urban expansion on terrestrial vertebrate diversity. *Nat Commun* 13(1):1–12.
2. Seto KC, Güneralp B, Hutyra LR (2012) Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proc Natl Acad Sci*:1–6.
3. Bratman GN, et al. (2019) Nature and mental health: An ecosystem service perspective. *Sci Adv* 5(7):eaax0903.
4. Cazalis V, Loreau M, Barragan-Jason G (2023) A global synthesis of trends in human experience of nature. *Front Ecol Environ* 21(2):85–93.
5. Simkin RD, Seto KC, McDonald RI, Jetz W (2022) Biodiversity impacts and conservation implications of urban land expansion projected to 2050. *Proc Natl Acad Sci U S A* 119(12):1–10.
6. Hahs AK, et al. (2023) Urbanisation generates multiple trait syndromes for terrestrial

animal taxa worldwide. *Nat Commun* 14(1). doi:10.1038/s41467-023-39746-1.
7. Lambert M, Schell CJ (2023) Cities as the Solution to the Biodiversity Crisis. *Urban Biodiversity and Equity*, eds Lambert M, Schell CJ (Oxford University Press, Oxford), pp 1–21.
8. Gerrish E, Watkins SL (2018) The relationship between urban forests and income: A meta-analysis. *Landsc Urban Plan* 170(April 2017):293–308.
9. Watkins SL, Gerrish E (2018) The relationship between urban forests and race: A meta-analysis. *J Environ Manage* 209(April 2017):152–168.
10. Schwarz K, et al. (2015) Trees Grow on Money: Urban Tree Canopy Cover and Environmental Justice. *PLoS One* 10(4):e0122051.
11. Leong M, Dunn RR, Trautwein MD (2018) Biodiversity and socioeconomics in the city: a review of the luxury effect. *Biol Lett* 14(5):20180082.
12. Schell CJ, et al. (2020) The ecological and evolutionary consequences of systemic racism in urban environments. *Science (80-)* 369(6509):1–19.
13. Magle SB, et al. (2021) Wealth and urbanization shape medium and large terrestrial mammal communities. *Glob Chang Biol* (July):1–14.
14. Chamberlain D, et al. (2020) Wealth, water and wildlife: Landscape aridity intensifies the urban luxury effect. *Glob Ecol Biogeogr* (December 2019):1–11.
15. Des Roches S, et al. (2020) Socio-Eco-Evolutionary Dynamics in Cities. *Evol Appl*. doi:10.1111/eva.13065.
16. Alberti M, et al. (2020) The Complexity of Urban Eco-evolutionary Dynamics. *Bioscience* 70(9):772–793.
17. Szulkin M, Munshi-South J, Charmantier A (2020) *Urban Evolutionary Biology* doi:10.1093/oso/9780198836841.001.0001.
18. Callaghan CT, Ozeroff I, Hitchcock C, Chandler M (2020) Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework. *Biol Conserv* 251(September):108753.
19. Mesaglio T, Callaghan CT, Samonte F, Gorta SBZ, Cornwell WK (2023) Recognition and completeness: two key metrics for judging the utility of citizen science data. *Front Ecol Environ* 21(4):167–174.
20. Hughes AC, et al. (2021) Sampling biases shape our view of the natural world. *Ecography (Cop)* 44(9):1259–1269.
21. Chapman M, et al. (2024) Biodiversity monitoring for a just planetary future. *Science* 383(6678):34–36.
22. Grade AM, Chan NW, Gajbhiye P, Perkins DJ, Warren PS (2022) Evaluating the use of semi-structured crowdsourced data to quantify inequitable access to urban biodiversity: A case study with eBird. *PLoS One* 17(11):e0277223.
23. Perkins D Blind Spots in Citizen Science Data: Implications of Volunteer Biases in eBird Data. Available at: http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf.
24. Ellis-Soto D, Chapman M, Locke DH (2023) Historical redlining is associated with increasing geographical disparities in bird biodiversity sampling in the United States. *Nat Hum Behav* 7(11):1869–1877.
25. Pickett STA, Grove JM, Boone CG, Buckley GL (2023) Resilience of racialized segregation is an ecological factor: Baltimore case study. *Build Cities* 4(1):783–800.
26. Grove JM, et al. (2018) The Legacy Effect: Understanding How Segregation and Environmental Injustice Unfold over Time in Baltimore. *Ann Am Assoc Geogr* 108(2):524–537.
27. Massey D, Denton N (1993) *American Apartheid: Segregation and the Making of the Underclass.* (Harvard University Press, Cambridge, Massachusetts).
28. Aaronson D, Hartley D, Maxumder B (2017) *The Effects of the 1930s HOLC 'Redlining' Maps*.
29. Gee GCGC (2002) A multilevel analysis of the relationship between institutional and individual racial discrimination and health status. *Am J Public Health* 92(4):615–623.
30. Swope CB, Hernández D, Cushing LJ (2022) The Relationship of Historical Redlining with Present-Day Neighborhood Environmental and Health Outcomes: A Scoping Review and Conceptual Model. *J Urban Heal* 99(6):959–983.

