INTRODUCTION

Natural history collections (NHCs) are an irreplaceable record of global biodiversity because they contain data spanning decades to millennia (Ewers-Saucedo et al., 2021; Kharouba et al., 2019; Pyke & Ehrlich, 2010). NHCs have advanced our understanding of long-term changes in phenology (e.g. Brooks et al., 2014), morphology (MacLean et al., 2019), species interactions (Meineke et al., 2018; Miller-Struttmann et al., 2015), disease dynamics (e.g. Zeilinger et al., 2017) and trait evolution (Holmes et al., 2016). Recent increases in the digital accessibility of these data have generated new interest in the possibility of using NHCs to estimate both species' distributions in historical time periods as well as trends in either distributions or abundances, particularly over longer time horizons than what is currently possible with contemporary survey data (e.g. Bartomeus et al., 2013; Boyle et al., 2019; Cameron et al., 2011;
This interest in using historical NHC data coincides with a pressing concern about annual variations in insect abundances or the factors influencing their population trends (Cardoso & Leather, 2019). Insects comprise over half of all described species on Earth but are vastly understudied, primarily because they are small in size, commonly considered as pests, and undervalued by the public compared to larger, charismatic species. As a result of these data limitations, the true incidence or rate of decline for the vast majority of insect species is unknown (Cardoso & Leather, 2019). For many insects, NHCs are the only long-term datasets available. However, like all NHC data, historical insect specimens were often collected with a specific purpose in mind (e.g. taxonomic inventory) and, as a result, exhibit spatiotemporal biases in collection effort when collated across research collections (Boakes et al., 2010). Ultimately, these biases limit the use of NHC data to estimate long-term population trends. Even for a well-loved, high-profile species such as the monarch butterfly Danaus plexippus, which contains thousands of recorded specimens across the continental U.S. over more than a century, it is currently impossible to estimate trends in abundance using data only from digitized museum records (Ries et al., 2019; Wepprich, 2019).

The aim of this review is to provide an overview of the challenges associated with using data from NHCs to track long-term population trends (i.e. changes in distribution or abundance through time, see Glossary in Box 1), highlight recent methodological advancements that aim to overcome these challenges, and discuss emerging research opportunities. Specifically, we examine the potential of integrating digitized museum records and other contemporary data sources...
(e.g. collected via structured, designed surveys and opportunistic citizen science programs) in a unified analytical framework that explicitly accounts for the unique sampling biases associated with each data source. The emerging field of integrated modelling provides a promising framework for leveraging the wealth of NHC data to accurately estimate long-term trends of insect populations but requires further development and assessment. We conclude with a discussion on future research that would further increase the value of NHCs for understanding the impacts of global change on insect populations.

2 CHALLENGES ASSOCIATED WITH ESTIMATING POPULATION TRENDS FROM NHC DATA

Natural history specimens were originally collected for a wide variety of purposes, and so can exhibit spatial, temporal, and taxonomic biases in collection effort when collated across research collections (Boakes et al., 2010; Meineke & Daru, 2021). Differences in collection methodology as well as non-random and non-representative sampling can make it difficult to extract signals of biological change from sampling noise (Isaac et al., 2014). For example, specimen records may arise from the personal interests of collectors (e.g. affinities for certain species) and be spatially biased toward locations that are easily accessible (e.g. roadsides, cities), contain high species diversity, or have a disproportionate number of rare species (e.g. protected areas; Girardello et al., 2019; Haque et al., 2020; Shirey et al., 2021). Spatial bias can result in environmental bias when records represent a nonrandom sampling of habitats across a species’ range or study region, which may lead to spurious patterns in estimated covariate relationships (Graham et al., 2004; Williams et al., 2002). There are also historical sampling biases in the ecosystems surveyed globally, with very few insect specimens collected in desert, boreal, and tropical forest biomes (Shirey et al., 2021). These biases, combined with the sparse spatial and temporal coverage even in well-sampled areas, can make robust comparisons of population trends across multiple ecosystems, regions, and species difficult. Seasonal biases in collection effort corresponding to the times of year when a particular life stage is most prevalent (e.g. breeding season aggregations) or during favourable sampling conditions (e.g. summer months) can further hinder our ability to understand broad spatial patterns in trends and environmental relationships across a species’ range (e.g. Davis et al., 2015).

