DATA PAPER

Freshwater insects CONUS: A database of freshwater insect occurrences and traits for the contiguous United States

Laura Twardochleb1,2 | Ethan Hiltner1 | Matthew Pyne3 | Phoebe Zarnetske2,4

1Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA
2Ecology, Evolutionary Biology and Behavior Program, Michigan State University, East Lansing, MI, USA
3Department of Biology, Lamar University, Beaumont, TX, USA
4Department of Integrative Biology, Michigan State University, East Lansing, MI, USA

Correspondence
Laura Twardochleb, California Department of Water Resources, 3500 Industrial Boulevard, West Sacramento, CA 95799, USA.
Email: laura.twardochleb@water.ca.gov

Funding information
National Aeronautics and Space Administration, Grant/Award Number: 80NSSC17K0395

Editor: Andres Baselga

Abstract
Motivation: Freshwater insects comprise 60% of freshwater animal diversity; they are widely used to assess water quality, and they provide prey for numerous freshwater and terrestrial taxa. Our knowledge of the distribution of freshwater insect diversity in the USA is incomplete because we lack comprehensive, standardized data on their distributions and functional traits at the scale of the contiguous United States (CONUS). We fill this knowledge gap by presenting Freshwater insects CONUS: A database of freshwater insect occurrences and traits for the contiguous United States. This database includes 2.05 million occurrence records for 932 genera in the major freshwater insect orders, at 51,044 stream locations sampled between 2001 and 2018 by federal and state biological monitoring programmes. Compared with existing open-access databases, we tripled the number of occurrence records and locations and added records for 118 genera. We also present life-history, dispersal, morphological and ecological traits and trait affinities (analogous to fuzzy-coded traits) for 1,007 stream insect genera, assembled from existing databases, reference books and the primary literature. We nearly doubled the number of traits for 11 trait groups and added traits for 180 genera that were not available from open-access databases. Our database, Freshwater insects CONUS, facilitates the mapping of freshwater insect taxonomic and functional diversity and, when paired with environmental data, will provide a powerful resource for quantifying how the environment shapes stream insect diversity and taxon-specific distributions.

Main types of variables contained: Georeferenced occurrence records and traits for stream insects.

Spatial location and grain: Contiguous United States at a grain of c. 1 m².

Time period and grain: Occurrence records from January 2001 to December 2018, with 1-day temporal resolution. Traits from January 1911 to December 2018.

Major taxa and level of measurement: Genera from the orders Coleoptera, Diptera, Ephemeroptera, Hemiptera, Lepidoptera, Megaloptera, Neuroptera, Odonata, Plecoptera and Trichoptera.

Software format: .csv.
Understanding the distribution of biological diversity at continental scales is a key goal of biogeography, community ecology and conservation research (Pereira et al., 2013; Ricklefs et al., 1993; Wiens & Donoghue, 2004). Species occurrence records and functional traits are needed to quantify and map taxonomic and functional diversity for monitoring and assessing environmental influences on populations, ecological communities and ecosystem functioning (Jetz et al., 2019; Pereira et al., 2013). Taxon-specific distribution data are also essential for predicting geographical ranges and species responses to global change, which are important facets of conservation planning (Rodríguez et al., 2007; Serra-Díaz & Franklin, 2019). Ecologists have made progress towards assembling taxonomic occurrence and trait datasets that enable the mapping of broad-scale biodiversity patterns of terrestrial organisms (e.g., Belmaker & Jetz, 2011; Butler et al., 2017), marine organisms (Grady et al., 2019) and freshwater fish (Comte & Olden, 2017). Despite this progress, open-access biodiversity datasets for freshwater insects, such as the U.S. Environmental Protection Agency (USEPA) Freshwater Biological Traits database (USEPA database; Poff et al., 2006; U.S. EPA, 2012; Vieira et al., 2006) and the Water Quality Portal (WQP; https://www.waterqualitydata.us), are not easily combined for biodiversity mapping, because they contain outdated taxonomic names and trait terminology and have gaps in trait assignment for many taxa. In addition, they do not provide fuzzy traits commonly used by researchers in Europe and other regions that would facilitate cross-continental comparisons and assembly of global trait databases (Schmera et al., 2015). Below, we briefly describe the history, uses and limitations of existing databases and the need for integrated, comprehensive and standardized trait and occurrence datasets for mapping taxonomic and functional diversity and taxon-specific distributions of freshwater insects in the contiguous United States.

Freshwater insects are indicators of ecosystem health, and changes to their biodiversity can signal wider shifts in biodiversity of other taxonomic groups and ecosystem functioning (Bonada et al., 2006; Cardinale et al., 2002; Covich et al., 1999; Perkins et al., 2015; Suter & Cormier, 2015). In addition, populations of freshwater invertebrates are already declining globally owing to global change (Reid et al., 2019). The biodiversity and population health of freshwater insects is consequential for other aquatic and terrestrial organisms, because freshwater insects provide prey for numerous taxa, including freshwater fish, riparian birds, bats and lizards (Baxter et al., 2005), and they are used to assess water quality (Barbour et al., 2000; Bonada et al., 2006). Freshwater insects are also important drivers of nutrient transport within river networks and between terrestrial and aquatic habitats because of their capacity to fly and other life histories (Gounand et al., 2018).

Despite the importance of freshwater insects in both aquatic and terrestrial realms (Baxter et al., 2005; Covich et al., 1999), there are significant gaps in our knowledge of their biodiversity patterns (Balian et al., 2008). Without data on their occurrences and traits, it is difficult to map distributions of freshwater insect taxonomic and trait diversity, especially at broad scales (Balian et al., 2008; Troia & McManamay, 2016).