31. Nardone A, Rudolph KE, Morello-Frosch R, Casey JA (2021) Redlines and greenspace: The relationship between historical redlining and 2010 greenspace across the United States. *Environ Health Perspect* 129(1):1–9.
32. Nowak DJ, Ellis A, Greenfield EJ (2022) The disparity in tree cover and ecosystem service values among redlining classes in the United States. *Landsc Urban Plan* 221(January):104370.
33. Locke DH, et al. (2021) Residential housing segregation and urban tree canopy in 37 US Cities. *npj Urban Sustain* 1(1). doi:10.1038/s42949-021-00022-0.
34. Namin S, Xu W, Zhou Y, Beyer K (2020) The legacy of the Home Owners' Loan Corporation and the political ecology of urban trees and air pollution in the United States. *Soc Sci Med* 246(xxxx). doi:10.1016/j.socscimed.2019.112758.
35. Hoffman JS, Shandas V, Pendleton N (2020) The effects of historical housing policies on resident exposure to intra-urban heat: A study of 108 US urban areas. *Climate* 8(1):12.
36. Nelson-Olivieri JR, et al. (2024) Inequalities in noise will affect urban wildlife. *Nat Ecol Evol* 8(1):163–174.
37. Burghardt KT, et al. (2023) Current street tree communities reflect race-based housing policy and modern attempts to remedy environmental injustice. *Ecology* 104(2):1–11.
38. Estien CO, Wilkinson CE, Morello-frosch R, Schell CJ (2024) Historical Redlining Is Associated with Disparities in Environmental Quality across California. doi:10.1021/acs.estlett.3c00870.
39. Wood EM, et al. (2023) Historical racial redlining and contemporary patterns of income inequality negatively affect birds, their habitat, and people in Los Angeles, California. *Ornithol Appl* 126:1–29.
40. Estien CO, Fidino M, Wilkinson CE, Morello-Frosch R, Schell CJ (2019) Historical redlining impacts wildlife biodiversity across California. 53(5):323–338.
41. GBIF.Org (2023) Occurrence Download. doi:10.15468/DL.A2XWXV.
42. Aronson MFJ, et al. (2014) A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proceeding R Soc B* 281(1780):20133330.
43. Knapp S, et al. (2021) A Research Agenda for Urban Biodiversity in the Global Extinction Crisis. *Bioscience* 71(3):268–279.
44. Ceballos G, et al. (2015) Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci Adv* 1(5):9–13.
45. Cox N, et al. (2022) A global reptile assessment highlights shared conservation needs of tetrapods. *Nature* 605(7909):285–290.
46. Rega-Brodsky CC, et al. (2022) Urban biodiversity: State of the science and future directions. *Urban Ecosyst* (February). doi:10.1007/s11252-022-01207-w.
47. The Smithsonian Institution (1996) Numbers of Insects (Species and Individuals). *BugInfo*. Available at: https://www.si.edu/spotlight/buginfo/bugnos [Accessed March 4, 2024].
48. eBird (2024) Sightings. *United States, Illus Checkl*.
49. NatureServe (2023) Biodiversity in Focus: United States Edition. 28.
50. Rutter JD, et al. (2021) Racial, ethnic, and social patterns in the recreation specialization of birdwatchers: An analysis of United States eBird registrants. *J Outdoor Recreat Tour* 35(February):100400.
51. Blake C, Rhanor A, Pajic C (2020) The demographics of citizen science participation and its implications for data quality and environmental justice. *Citiz Sci Theory Pract* 5(1):1–10.
52. Farnsworth A, Horton KG, Marra PP (2024) To mitigate bird collisions, enforce the Migratory Bird Treaty Act. *Proc Natl Acad Sci U S A* 121(9):1–4.
53. Maitner BS, et al. (2018) The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. *Methods Ecol Evol* 9(2):373–379.
54. Ellis-Soto D, et al. (2023) A vision for incorporating human mobility in the study of human-wildlife interactions. *Nat Ecol Evol*:13–17.
55. Walker K, Herman M (2021) tidycensus: Load US Census Boundary and Attribute Data as "tidyverse" and 'sf'-Ready Data Frames. Available at: https://cran.r-project.org/package=tidycensus.
56. Boettiger C (2023) gbifdb: High Performance Interface to 'GBIF'_. Available at:

https://cran.r-project.org/package=gbifdb.
57. Nelson RK, et al. (2018) Mapping Inequality. *American Panorama*, eds Nelson RK, Ayers EL Available at: https://dsl.richmond.edu/panorama/redlining/#loc=11/40.1012/-75.1297&opacity=0.14&sort=138&city=philadelphia-pa&text=about [Accessed May 30, 2019].
58. Lobo JM, et al. (2018) KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. *Ecol Indic* 91(April):241–248.
59. Goodchild MF, Anselin L, Deichmann U (1993) A Framework for the Areal Interpolation of Socioeconomic Data. *Environ Plan A* 25(3):383–397.
60. Browning MHEM, et al. (2022) The PAD-US-AR dataset: Measuring accessible and recreational parks in the contiguous United States. *Sci Data* 9(1):1–15.
61. Bates DM (2010) lme4: Mixed-effects modeling with R. 1–145.
62. Lüdecke D (2018) ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *J Open Source Softw* 3(26):772.

Figures and Tables

Figure 1. Spatial extent of 195 cities assessed across the United States. I) Metropolitan Statistical Areas (MSAs; n = 145) included in the study. II) Within MSAs, Urban Areas are smaller, as defined by the US Census Bureau. Home Owners Loan Corporation are within UAs, which are in tern within MSAs, though there are a few instances where small parts of UA's extend beyond an MSA boundary.