Digitized specimen records are also considered a ‘presence-only’ data source because they do not provide information on where a species was observed but not collected (Graham et al., 2004). Presence-only data are the most abundantly available data type and frequently the only data available for many species worldwide. As such, there is growing interest in using such data to model changes in species’ distributions and abundances but doing so requires assumptions about how the data were collected (Dorazio, 2012; Pearce & Boyce, 2006). A key limitation of modelling presence-only data is that it fails to account for the effects of imperfect detection (Dorazio, 2012; Yoccoz et al., 2001). A species’ non-detection can represent a true absence, a false negative observation (i.e. species was present but not collected), or a lack of sampling effort (i.e. the site was not sampled). Species that are elusive, occur at low densities, or exhibit cryptic coloration can be difficult to detect and may therefore be under-represented in collections relative to their true abundances, whereas species that are rare or highly prized by collectors may be over-represented (see Gotelli et al., 2021; Guralnick & Van Cleve, 2005). Without additional information on where a species was not observed or observed but not collected, as well as the spatial areas targeted for collection, it is hard to disentangle these potential effects (Löw et al., 2007). For this reason, some biologically relevant parameters (e.g. baseline abundance) cannot be reliably estimated with just presence-only data (Dorazio, 2014; Farr et al., 2020).

Temporal changes in collection effort and spatial coverage present an additional challenge and may obscure real patterns of population change, particularly if changes in effort and coverage are not well documented (Graham et al., 2004; Stockwell & Peterson, 2002). Furthermore, improved collection methods and knowledge of species’ biology over time may result in apparent positive population trends that are simply an artefact of improved detection and collection of species. Collectors may also seek out species according to their rarity (e.g. species of conservation concern) or societal importance (e.g. agricultural pests), which also varies through time (Jepsson et al., 2010). Conversely, species may be collection-restricted in parts of their range and at different points in time based on population status (e.g. red-listed species). These changes in collecting behaviour can then be confounded with real population change resulting in misleading inferences (e.g. inferring population gain when, in fact, a species is just more likely to be collected due to changing affinities).

Finally, identification errors can arise due to species misidentification or incomplete taxonomic knowledge (e.g. cryptic species) either at the time of collection or during the accessioning and publication of species records (Goodwin et al., 2015; Graham et al., 2004). When possible, digitized records or physical specimens should be re-examined to verify species identification. In practice, this is challenging to do comprehensively, and is still not error-free (e.g. in the case of cryptic species). Frequent updates to taxonomy can also result in species synonyms when records are aggregated across different institutions and time periods. Spatial error and uncertainty pose additional challenges. Specimens are often associated with a written description or geospatial coordinates of the sampled location, but the amount and quality of locality information varies widely between specimens and collections, influencing the accuracy of georeferencing efforts (i.e. associating a specimen with a particular latitude and longitude). Older records are more difficult to accurately georeference as place names, and distinctive landmarks or habitat features may change through time, and these records are less likely to already be associated with geospatial coordinates (Murphey et al., 2004). Specimens with incomplete locality information are either georeferenced
to the centroid of larger geo-political units (e.g. counties; Collins et al., 2017; Erickson & Smith, 2021; Park & Davis, 2017; Pender et al., 2019), or should be removed from the aggregated dataset prior to analysis. Potential inaccuracies and imprecision in survey route or collection location are also problematic for insects because many species are ecologically specialized with small home ranges, so small-scale differences in the environmental conditions can influence local distributions and abundances (Harrison et al., 2015; Nice et al., 2019).