Systematic surveys of ecological communities provide some of the highest quality occurrence data for assessing biodiversity, but few of these datasets have been integrated over large spatial scales, especially for freshwater insects (Jetz et al., 2019; Troia & McManamay, 2016). Incidence data and range maps are also limiting for freshwater insects. For example, insect occurrence records from the Global Biodiversity Information Facility (GBIF), derived primarily from museum collections, are sparse (Troia & McManamay, 2016), and expert range maps from the International Union for the Conservation of Nature (IUCN) are available for only one of the nine major freshwater insect orders, damselflies and dragonflies (Odonata) (IUCN, 2020). Ecologists still lack a dataset of systematically surveyed freshwater insect occurrence records covering the major freshwater insect orders and spanning the contiguous United States. As a consequence, previous studies have mapped stream insect diversity for only a subset of insect orders (e.g., Ephemeroptera, Plecoptera, Trichoptera; Shah et al., 2014; Vinson & Hawkins, 2003) or for regions of the USA (Poff et al., 2010; Pyne & Poff, 2017). One of our goals was to integrate systematically surveyed community data for freshwater insects into occurrence datasets for biodiversity mapping.

Environmental agencies in the USA and in countries throughout the world use macroinvertebrates in bioassessment of stream condition in compliance with mandates to protect the ecological integrity of surface waters (Barbour et al., 2000; Bonada et al., 2006). In the USA, local, tribal, state and federal agencies have monitored macroinvertebrate community composition at georeferenced stream locations since the passage of the Clean Water Act in 1972 (Barbour et al., 2000). These systematic community surveys provide a rich source of information about stream insect occurrences. Some of these data are already publicly available online through the WQP, including data from the U.S. Geological Survey (USGS) National Water Quality Assessment and the USEPA National Aquatic Resource Surveys. However, additional monitoring data from state agencies have yet to be integrated and released as open-access datasets. A database is needed that collates and standardizes the biological monitoring data from these disparate sources and integrates them with trait databases using consistent and updated trait terminology (Schmera et al., 2015) and up-to-date taxonomy. It is important to standardize and integrate traits with freshwater insect occurrence records, because trait distributions are needed to assess biodiversity patterns and monitor the ecological integrity of surface...
waters (Schmera et al., 2017; Statzner & Bêche, 2010). An integrated database of occurrence records and functional traits will facilitate the mapping of stream insect diversity in the USA.

There is a long history in stream ecology of using functional traits of stream macroinvertebrates to measure aquatic community and ecosystem responses to environmental stressors (Dolédec et al., 1999; Statzner & Bêche, 2010). The composition of insect traits, such as body size, functional feeding group and morphology, is influenced by in-stream habitat measures, including velocity and timing of stream flow (the habitat template; Townsend & Hildrew, 1994), and by landscape filters, including climate and human activity (Poff, 1997). Therefore, the trait composition of stream insect communities is often used to infer the impacts of human disturbance (Bonada et al., 2006), and traits are widely incorporated into indicator analyses by state and federal agencies for assessing stream condition (e.g., Mazor et al., 2016; Stoddard et al., 2008).

Previous efforts to standardize and document traits of stream insects for the USA have resulted in a widely used, publicly available dataset, the USEPA Freshwater Biological Traits database (U.S. EPA, 2012). The initial data for the USEPA database were compiled for the USGS by Vieira et al. (2006) and subsequently reclassified by Poff et al. (2006) to reflect functional trait niches of lotic insects. However, there remain significant gaps in trait coverage. Many insect taxa were never assigned traits, and many more have assignments for only a single trait, such as body size. The USEPA database also contains limited data on trait variation within genera, by species, literature source or geographical region, and the database does not summarize this variation using fuzzy trait assignments commonly used by researchers in Europe and other regions (Schmera et al., 2015). Moreover, the trait assignments are not consistent with a recently proposed unified terminology for traits of stream organisms (Schmera et al., 2015). Therefore, the U.S. traits are not compatible with those used in Europe and other regions. In addition, there have been recent efforts to update functional trait databases of European freshwater macroinvertebrates (Mürria et al., 2020; Sarremejane et al., 2020). Updating and expanding on the USEPA traits database by increasing the number of trait assignments, standardizing taxonomy and trait terminology and providing trait variation in the form of fuzzy traits would facilitate macroecological (continental to global) mapping and assessments of stream insect trait composition and functional diversity.

We present a database, Freshwater insects CONUS: A database of freshwater insect occurrences and traits for the contiguous United States, for genera from the major freshwater insect orders: Coleoptera, Diptera, Ephemeroptera, Hemiptera, Lepidoptera, Megaloptera, Neuroptera, Odonata, Plecoptera and Trichoptera. Our occurrence dataset contains >2.05 million occurrence records for 932 genera sampled from 51,044 stream locations between 2001 and 2018. Our trait dataset includes dispersal, ecological, life-history and morphological traits (Table 1) assigned at the genus level for 1,007 freshwater insect genera, including the 932 genera in our occurrence dataset. Our occurrence records are primarily from wadeable streams, and our trait dataset is primarily for stream insects, although some occurrence records are from larger rivers, and some insects assigned traits also occur in ponds, lakes or rivers. We build upon the foundational occurrence and trait databases described above by integrating occurrence records from state agencies that were not accessible online and by providing updated, standardized taxonomy and trait terminology. We also greatly expand the number of insect genera with trait assignments and provide fuzzy traits to facilitate integration and comparison with trait databases in other regions of the world. Together, these datasets facilitate mapping of the geographical distributions of stream insect diversity, in addition to distributions of individual insect genera and traits.

2 | METHODS

We implemented five sequential steps of compiling data sources, digitalizing data, data cleaning, taxonomic harmonization and trait assignment (Figure 1). We detail these steps below.

2.1 | Data sources

We compiled our freshwater insect occurrence database by downloading records from the WQP in February 2017 as follows. We selected “All” for Location, Site and Sampling parameters. We selected “Invertebrates” and “Benthic macroinvertebrates” for the Assemblage and “All” for the Taxonomic Name under Biological Sampling parameters. This resulted in a dataset of 2,738,480 records for macroinvertebrate taxa identified to order, family, genus or species from 66,356 sampling locations, before data cleaning and taxonomic harmonization. To fill spatial gaps in occurrence records, we requested biomonitoring data from 30 state agencies and downloaded or received records from 19 agencies. This added 6,067,204 records from 55,791 locations, some of which were duplicates of the WQP data.