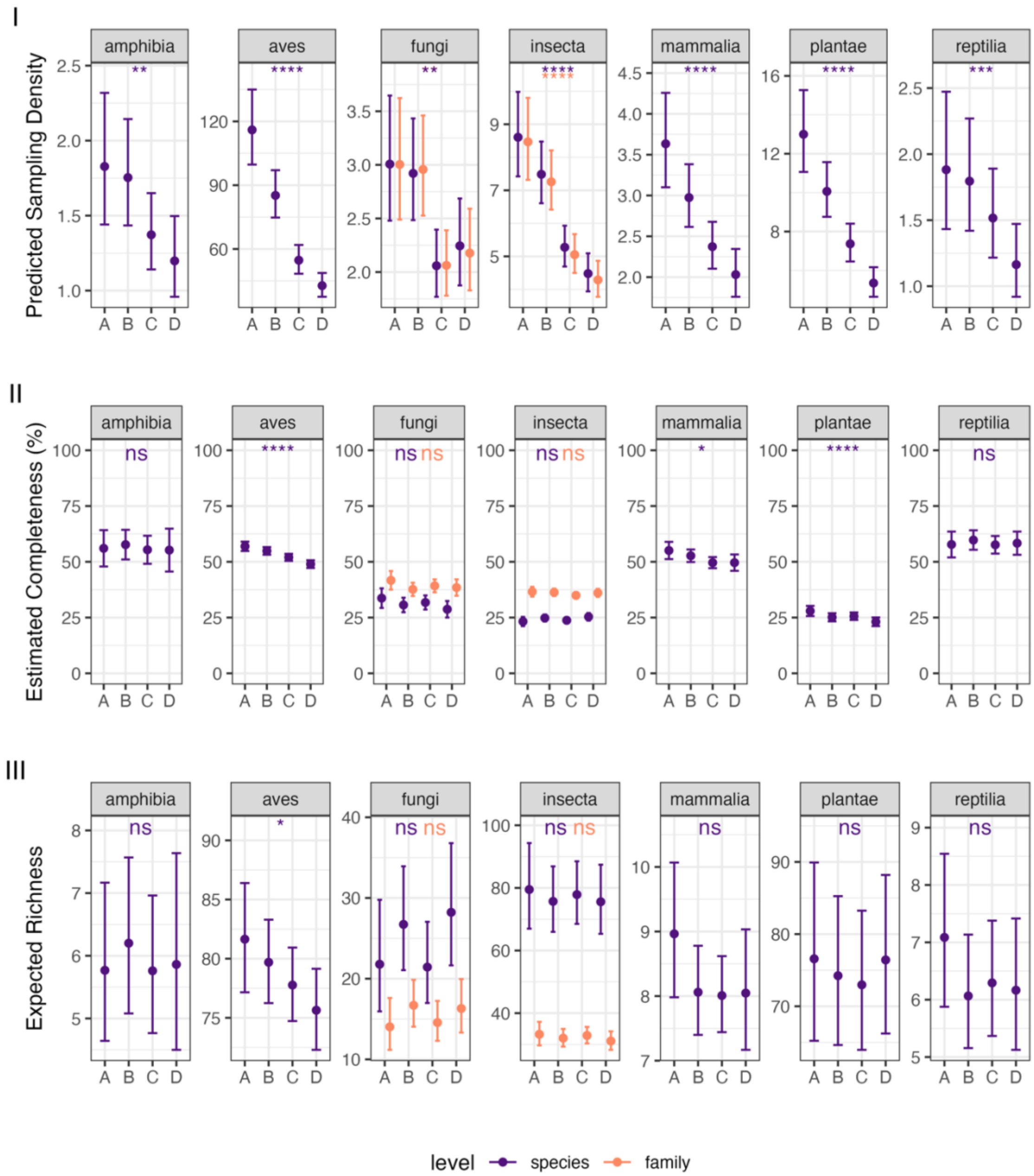

Figure 2. Model-adjusted predicted sampling density varies significantly across HOLC grade for all 9 taxanomic groups (I). Overall estimated completeness is low, and only varies for aves, mammalia, and plantae (II). The observed differences in sampling density and estimated completeness to not translate to differences by HOLC Grade for expected richness except for birds (III). Note the different vertical axes lengths.

SUPPLEMENTAL MATERIALS

A Multi-taxa Analysis of Residential Segregation across the Urban United States

1 USDA Forest Service, Northern Research Station, Baltimore Field Station, Suite 350, 5523 Research Park Drive, Baltimore, MD21228, USA ORCID: 0000-0003-2704-9720

2 National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, Santa Barbara, CA, USA ORCID: 0000-0002-1377-1539

3 Department of Ecology and Evolutionary Biology, Yale University, New Haven, USA ORCID: 0000-0003-4766-021X

Contents:

Table S1. Descriptive Statistics

Characteristic	amphibia: species, N = 503[1]	aves: species, N = 7,717[1]	fungi: family, N = 1,374[1]	fungi: species, N = 1,341[1]	insecta: family, N = 4,473[1]	insecta: species, N = 4,432[1]	mammalia: species, N = 1,943[1]	plantae: species, N = 5,774[1]	reptilia: species, N = 861[1]
Sampling Density (log)	0.55 (-0.16, 1.23)	4.38 (3.16, 5.49)	1.04 (0.31, 1.87)	1.05 (0.31, 1.88)	1.93 (1.12, 2.86)	1.93 (1.12, 2.88)	1.06 (0.36, 1.83)	2.26 (1.40, 3.31)	1.06 (0.26, 1.85)
Estimated Completeness	56 (40, 69)	58 (38, 74)	39 (28, 49)	29 (18, 40)	36 (24, 48)	23 (15, 35)	48 (40, 65)	22 (13, 39)	62 (42, 76)
Unknown	360	1,327	842	910	1,945	2,407	1,054	2,930	460
Expected Richness (log)	1.70 (1.26, 2.12)	4.44 (4.06, 4.81)	2.88 (2.13, 3.49)	3.47 (2.63, 4.19)	3.64 (3.06, 4.09)	4.47 (3.78, 5.07)	2.01 (1.60, 2.53)	4.72 (4.04, 5.32)	1.90 (1.33, 2.26)
Unknown	360	1,327	842	910	1,945	2,407	1,054	2,930	460
HOLC Grade									
A	95 (19%)	929 (12%)	214 (16%)	211 (16%)	555 (12%)	552 (12%)	274 (14%)	726 (13%)	132 (15%)
B	140 (28%)	2,009 (26%)	397 (29%)	383 (29%)	1,208 (27%)	1,198 (27%)	578 (30%)	1,530 (26%)	244 (28%)
C	177 (35%)	3,010 (39%)	524 (38%)	514 (38%)	1,796 (40%)	1,782 (40%)	761 (39%)	2,325 (40%)	305 (35%)
D	91 (18%)	1,769 (23%)	239 (17%)	233 (17%)	914 (20%)	900 (20%)	330 (17%)	1,193 (21%)	180 (21%)
Population / km^2	647 (313, 1,259)	1,324 (659, 2,694)	852 (410, 1,650)	845 (408, 1,636)	1,012 (513, 1,991)	1,009 (513, 1,978)	855 (438, 1,667)	1,137 (578, 2,283)	798 (389, 1,578)
NDVI (mean)	0.43 (0.35, 0.50)	0.41 (0.32, 0.48)	0.40 (0.29, 0.49)	0.40 (0.29, 0.49)	0.40 (0.30, 0.47)	0.39 (0.30, 0.47)	0.38 (0.28, 0.47)	0.40 (0.31, 0.48)	0.37 (0.28, 0.46)
Protected Open, Accessible Space (%)	2.5 (0.9, 7.3)	1.5 (0.1, 4.4)	2.6 (0.9, 6.4)	2.6 (0.9, 6.4)	1.9 (0.5, 5.0)	1.9 (0.5, 5.0)	2.4 (0.8, 5.6)	1.8 (0.4, 4.8)	1.7 (0.4, 4.8)
Water (%)	0.19 (0.00, 0.97)	0.00 (0.00, 0.67)	0.02 (0.00, 0.67)	0.02 (0.00, 0.67)	0.00 (0.00, 0.62)	0.00 (0.00, 0.62)	0.01 (0.00, 0.69)	0.00 (0.00, 0.65)	0.09 (0.00, 0.69)

[1] Median (IQR); n (%)