3 | METHODS TO OVERCOME THESE CHALLENGES

Methods that are commonly used to overcome some of these challenges can be split into two categories: data-level and model-level approaches. Data-level approaches include a variety of data filtering and correction techniques that aim to reduce spatiotemporal biases in the collection process (Isaac et al., 2014). For example, systematic subsampling (i.e. thinning) of data in geographic or environmental space can reduce the spatial aggregation of records in highly sampled locations (e.g. cities). Filtering data to the core spatial and temporal extent of species’ ranges can further reduce collection biases that occur at range edges where species are less abundant and sporadically distributed (Ries et al., 2019). However, if the goal is to look at trends (e.g. in distribution), then it may be important to include observations along established or hypothesized range boundaries where a signal of population change may be easier to detect. Data filtering will almost always be a necessity, but how it is done should depend on both the specifics of the data and the question at hand. To address challenges associated with presence-only data, auxiliary information can be used to infer pseudo-absences at locations where the target species was not collected. The selection of pseudo-absence points can be inferred from occurrence data on other, non-target but closely related species, or using additional information on sampling effort (Van Der Wal et al., 2009). Data describing the collection process, such as the length of time a collector was sampling or the distance a collector travelled, are ideal, but rarely accompany digitized specimens. Data on other organisms caught at the same time and place using the same collection methods as the target species (i.e. background-group taxa) are assumed to share the same bias and therefore serve as a proxy for sampling effort (Ponder et al., 2001). Data filtering and correction techniques can be an effective tool for removing some biases and improving overall data quality but come at the cost of reduced sample sizes (Van Eupen et al., 2021) and may not account for all biases when estimating species’ trends, particularly because processes for collecting data may change through time and therefore be confounded with signals of biological change if not accounted for (Isaac et al., 2014).

Model-level approaches are arguably more robust than data-level approaches because they explicitly account for collection biases and changes in sampling effort over time via random effects structures, covariates, or an explicit observation process model component (e.g. Erickson & Smith, 2021). Random effects (e.g. for site and/or year) and sampling covariates (e.g. time of year) can help account for variation in sampling effort across space and time and generally do not require throwing out data (Meineke & Daru, 2021). The inclusion of covariates such as collector and list-length (i.e. the number of other species collected at the same location and time by the same collector) can also help account for collector-specific biases in survey-site selection (Roberts et al., 2007; van Strien et al., 2010). Recent approaches have used information on sampling effort to construct collection histories for individual collectors, allowing for explicit estimates of imperfect detection when modelling species’ occurrence rates (Erickson & Smith, 2021). Occupancy models are commonly used to estimate species occurrence and detection probabilities using systematic surveys (MacKenzie et al., 2002, 2018; Royle & Dorazio, 2008), and semi-structured citizen science observations (e.g. Altwegg & Nichols, 2019; Johnston et al., 2021). However, estimating the probability of detection requires independent, repeat visits to a single site in a short time frame, which may not be possible with NHC data in scarcely sampled regions or time periods. Combining data- and model-level approaches has shown to improve accuracy for modelling species’ distributions using NHC data (e.g. Erickson & Smith, 2021), but has yet to be explored for estimating population trends and still does not help overcome challenges associated with large data gaps across space and time.

4 | DATA INTEGRATION AS AN EMERGING RESEARCH OPPORTUNITY

One possible solution to estimate insect population trends using NHC data is to combine specimen records with other types of observational data. The emerging field of integrated modelling (reviewed in Isaac et al., 2020) can provide a promising framework for combining multiple datasets to estimate species distribution patterns and covariate effects on occurrence or abundance (Fletcher et al., 2019). Data integration methods typically use a joint likelihood approach, where each dataset provides additional or complimentary information on the biological process (e.g. occurrence and/or abundance) through observation sub-models that account for imperfect detection and other sampling biases (Box 2: Kéry & Royle, 2015). The integration of multiple data sources and different data types (e.g. count and presence-only data) can provide several inferential benefits (Zipkin et al., 2021), including increased precision of parameter estimates (Miller et al., 2019) and improved identifiability of parameters (Dorazio, 2014). Data integration methods can also account for uncertainties and correlations among datasets (Pacifici et al., 2017). Despite recent modelling advances, an evaluation of the advantages and unique challenges associated with the integration of museum records has yet to be undertaken.

We examine the potential advantages and methodological challenges associated with the integration of historical museum
BOX 2 An overview of integrated modelling

Background—Integrated modelling provides a framework for combining multiple datasets in a single statistical model to estimate species distribution (or abundance) patterns and covariate effects on occurrence (or abundance, including on occasion demographic rates). Data integration methods build on recent advances in hierarchical modelling that separate the biological and observation processes via sub-models that are linked together via a joint likelihood (Kéry & Royle, 2015). Integrated models are commonly fit using a Bayesian approach, which allows for maximum flexibility in the overall model structure and the propagation of uncertainty for all population-level parameters.