We began to assemble the freshwater insect trait dataset by downloading records from the USEPA database in September 2017. The USEPA database contains trait information from 967 publications and government reports spanning 2005–2017, but primary data sources are Vieira et al., (2006) and Poff et al., (2006). The database includes habitat, life-history, mobility, morphological and ecological trait data for 1,343 North American macroinvertebrate genera, including freshwater insects, molluscs and arachnids. We subset the USEPA database to include only insect taxa, which resulted in a dataset of traits for 908 insect genera before harmonizing genus names with the latest taxonomic designations. We cross-referenced genera between the USEPA database and our occurrence dataset to search for taxa without trait assignments, identified the needed traits for those taxa, and filled the gaps in trait data through systematic literature review. We also added trait data for taxa already in the USEPA database that were missing assignments for some traits.

We began by merging unpublished trait data compiled in 2014 for a Californian project on stream hydrology (Mazor et al., 2016; Stein et al., 2017). This dataset focused on macroinvertebrates found in Californian streams that were not represented in the trait database of
| Grouping feature | Trait group         | Trait             | Definition                                      | Definition citation |
|------------------|---------------------|-------------------|------------------------------------------------|---------------------|
| Life history     | Number of generations per year | Semivoltine       | Less than one generation per year              | Poff et al. (2006)  |
|                  |                     | Univoltine        | One generation per year                        | Poff et al. (2006)  |
|                  |                     | Bi_multivoltine   | More than one generation per year              | Poff et al. (2006)  |
| Synchronization of emergence |         | Well              | Emergence occurs within a matter of days      | Poff et al. (2006)  |
|                  |                     | Poorly            | Emergence occurs within a matter of weeks or months | Poff et al. (2006)  |
| Emergence season | Spring              |                  | Emergence between the months of March and May |                     |
|                  |                     | Summer            | Emergence between the months of June and August|                     |
|                  |                     | Fall              | Emergence between the months of September and November|                     |
|                  |                     | Winter            | Emergence between the months of December and February|                     |
| Dispersal        | Female dispersal    | Low               | <1 km flight before laying eggs                | Poff et al. (2006)  |
|                  |                     | High              | >1 km flight before laying eggs                | Poff et al. (2006)  |
|                  | Adult flying strength | Weak              | Taking frequent breaks while flying, or flight is low to the ground | Poff et al. (2006)  |
|                  |                     | Strong            | Able to fly into a light breeze or fly for several miles without breaks | Poff et al. (2006)  |
| Morphology       | Maximum body size   | Small             | <9 mm                                           | Poff et al. (2006)  |
|                  |                     | Medium            | 9–16 mm                                         | Poff et al. (2006)  |
|                  |                     | Large             | >16 mm                                          | Poff et al. (2006)  |
| Respiration mode | Tegument            |                  | An outer covering, outer enveloping cell layer or membrane used to acquire oxygen | Merritt et al. (2008) |
|                  |                    | Gill              | A thin-walled structure with trachea, used for the absorption of oxygen | Arnett (2000)       |
|                  |                    | Plastron, spiracle| Oxygen is absorbed from the atmosphere, from aquatic plants or from a temporary air store, such as an air film or bubble on the surface of the body, or a permanent air store (a plastron) | Merritt et al. (2008)  |
| Ecology          | Rheophily           | Depo              | Occupies running-water pools or margins with fine sediments (sand and silt)| Merritt et al. (2008) |
|                  |                     | Depo_eros         | Occupies both erosional and depositional habitats| Merritt et al. (2008) |
|                  |                     | Eros              | Occupies running-water riffles with coarse sediments (cobbles, pebble, gravel)| Merritt et al. (2008) |
| Thermal preference |                    | Cold stenothermal | <5 °C                                           | Vieira et al. (2006)  |
|                  |                    | Cold- cool eurythermal | 0–15 °C                                       | Vieira et al. (2006)  |
|                  |                    | Cool- warm eurythermal | 5–30 °C                                       | Vieira et al. (2006)  |
|                  |                    | Warm eurythermal  | 15–30 °C                                        | Vieira et al. (2006)  |
|                  |                    | Hot eurythermal   | >30 °C                                          | Vieira et al. (2006)  |
| Habit            | Crawler            |                  | Adapted for crawling on the surface of floating leaves of vascular hydrophytes or fine sediments on the bottom of water bodies | Merritt et al. (2008) |

(Continues)
Poff et al., (2006) and included trait assignments for 73 insect genera not in the USEPA database. The 2014 trait data were compiled using a systematic search of: (a) the trait databases of Vieira et al., (2006) and USEPA; (b) freshwater entomology books and taxonomic identification manuals; and (c) peer-reviewed articles of each taxon (mostly at genus level) that contained life-history information (for citations of data sources used in this 2014 trait compilation, see the Appendix). If there were gaps remaining in trait information, an expert taxonomist was consulted to fill in the gaps (Boris Kondratieff, personal communication).

After merging the unpublished trait data, we conducted an initial search of the freshwater insect trait literature in the contiguous United States. We began by searching freshwater entomology books and published and online taxonomic identification manuals. We then followed established guidelines for conducting a systematic search of the primary literature (Pullin & Stewart, 2006). We searched Web of Science, Google Scholar and the library catalogue at our university to identify peer-reviewed papers containing information on the ecology of freshwater insects. This search was conducted from September 2017 to December 2018 and referenced papers from 1911 to 2018. We used the following search terms: genus AND Emergence synchron* OR emergence season* OR feed mode* OR dispersal* OR flight strength OR flying strength OR voltinism OR thermal preference OR rheophil* OR respir* OR body size OR habit OR larvae OR gill OR tegument OR plastron OR depositional OR erosional. We retained sources published in English with one or more of these search terms in the abstract, title or key words and that contained dispersal, ecological, life history or morphological information for freshwater insects of North America. In addition to published trait sources, we used iNaturalist citizen science data (https://www.inaturalist.org/, accessed in 2018) to assign the emergence season. These data consist of time-stamped occurrence records submitted by commercial and recreational fisherman since May 2013 in order to track the emergence dates of freshwater insects across North America. Sources for trait data are provided in the final data tables (Figure 2).