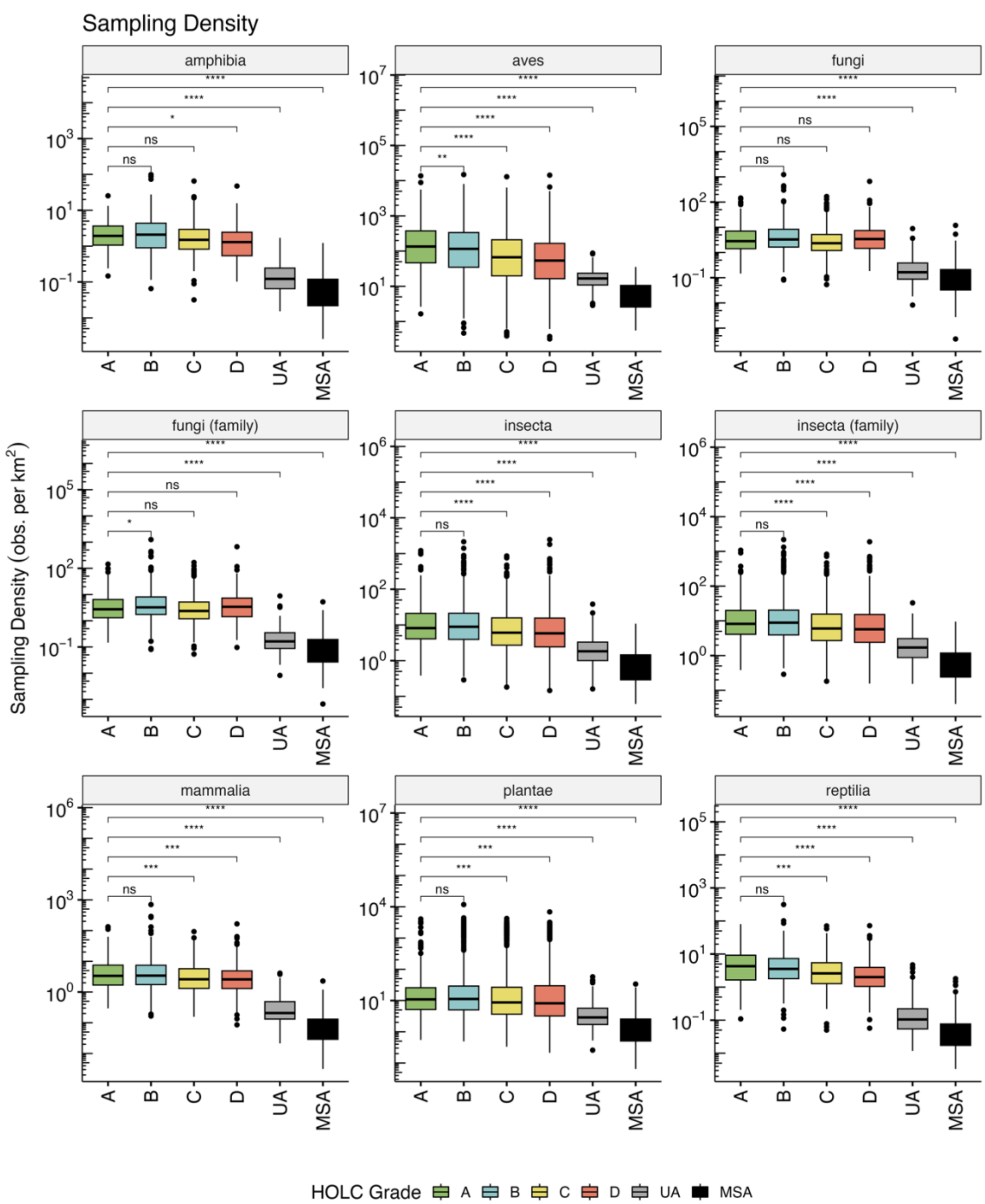

Figure S1. Sampling Density (the number of volunteered-collected observations per area) vary by Home Owners Loan Corporation neighborhoods, with areas formerly A-Graded having more biodiversity information than areas formerly D-Graded for all taxon except for fungi at both species and family levels. Sampling density in HOLC polygons, was greater than their encompassing Census-defined Urban Areas (UA) and Metropolitan Statistical Areas (MSA).