Basic Model Structure—We illustrate integrated modelling using a hierarchical framework (left panel) to make inferences about an underlying biological process (abundance at location \( j \), \( N_j \)) that is imperfectly observed in the presence-only dataset \( y \), and structured count dataset \( z \). Each dataset provides additional or complimentary information on the biological process (yellow) through observation sub-models that account for imperfect detection and other sampling biases (grey). The equations (right panels) provide an example using a spatial point process approach (Miller et al., 2019) to model abundance, where \( \lambda \) is the mean expected abundance, \( \sigma \) is the observation error for the presence-only data, and \( p \) is the detection probability for the structured count data. Link functions can also be used to incorporate relevant covariates that influence species dynamics and detection across space and time. Note that this framework can be easily extended to estimate changes in abundance through time (rather than a single time step, as shown here) and accommodate other data sources via additional observation sub-models.

Benefits of data integration—The benefit of data integration lies in the estimation of parameters that are shared across sub-models. By using all available information, population parameters (e.g. abundance) can be estimated more precisely than if estimated from each data source independently (Zipkin et al., 2021). Integrated modelling also provides a robust framework by which to capitalize on the individual strengths of different data types while minimizing their weaknesses. For example, presence-only data are easier to collect across broad spatial and temporal extents than standardized data but information-poor relative to structured counts. Integrating presence-only and structured survey data can expand the spatiotemporal scope of inference while allowing for the robust estimation of parameters that cannot be estimated using presence-only data (Dorazio, 2014; Farr et al., 2020). [Correction added on 4-July-2022, after first online publication: The layout of Box 2 has been corrected.]

records and contemporary data sources using the monarch butterfly Danaus plexippus as an example species. Monarchs are one of the most recognizable and well-studied butterflies in North America and are experiencing steep declines throughout their overwintering and breeding ranges (Agrawal & Inamine, 2018; Brower et al., 2012; Schultz et al., 2017). Long-term changes in climate, land-use, and agricultural practices (e.g. pesticide application practice) are all contributing factors to observed declines (Pleasants & Oberhauser, 2013; Saunders et al., 2018; Zipkin et al., 2012), but survey data limitations have restricted retrospective analyses of population trends to only the past 25 years (e.g. Zylstra et al., 2021). Boyle et al. (2019) recently attempted to use historical NHC data to estimate monarch population trends over the last century but failed to appropriately account for spatiotemporal collection biases (Ries et al., 2019; Wepprich, 2019). After correcting for these biases, Ries et al. (2019) showed that it is not currently possible to
estimate long-term abundance trends using only digitized museum records because of data limitations, with a median number of a single specimen collected per year over the 20th century. Data integration methods would enable us to leverage all available data from both historical and contemporary time periods, which may allow for trend estimation over longer time horizons and reveal important relationships with recent land-use and/or climate changes.

Data are available for the eastern population of adult monarch butterflies from a variety of sources from 1900 to 2018. For the primary breeding grounds in the Midwestern United States, these data include historical museum records of 1940 specimens aggregated across 52 NHCS in North America, presence-only opportunistic observations from iNaturalist, and counts from five structured monitoring programs, including Pollard data from four state monitoring networks in Illinois, Iowa, Michigan, and Ohio, and population count data from the North American Butterfly Association (NABA). Additional information about each of the datasets can be found in Supporting Information S1. To examine the potential advantages and challenges associated with the use of data integration methods to estimate long-term population trends, we aggregated all these available data over the last ~120 years from 1900 to 2018. We then filtered all datasets to the core spatial (39°N to 49°N latitude and 80.5°W to 96°W longitude) and temporal (June to August) extent of the summer Midwestern breeding population.

As expected, NHCS provide the only data available for monarch populations in historical time periods (i.e. before 1990; Figure 1) but are spatiotemporally biased. Approximately 50% of all monarch specimens across the Midwestern U.S. were collected in just three time periods (1930–1931, 1977–1982, and 1995–1999). Contemporary data sources significantly increase spatial data coverage and are the only reliable source of information on monarchs in more recent time periods, as the number of digitized museum records has decreased dramatically from 2001 to 2018 (Figure 2). Opportunistic iNaturalist data serve to both increase sample size and fill in large spatial data gaps (e.g. in Minnesota and Indiana) from 2010 to 2018, but are primarily restricted in temporal coverage to the last 10 years (Figure 1). Structured survey data provide the most consistent and standardized perspective on contemporary populations from 1991 to 2018. These surveys are still limited, however, with large gaps in spatial coverage in certain areas of the Midwest that have prevented the estimation of relative covariate effects prior to 2004 (see Żylstra et al., 2021). In theory, data integration methods should be able to combine the various strengths of these independent data sources to expand the spatiotemporal scope of possible analysis. In practice, however, there are several potential challenges to estimating long-term population trends using these data and methods, including: (a) accounting for the appropriate sampling biases and observation errors in each data source; (b) differences in the information content and data quantity among integrated datasets; (c) spatial misalignment of datasets or of observational data and environmental covariates; and (d) temporal mismatch and extent of temporal overlap among available datasets.