| Grouping feature | Trait group | Trait | Definition | Definition citation |
|------------------|------------|-------|------------|---------------------|
| Burrower         |            | Inhabiting the fine sediment of streams and lakes | Merritt et al. (2008) |
| Clinger          |            | Representatives have behavioural and morphological adaptations for attachment to surfaces in stream riffles and wave-swept rocky littoral zones of lakes | Merritt et al. (2008) |
| Skater           |            | Adapted for skating on the surface, where they feed as scavengers on organisms trapped in the surface film | Merritt et al. (2008) |
| Swimmer          |            | Adapted for fish-like swimming in lotic or lentic habitats | Merritt et al. (2008) |
| Sprawler         |            | Inhabiting the surface of floating leaves of vascular hydrophytes or fine sediments | Merritt et al. (2008) |
| Climber          |            | Adapted for living on vascular hydrophytes or detrital debris, with modifications for moving vertically on stem-type surfaces | Merritt et al. (2008) |
| Planktonic       |            | Inhabiting the open water limnetic zone of standing waters | Merritt et al. (2008) |

| Feeding style    |            | Predators that ingest prey whole or in parts (engulfers) or that pierce prey tissues and suck fluids (piercers) | Merritt et al. (2008) |
| Collector-gatherer |        | Insects that collect and consume decomposing organic matter | Cummins (1973) |
| Collector-filterer |        | Insects that collect and filter living algal cells or detritus | Merritt et al. (2008) |
| Herbivore        |            | Insects that scrape algae or that shred or pierce living aquatic plants | Merritt et al. (2008); Poff et al. (2006) |
| Shredder         |            | Insects that shred decomposing vascular plant tissue (detritivores) | Poff et al. (2006) |
| Parasite         |            | Parasites that consume living animal tissue | Merritt et al. (2008) |

Note: To be consistent with the unified trait terminology for stream organisms proposed by Schmera et al., (2015), we have reorganized traits by grouping feature and trait groups. A definition for each trait and literature citation for that definition are provided.
2.2 | Data digitalization

We digitized details about sampling methodology that were absent from state datasets. We requested the geodetic datum of horizontal coordinates for sampling locations through e-mails with agencies. We also recorded the sampling equipment and area of the stream bottom sampled by requesting methodology directly from agencies or by digitizing information in state field sampling manuals. These details could potentially be used to estimate sampling effort across sites. However, many agencies did not record their sampling methods for some samples, and thus gaps remain in the documentation of sampling methodology.

When digitizing trait information, we focused on a subset of the traits originally documented in the USEPA database that should be influenced by environmental gradients of climate, land use, topography and base flow that are important predictors of stream insect functional composition at broad spatial extents (Bonada et al., 2007; Diaz et al., 2008; Lawrence et al., 2010; Poff et al., 2010; Pyne & Poff, 2017;
Statzner & Bêche, 2010). We organized traits following recommendations for a global, unified trait terminology for stream ecology (Schmera et al., 2015). We summarized traits into “trait groups” of closely related traits (e.g., “small”, “medium” and “large” are traits of the trait group “maximum body size”) and grouped related trait groups into “grouping features” of life history, dispersal, morphology or ecology (Table 1).

When digitizing traits from entomology books and taxonomic guides, we reviewed each source for all genera in our database with missing traits. When pulling information from the primary literature, we searched systematically for traits one genus at a time. Where possible, we also converted trait textual descriptions in the “comments” column of the USEPA database into trait assignments. We recorded traits at the genus or the species level using accepted trait definitions (Table 1). We documented trait variation within each genus by separating sources by row when compiling traits from multiple literature sources for a single genus. In addition, traits for the same genus from different geographical regions and traits for different species within a genus were separated by row. Thus, each genus could have a different trait recorded for each row based on the species, region and literature source. If, for any source (within a row) there were two or more possible traits from the same trait group (Table 1), we recorded the most commonly occurring trait documented by the source while also noting all other possible traits as “trait comments”. Although the traits recorded as “comments” did not influence final trait assignments,

![Database layout, with connecting lines indicating relationships among tables. Orange boxes are the “raw” community and trait datasets cleaned from data in the Data_Sources table (purple) using R scripts. “Cleaned” trait tables are shown in green and occurrence records in blue. Tables of ancillary information are in grey. From left to right: Raw_Traits contains data for each genus varying by location, species and literature source, which we digitized and cleaned during steps 2, 3 and 4 of database assembly (Figure 1). Genus_Traits and Genus_Trait_Affinities contain modal traits and trait affinities that we produced from Raw_Traits using R scripts during step 5. Ancillary_Trait contains information about each trait (Table 1). Genus_Occurrences contains occurrence records that we produced from Raw_Community_Data using R scripts in database assembly steps 3 and 4. Ancillary_Taxonomy contains taxonomic names recorded in the Water Quality Portal (WQP), state data and U.S. Environmental Protection Agency (USEPA) database, with their corresponding accepted names, taxonomic serial numbers and higher taxonomic designations obtained during step 4. Raw_Community_Data contains occurrence data from the WQP and state agencies supplied in data tables listed in Data_Sources. We recorded additional data about state sampling methodology in Ancillary_Sample_Method during step 2. We cleaned the data files in Data_Sources using R scripts during steps 3 and 4 [Colour figure can be viewed at wileyonlinelibrary.com]
they are provided with the final datasets as additional natural history
information. We summarized trait variation across rows within a genus
(across species, regions and literature sources) into trait affinities (anal-
gous to fuzzy-coded traits; see Assigning trait membership, below).

One limitation of both the USEPA database and our database
is that traits are not well defined by life-history stage for certain
taxa. Most freshwater insects have an obligate aquatic larval stage
transitioning to a terrestrial adult stage, and traits for these insects
are assigned for the aquatic larval stage. However, many insects in
the orders Coleoptera and Hemiptera are aquatic in both larval and
adult stages and have traits that differ by stage (Merritt et al., 2008).
Most trait entries for these taxa in the USEPA database are for the
adult stage. Likewise, we found during our systematic search of the
trait literature that adult traits for Coleoptera and Hemiptera were
more commonly available than larval traits. Therefore, there is a bias
toward traits for adult stages of Coleoptera and Hemiptera in our
database. In addition, traits defining reproduction and life span are
not well represented in our database because these traits were not
readily available in the primary literature or the USEPA database for
the majority of taxa.