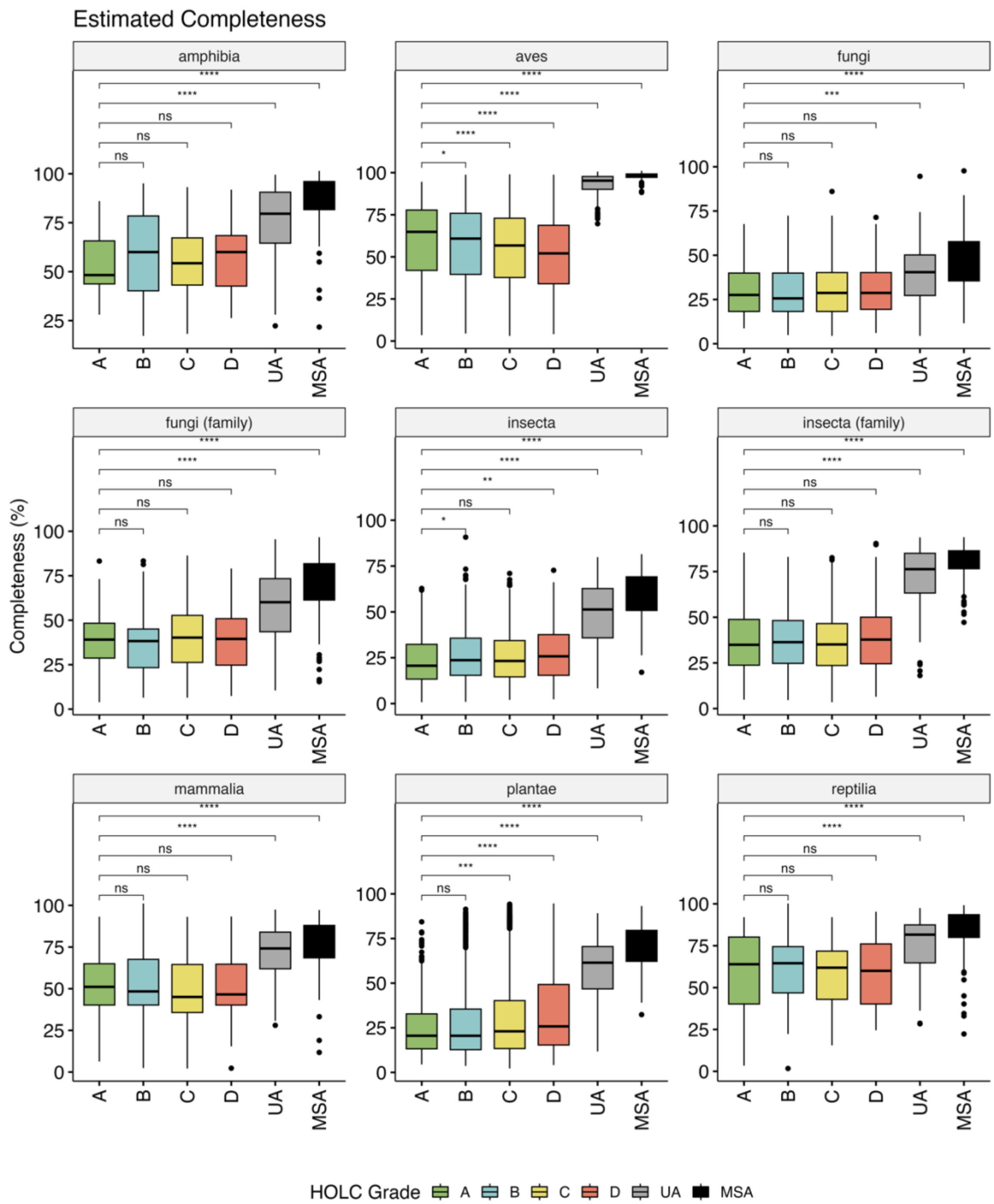

Figure S2. When sampling density is used to estimate percent completeness, few statistically significant differences emerged. A-Graded areas have more complete biodiversity data than D-Graded areas for birds (aves), but the association is reversed for insects (at species and family levels) and for plants. The percent completeness is relatively low overall, and especially for insects, fungus, and plants. Despite fewer observations per area for sampling density, the percent completeness is greater in Urban Areas and Metropolitan Statistical Areas than HOLC polygons, owing to their larger size.