The development of modelling approaches that integrate different data types while accounting for sampling biases is an ongoing challenge (Fletcher et al., 2019; Isaac et al., 2020; Miller et al., 2019). Part of the challenge comes from having to identify and then account for each dataset’s unique sampling biases and observation errors. These processes are rarely known with certainty and often require a thorough understanding of how and for what purpose the data were collected. This is particularly difficult for opportunistic, presence-only datasets, like those derived from NHCS or iNaturalist, that do not have a defined sampling scheme and often have limited metadata describing the sampling process. Research-grade observations from iNaturalist exhibit similar sampling biases to those previously discussed for museum collections, including spatial biases toward easily accessed locations (e.g. urban centers; Di Cecco et al., 2021) and taxonomic biases toward species that are more easily detected (e.g. of larger body size, locally abundant; Callaghan et al., 2021). So, while integrating museum records with contemporary presence-only datasets can greatly expand the scope of study, it may also add more noise and complicate inference. Careful considerations must be made when choosing which datasets to integrate, particularly if there is difficulty or uncertainty in how best to account for potential sampling biases and observation errors.

Independent datasets also differ in information content and quantity of data. Well-designed structured surveys provide a rich source of high-quality data on target species, but are usually only collected over a small, pre-defined study extent because of financial and logistical constraints (Dornelas et al., 2018; Lawton et al., 1998). Opportunistic data are information-poor because they only provide data on where species were observed and recorded (as reviewed above in the context of NHC data) but are often easy and cheap to collect over broad spatial and temporal extents (Kosmala et al., 2016). In the context of data integration, uneven quantities of data among different data sources can result in one data source overwhelming another (Pacifici et al., 2017), with parameter estimates biased toward the most abundant data source regardless of its reliability or information content (Zipkin et al., 2021). One potential solution is to differentially weigh datasets according to either sample size, measures of variability or reliability (Fletcher et al., 2019; Francis, 2011; Punt, 2017). However, the validity of this solution has not been explored when the relative quantities of data change through time. For example, the number of independent locations where monarchs were recorded via iNaturalist clearly outnumber all other data sources from 2010 to 2018, but this is not the case in previous time periods (Figure 2). Capturing these types of temporal dynamics in relative data quantities will ensure that the best possible inferences are made using all available data sources.

A third challenge is the integration of data collected at different spatial scales. Misalignment can occur among the different datasets and/or between observed data and environmental covariates. Mismatches in spatial scale can result from uncertainty in the precise location at which a species was collected or observed. The precision of specimen georeferencing efforts can vary, meaning that some museum records are linked to a very precise location and others are only referenced to large geographic features such as cities or counties (Murphey et al., 2004). Similarly, research-grade iNaturalist observations include a georeferenced location that is either provided automatically via GPS-enabled devices (e.g. cellphones)
or entered manually by observers. While observers can provide a measure of location accuracy, this is not a common practice. Because of uncertainties in where a species was collected or observed, these presence-only data sources are often modelled at larger spatial units (e.g. at a county-level) or spatial grains (e.g. 100-km resolution), whereas environmental covariate data are often gridded at a higher

FIGURE 1 Spatiotemporal coverage of preserved monarch specimens (left column, black), presence-only observations from iNaturalist (center column, red), and structured count data from Pollard and NABA survey programs (right column, blue) in four time periods from 1900–2018. NHC and iNaturalist observations show the number of locations where at least 1 adult monarch was collected/observed. For the structured count data, the number of observations is the number of surveys with >0 monarch counts, and the number of surveys represents the total number of surveys conducted within the particular time period. Bubble size represents the mean relative count (i.e. number of adult monarchs counted/number of survey hours) at each location.
spatial resolution (e.g., ~1-km resolution for WordIClim climate surfaces; Fick & Hijmans, 2017). If unaccounted for, these mismatches in spatial scale can result in biased inferences or overstated precision of model parameters (Gotway & Young, 2002). ‘Change of support’ techniques can be used to address various kinds of spatial misalignment by up- or down-scaling datasets accordingly to achieve a single, shared spatial extent and resolution (Cressie, 1996; Gotway & Young, 2002; Pacifi et al., 2019).