### 2.3 Data cleaning

During the first step of data cleaning, we removed duplicate occurrence
records and those with missing coordinates. Next, we examined
records visually for georeferencing errors by mapping all occurrence
locations for each insect family and comparing maps of their distri-
butions with GBIF range maps (GBIF, 2020). This represents an inde-
pendent assessment of range, because most GBIF records are from

| Data table name | Content | Links to other tables | Database assembly steps |
|-----------------|---------|-----------------------|-------------------------|
| Raw_Traits      | Cleaned trait data using R scripts for each taxonomic name (“Submitted name Trait”, usually genus, occasionally species or family) recorded in datasets from the WQP, state agencies or USEPA. There are multiple trait entries separated by row for each taxon, with each row presenting trait data recorded from a different location, species or literature source | Ancillary_Taxonomy through “Submitted_name” column | 1, 2, 3, 4 |
| Genus_Traits    | Modal traits for each genus assigned from data in Raw_Traits using R scripts | Genus_Trait_Affinities and Ancillary_Trait through “Trait” column. Genus_Occurrences and Ancillary_Taxonomy through “Genus” column | 5 |
| Genus_Trait_Affinities | Trait affinities for each genus assigned from data in Raw_Traits using R scripts | Linkages are the same as for Genus_Traits, above | 5 |
| Ancillary_Trait | Information about traits contained in Table 1 | Genus_Traits and Genus_Trait_Affinities through “Trait” column | 1 |
| Genus_Occurrences | Genus occurrence records produced from Raw_Community_Data using R scripts | Genus_Traits through “Trait” column and Ancillary_Taxonomy through “Genus” column | 3, 4 |
| Ancillary_Taxonomy | Data from taxonomic harmonization, including taxonomic names (“Submitted_name”) recorded in the WQP, state data and USEPA database and the corresponding accepted names, taxonomic serial numbers and higher taxonomic designations. Users can search on any column in Ancillary_Taxonomy and find corresponding occurrence and trait records in other tables | Raw_Traits and Raw_Community_Data through “Submitted_name” column. Genus_Traits, Genus_Trait_Affinities and Genus_Occurrences through “Genus” column | 4 |
| Data_Sources | Information about source data files, state agency websites and agency contacts | | 1 |
| Raw_Community_Data | Cleaned occurrence data from the WQP and state agencies using R scripts. Includes records for taxa identified to species, genus, family or order | Genus_Occurrences through “Unique_ID”, Ancillary_Taxonomy through “Submitted_name”. Ancillary_Sample_Method through “Sample_method” | 2, 3, 4 |
| Ancillary_Sample_Method | Detailed methodology for sample methods in Raw_Community_Data | Raw_Community_Data through “Sample_method” and Data_Sources through “Data_source” | 2 |

Note: “Links to other tables” indicates which columns can be used to join related tables. The database assembly steps (Figure 1) involved in creating each table are also provided.

Abbreviations: USEPA, U.S. Environmental Protection Agency; WQP, Water Quality Portal.
museum collections. In addition, we searched data providers and datasets in GBIF for the agencies that provided our occurrence records and found no records of data contributions to GBIF from those providers. We removed obvious geographical outliers (e.g., points in the ocean) and corrected transposed latitude and longitude coordinates and those coordinates with an incorrect sign on the decimal degrees of latitude or longitude. We also mapped data by state to assess georeferencing errors (records falling outside state bounds). In total, we removed 5,325,297 duplicate records and 836,310 records that were missing sampling coordinates or contained georeferencing errors. We then removed an additional 211,627 records during taxonomic harmonization (see Taxonomic harmonization, below) either because records were for non-insect taxa or because misspellings or other errors rendered the taxa unidentifiable. This resulted in a dataset of 2,432,450 occurrence records of insects identified to order, family, genus or species, from 55,791 sampling locations. We performed data cleaning and taxonomic harmonization in R v.3.5.3 (R Core Team, 2019). Scripts with R code for data cleaning are provided through GitHub and the Environmental Data Initiative (see Data organization and usage, below).

### 2.4 Taxonomic harmonization

After data cleaning, we verified and harmonized taxonomic names between the occurrence and trait datasets using the "taxize" package v.0.9.92 in R (Chamberlain et al., 2019). We used "taxize" to search the database of the Integrated Taxonomic Information System (ITIS) to extract updated genus names, taxonomic serial numbers and upstream names (Family, Order) for each taxon. Some names were not found in the ITIS database owing to misspellings, missing data in ITIS (e.g., for recently identified taxa) or because names were invalid and ITIS contained no valid synonyms. For these cases, we verified names by manually searching other online sources, including GBIF (GBIF.org, 2020), IUCN (IUCN, 2020) and the primary literature. We accepted names that were listed as valid U.S. taxa by the majority of sources. Although we assigned an accepted name for those taxa, we could not assign an accepted taxonomic serial number from ITIS. In addition, some names could not be verified using any source. For those taxa, we assigned the valid upstream name (Family or Order) from ITIS. In total, we re-assigned names for 413 taxa in the trait dataset, including 58 changes to the genus name. In the occurrence dataset, we re-assigned 704 names, including 177 genus names, 96 of which were combined into 36 genera.

### 2.5 Assigning trait membership

We assigned membership to traits at the genus level in two ways, as modal traits and as trait affinities (Figure 1). We assigned modal traits as the most frequently occurring trait in a trait group (Table 1) across all species, geographical regions and literature sources (rows) for a genus. Affinity scores account for trait variation within a genus by species, geographical region or literature source and are analogous to fuzzy-coded traits used by researchers in Europe and other regions (Schmera et al., 2015; Usseglio-Polatera et al., 2000). Trait affinities differ from fuzzy-coded traits in that they are assigned as proportions, whereas fuzzy-coded traits are typically assigned using an ordinal scale of zero to three or five and are also occasionally expressed on a continuous scale from 0 to 100% (Schmera et al., 2015; Usseglio-Polatera et al., 2000). They were assigned by computing the proportion of rows for each genus that were assigned to each trait in a trait group, such that each row counted as a single trait contribution. Thus, each species, geographical location and literature source for a genus contributed a single value toward the affinity score. Affinity scores sum to one across all traits in a trait group for each genus.