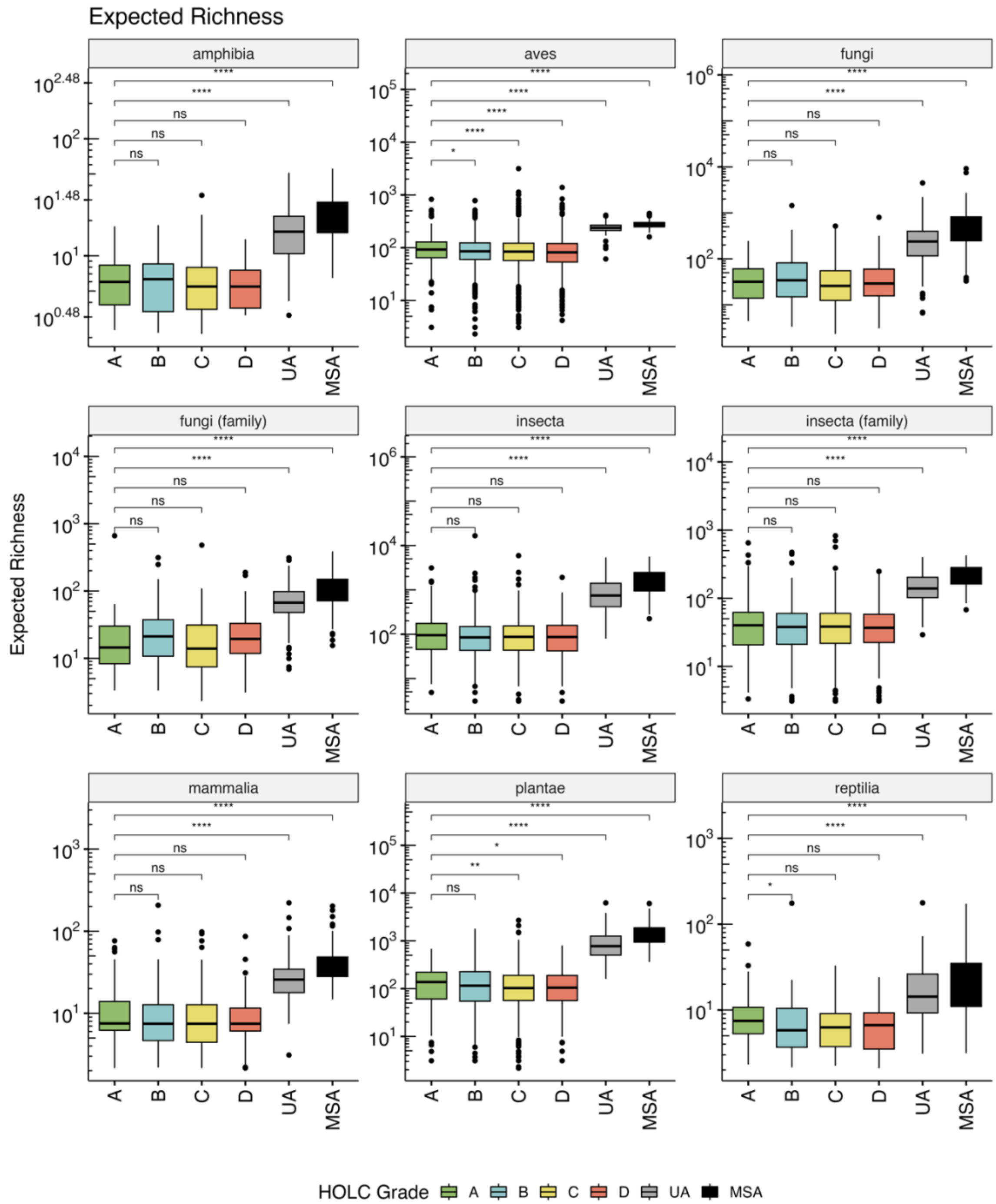

Figure S3. Only Aves and Plantae expected richness vary by HOLC grade, the other taxon are invariant to the neighborhood classification system.

```
# end April 22, 2024
```