Temporal mismatch among datasets has been far less studied and is still an active and important area of research (Fletcher et al., 2019). Current data integration methods are not yet capable of combining data collected on very different temporal scales. To some degree, the utility of data integration will depend on the relative length of available time-series (Figure 2). For example, coupling decades of museum data with a single year of structured data is not likely to provide inferential benefits and may require additional assumptions about environments and covariate effects being constant over time. It is also unclear to what extent the amount of temporal overlap between datasets may influence model performance and resultant inferences, or how minimum requirements of temporal overlap may change based on the biology of the species or taxon of interest. In the Midwestern United States, there is a ~30-year period of overlap between historical museum records and contemporary survey data, and a 10-year period of overlap where all three data types are present though new museum records are particularly scarce from 2010 to 2018 when iNaturalist observations are most abundant (Figure 2). These periods of temporal overlap between historical and contemporary data sources may provide an anchor point from which to predict relative abundances, with associated uncertainties, back in time when only museum records are available. Estimating insect trajectories may require longer periods of temporal overlap between data sources than what is necessary for other taxa because of the highly dynamic nature of many insect populations (Fox et al., 2019; Hanski, 1990; Williams, 1961).

5 | FINAL REMARKS

The digitization of museum specimens provides new and exciting opportunities for understanding species’ trends over decades to centuries but incorporating NHC data into rigorous population analyses is an ongoing challenge. Spatiotemporal and collector-specific biases have limited the use of museum records for understanding temporal changes in species distributions and abundances. Yet, the untapped wealth of NHC data may be the only information available to evaluate long-term impacts of global changes on several species, including most insects (Kharouba et al., 2019). Data integration methods (Fletcher et al., 2019; Isaac et al., 2020) offer one potential framework for advancing the utility of NHC data for inferences on long-term population trends. However, the field of integrated modelling is still relatively new, and current methods are not yet capable of integrating datasets collected over different time periods without several strong assumptions. Furthermore, a thorough assessment of the conditions under which it is useful and reasonable to integrate NHC and other observational data is needed. For example, data integration may not be useful for taxa or regions with limited structured survey data. A formalized integration framework, however, could be used to identify cases where integration is not possible using existing data sources, and therefore provide information on where to best target contemporary population surveys or prioritize digitization efforts to fill in historical data gaps. Future work should first target the key methodological questions outlined in Box 3, but we also note that many more challenges and opportunities will likely arise as new applications of data integration using NHC data are explored.

One possible alternative to the population-level data integration methods we describe is the use of filtered and survey-calibrated NHC data to document community-level changes in relative abundances and species rank at different points in time using the FAMA (field abundance-museum abundance) approach recently described in Gotelli et al., 2021. This would require having survey and NHC data that were collected independently for multiple species at
roughly the same time and locations but could be a powerful tool for the taxa where such conditions are met. For most insect groups and geographic regions, however, there are not enough historical data currently available to disentangle low rates of annual decline from natural generation-to-generation changes in relative abundances. With only 10% of worldwide NHC data digitized (Ariño, 2010; Page et al., 2015), it is difficult to know if or when it may be possible to do so; the complete digitization of estimated NHC holdings is still projected to take several decades (Ariño, 2018). As such, the development of methodological tools that can effectively make use of all types of currently available population data - including specimen records - is an important and timely avenue for future ecological and collections-based research.

AUTHORS’ CONTRIBUTIONS

C.L.D., R.P.G. and E.F.Z. conceived the ideas; C.L.D. retrieved the monarch butterfly data from available sources, created the figures, and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank the countless museum staff and volunteer scientists for their time and expertise digitizing and documenting butterfly records. This work was supported by an NSF Postdoctoral Research Fellowship in Biological Collections (DBI-2010698) to C.L.D. and an NSF Infrastructure Innovations in Biological Research award (DBI-1954406) to E.F.Z.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

No original data were presented in this article.

ORCID

Courtney L. Davis  https://orcid.org/0000-0002-6467-4288
Robert P. Guralnick  https://orcid.org/0000-0001-6682-1504
Elise F. Zipkin  https://orcid.org/0000-0003-4155-6139


**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.