### 3 Results and Discussion

The freshwater insects CONUS database is organized as nine relational data tables with associated metadata (Figure 2; Table 2). Metadata accompanying the dataset include information on project funding, contributors, geographical and temporal scope, variable names, descriptions, measurement scales, missing values and trait codes. We provide our data tables as .csv files with metadata through the Environmental Data Initiative (EDI). The R scripts that we used for data cleaning are also available through the EDI and GitHub. We encourage submissions of occurrence and trait records for future updates to the database. A template and instructions for data submission are available at GitHub. See Data availability statement, below, for links to the EDI and GitHub repositories.

Here, we describe a few of the many uses for our database. In its "raw" form, users can extract trait data for insects identified to order, family, genus or species by merging the Raw_Traits table with the Ancillary_Trait and Ancillary_Taxonomy tables (Figure 2). We recorded trait variation in Raw_Traits, with each row for a genus presenting trait data for a different species, location or literature source. Users can thus extract traits by state ("Study_location_state") or literature source ("Study_citation") or can summarize trait variation within a genus, family or order when the Raw_Traits table is merged with the Ancillary_Taxonomy table (through the "Submitted_name" column). In addition, users can merge the Raw_Community_Data and Ancillary_Taxonomy tables (Figure 2) to find occurrence records for insect species and map their distributions across the USA (as in Figure 3, for insect genera). Searching columns in the Raw_Community_Data table enables users to extract and map insect records for each state ("Study_state" column), monitoring organization ("Monitoring_organization" column) or type of water body ("Location_description" column). Moreover, merging the Raw_Community_Data and Ancillary_Sample_Method tables will enable users to isolate records that were sampled using particular equipment or a particular protocol, such as a Hess sampler, D-frame...
In the “cleaned” form of the database, the Genus_Occurrences table enables users to map genus richness (Figure 3a) and the distributions of individual insect genera. In addition, when merged with the Ancillary_Taxonomy table, records in the Genus_Occurrences table enable the mapping of distributions of genus richness and individual insect genera by family or order (Figure 3b). A great strength of the “cleaned” data tables comes from merging the Genus_Occurrences table with the Genus_Traits or Genus_Trait_Affinities table using the aquatic dipnet or Hester-Dendy sampler, by searching the “Sample_method” column.

FIGURE 3 (a,b) Genus richness by occurrence location for all orders (a) and for each order individually (b). Dark points indicate low genus richness and red indicate high richness. Note that genus richness has not been corrected for sampling bias [Colour figure can be viewed at wileyonlinelibrary.com]
This enables users to examine the spatial distributions of insect traits and trait affinities by genus (Figure 4), and by family or order when also combined with the Ancillary_Taxonomy table. More nuanced mapping of trait distributions is also possible. For example, users could map trait distributions for a particular state, monitoring organization, water body type or sampling methodology, as described above for the “raw” data tables.

### 3.2 Biodiversity patterns in data

We mapped insect genus richness by location using data in Genus_Occurrences (Figure 3). By merging Genus_Occurrences and Genus_Traits, we also mapped distributions of freshwater insect functional traits for the contiguous United States (Figure 4). These maps reveal some obvious sampling biases (see Bias in occurrence and trait

**TABLE 3** Number of genus occurrence and trait records by insect order

| Order       | Number of genus occurrence records | Number of genus occurrence locations | Number of genera with occurrence records | Number of genera with trait records | Number of species with records in each order |
|-------------|------------------------------------|-------------------------------------|------------------------------------------|-------------------------------------|---------------------------------------------|
| Coleoptera  | 210,077                            | 44,669                              | 145                                      | 160                                 | 464                                         |
| Diptera     | 862,826                            | 49,572                              | 335                                      | 363                                 | 556                                         |
| Ephemeroptera | 381,077                          | 46,737                              | 93                                       | 100                                 | 426                                         |
| Hemiptera   | 22,528                             | 10,231                              | 48                                       | 56                                  | 140                                         |
| Lepidoptera | 3,756                              | 2,956                               | 18                                       | 4                                   | 9                                           |
| Megaloptera | 25,056                             | 15,302                              | 9                                        | 8                                   | 13                                          |
| Neuroptera  | 389                                | 362                                 | 3                                        | 2                                   | 4                                           |
| Odonata     | 73,760                             | 23,102                              | 66                                       | 73                                  | 427                                         |
| Plecoptera  | 137,377                            | 29,155                              | 97                                       | 99                                  | 401                                         |
| Trichoptera | 338,460                            | 46,682                              | 127                                      | 145                                 | 673                                         |
records, below) and interesting patterns in the distributions of functional traits. For example, insect genera with bivoltine or multivoltine life cycles, corresponding to short generation times, and genera that prefer warm eurythermal habitats are concentrated in warm, low-lying regions, including southern California and Florida (Figure 4). We see the opposite patterns for some rheophily and respiration traits, where gilled insects and those preferring erosional habitats are concentrated in mountainous regions of the western and northeastern USA. Previous studies suggest that gilled insects and those with adaptations to life in erosional habitats should be found in cool, well-oxygenated and fast-flowing waters, such as are found in high-elevation streams (Poff et al., 2010; Statzner & Béche, 2010). These hypotheses could be tested definitively by combining our database with environmental data.

3.3 | Bias in occurrence and trait records

Our maps of genus richness (Figure 3) clearly illustrate spatial bias in occurrence records. These biases are partly attributable to the fact that some state agencies have not digitized their biological monitoring data. Moreover, sampling effort, including the number of samples and the area sampled, varied within and among datasets. These sources of bias resulted in sparse genus occurrence records in several states in the Midwest, mountain West and Southeastern USA (Figure 3a). There are also obvious gaps in occurrences and traits for the insect orders Hemiptera, Lepidoptera, Megaloptera and Neuroptera (Figure 3b; Table 3). Fewer aquatic insect genera reside within these orders in comparison to the obligate aquatic orders Ephemeroptera, Plecoptera and Trichoptera or the other well-represented aquatic orders, Coleoptera and Diptera. Their relative rarity could have resulted in training biases, in which aquatic ecologists and taxonomists are less likely to identify uncommon taxa accurately, or targeted sampling biases, in which sampling methodology is designed to capture genera from common orders.

Another common source of bias originated when identifying specimens in the laboratory. Some state agencies identify all macroinvertebrate specimens to family, whereas others use inconsistent methodology by identifying some taxonomic groups (e.g., Dipterans) to family or order and other groups to genus. We removed all

![FIGURE 5](image-url)
records for insects identified to family when producing our Genus_Occurrences table, which effectively excluded whole state datasets. However, records for insects identified to family or order are still available in Raw_Community_Data.

Biases in occurrence records could be corrected by aggregating records using a larger spatial unit and then applying coverage-based rarefaction (Chao & Jost, 2012) to down-weight the influence of well-sampled areas on spatial patterns of genus richness. For example, one could aggregate occurrence records by watershed (e.g., USGS hydrological units), treating each occurrence location as a spatial replicate, and then compute the sample coverage in each watershed. One would then rarefy or interpolate genus richness for equal levels of sample coverage across watersheds. Coverage-based rarefaction can be performed with the "iNext" package in R (Hsieh et al., 2016). The R packages "biogeo" and "dismo" can also assist with assessment of bias in occurrence records and modelling genus distributions (Hijmans et al., 2017; Robertson et al., 2016).

Trait coverage was most complete for the ecological trait groups feeding style, habit and rheophily and the morphological trait group maximum body size (Figure 5a). The trait groups with the fewest genera having assignments included the following life-history and dispersal trait groups: synchronization of emergence, emergence season, female dispersal and adult flying strength (Figure 5a). Approximately half of the insect genera in our database are still missing assignments for these four trait groups. In addition, there are gaps in coverage for all traits; no trait group contains a trait assignment for every genus in our database. These gaps highlight the need for more trait measurements of freshwater insects, especially insects in the orders Hemiptera, Lepidoptera, Megaloptera and Neuroptera (Figure 3b; Table 3). Moreover, there is bias toward adult stage traits for Coleoptera and Hemiptera (see Data digitalization, above), which indicates that more trait measurements are also needed for larval stages of insects in these orders. We expect that additional trait data will be available in books and scientific articles that have yet to be digitized and standardized, and many research programmes have trait datasets that are not published in any form. We encourage submission of these unpublished datasets to future updates of our database (see the Data availability statement, below).

Data sources for most of our 11 trait categories were from every state in the contiguous United States. However, there are geographical biases in trait assignments for certain trait groups, including female dispersal and emergence synchrony, which we derived from studies conducted in < 30 states. Another source of geographical bias arises from the USEPA database (a major source for our database), which contains a large amount of trait information from insects in Maine, North Carolina and Utah. In addition, the trait data from the USEPA database were compiled by researchers from Colorado State University (Poff et al., 2006; Vieira et al., 2006). Geographical biases of the researchers and locations of trait source data could bias the assignment of modal traits or affinities for certain trait groups, such as thermal preference, that are spatially influenced by environmental variables. Over-representation of trait information for certain species within a genus could also skew trait assignments toward values for those species. Trait affinities (analogous to fuzzy-coded traits) help to account for these sources of bias by quantifying trait variation for each trait group within a genus across species, geographical areas and literature sources. Data users should compare modal traits with trait affinities and the data in Raw_Traits to gain insight into the sources of trait variation and biases for each genus. These sources of bias are not unique, and future updates to our database will improve the geographical scope and resolution of traits across species within each genus.

3.4 Comparison with other datasets

We tripled the number of occurrence records and locations from what was available in the WQP, and we added occurrence records for 118 genera that were not previously available in open-access databases. The WQP contained 677,005 genus occurrence records from 18,705 locations and 814 genera, after data cleaning and taxonomic harmonization. Our Freshwater insects CONUS database contains > 2.05 million genus occurrence records for 932 genera at 51,044 stream locations. Of the occurrence records, 565,376 are repeat detections of the same taxa over time.

We nearly doubled the number of trait records available for the 11 trait groups we considered, from 24,655 traits in the USEPA database to 47,000 in our Freshwater insects CONUS database (Raw_Traits; Figure 2). As a result, we increased the number of genera assigned a modal trait (Figure 5). After taxonomic harmonization and data cleaning, the USEPA database contained traits for 827 insect genera, to which we added traits for 180 genera, for a total of 1,007 insect genera with trait assignments (Figure 5; Table 3). We also updated taxonomic names to reflect the most current genus designations and trait assignments to align with the unified trait terminology for stream organisms (Schmera et al., 2015). Finally, we added trait affinities (Genus_Trait_Affinities; Figure 2), which were not included in the USEPA database, in order to facilitate conversion of U.S. traits to the European system of fuzzy coding and account for trait variation within genera.

3.5 Conclusions

Our Freshwater insects CONUS database provides the most comprehensive datasets of freshwater insect occurrence records and traits for the contiguous United States by including records for a majority of the estimated 1,160 freshwater insect genera in North America (Balian et al., 2008). Our occurrence dataset provides good spatial coverage of occurrence records for most of the major freshwater insect orders because our data are derived from systematic community surveys. Another strength of our database is that our trait data are more comparable to datasets used by researchers in Europe and other regions of the world by including trait variation as trait affinities, analogous to fuzzy-coded traits, and using unified trait terminology (Schmera et al., 2015). These components are included to
facilitate the linkage of our database to those in other countries for cross-continental analyses of functional composition and diversity in freshwater insects. We identified regions of the USA and taxa for which more occurrence and trait data are needed, and we encourage data submissions for future updates to our database. Our database can be used to map freshwater insect taxonomic and functional diversity and, when paired with environmental data, will provide a powerful resource for quantifying how the environment shapes diversity patterns, in addition to taxon-specific distributions, across the contiguous United States.

ACKNOWLEDGMENTS
We would like to thank the state agencies that shared their data, including AR Department of Environmental Quality, AZ Department of Environmental Quality, CA Environmental Data Exchange Network, FL Department of Environmental Protection, IA Department of Natural Resources, ID Department of Environmental Quality, MN Pollution Control Agency, MO Department of Natural Resources, MS Department of Environmental Quality, NE Department of Environmental Quality, NM Environment Department, NV Division of Environmental Protection, OH Environmental Protection Agency, OR Department of Environmental Quality, PA Department of Environmental Protection, TN Department of Environment and Conservation, TX Commission on Environmental Quality, WA State Department of Ecology and WV Department of Environmental Protection. We also thank Kate Boersma, Beth Gerstner and Dana Infante for discussions and feedback on the manuscript, Minali Bhatt and Erika Ralston for assistance with collecting trait data, and Pat Bills for help with data organization. L.T. was funded by a NASA Earth and Space Science Fellowship (NASA #80NSSC17K0395) and Michigan State University. P.Z. and E.H. were funded by Michigan State University. M.P. was funded by California State Water Resources Control Board grant #12-430-550. This research was additionally supported by United States Department of Agriculture National Institute of Food and Agriculture Hatch Project #1010055.

AUTHOR CONTRIBUTIONS
L.T. and P.Z. conceived the idea for the database and manuscript; L.T., E.H. and M.P. searched and compiled the data; L.T. designed the database, wrote the R scripts and performed taxonomic harmonization, data cleaning and database formatting; L.T. and E.H. wrote the metadata and manuscript draft; and all authors revised manuscript drafts.

DATA AVAILABILITY STATEMENT
The database is available as .csv files through the Environmental Data Initiative (EDI; https://doi.org/10.6073/pasta/8238ea9bc15840844b3a023bb6bed158) (Twardochleb et al., 2020). We also provide the R scripts used for data cleaning through EDI and GitHub (https://github.com/aquaXterra/freshwater_insects_CONUS). We invite data submissions for future updates to the database. Instructions and a data submission template are available through GitHub (https://github.com/aquaXterra/freshwater_insects_CONUS). To submit data or ask questions about the data submission process, email Dr Phoebe Zarnetske, Michigan State University SpaCE Lab (Spatial & Community Ecology Lab), plz@msu.edu. A link to this repository with updated information can be found at the MSU SpaCE Lab website: www.communityecologylab.com

ORCID
Laura Twardochleb https://orcid.org/0000-0002-8804-9399
Matthew Pyne https://orcid.org/0000-0002-4236-8749
Phoebe Zarnetske https://orcid.org/0000-0001-6257-6951

REFERENCES
Arnett, R. H. A., Jr. (2000). American insects: A handbook of the insects of America north of Mexico (2nd ed.). CRC Press.
Balian, E. V., Segers, H., Lévêque, C., & Martens, K. (2008). The freshwater animal diversity assessment: An overview of the results. Hydrobiologia, 595, 627–637. https://doi.org/10.1007/s10750-007-9246-3
Barbour, M. T., Sweeney, K. F., Jackson, S. K., Courtemanch, D. L., Davies, S. P., & Yoder, C. O. (2000). Measuring the attainment of biological integrity in the USA: A critical element of ecological integrity. In M. Jungwirth, S. Muhar, & S. Schmutz (Eds.), Assessing the ecological integrity of running waters (pp. 453–464). Springer.
Baxter, C. V., Fausch, K. D., & Saunders, W. C. (2005). Tangled webs: Reciprocal flows of invertebrate prey link streams and riparian zones. Freshwater Biology, 50, 201–220. https://doi.org/10.1111/j.1466-8247.2004.00328.x
Belmaker, J., & Jetz, W. (2011). Cross-scale variation in species richness-environment associations. Global Ecology and Biogeography, 20, 464–474. https://doi.org/10.1111/j.1466-8238.2010.00615.x
Bonada, N., Dolédec, S., & Statzner, B. (2007). Taxonomic and biological trait differences of stream macroinvertebrate communities between Mediterranean and temperate regions: Implications for future climatic scenarios. Global Change Biology, 13, 1658–1671. https://doi.org/10.1111/j.1365-2486.2007.01375.x
Bonada, N., Pratt, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. Annual Review of Entomology, 51, 495–523. https://doi.org/10.1146/annurev.ento.51.110104.151124
Butler, E. E., Datta, A., Flores-Moreno, H., Chen, M., Wythers, K. R., Fazayeli, F., Banerjee, A., Atkin, O. K., Kattge, J., Amiaud, B., Blonder, B., Boenisch, G., Bond-Lamberty, B., Brown, K. A., Byun, C., Campetella, G., Cerabolini, B. E. L., Cornelissen, J. H. C., Craine, J. M., ... Reich, P. B. (2017). Mapping local and global variability in plant trait distributions. Proceedings of the National Academy of Sciences of the United States of America, 114, E10937–E10946. https://doi.org/10.1073/pnas.1708984114
Cardinale, B. J., Palmer, M. A., & Collins, S. L. (2002). Species diversity enhances ecosystem functioning through interspecific facilitation. Nature, 415, 426–429. https://doi.org/10.1038/415426a
Chamberlain, S., Szoecs, E., Foster, Z., Arndtsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O’Donnell, J., Oksanen, J., Tzovaras, B. G., Marchand, P., Tran, V., Salmon, M., Li, G., & Grenié, M. (2019). taxize: Taxonomic information from around the web. R package version 0.9.9. https://github.com/ropensci/taxize
Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing sampling by completeness rather than size. Ecology, 93, 2533–2547. https://doi.org/10.1890/11-1952.1
Comte, L., & Olden, J. D. (2017). Climatic vulnerability of the world’s freshwater and marine fishes. Nature Climate Change, 7, 718–722. https://doi.org/10.1038/nclimate3382
Covich, A. P., Palmer, M. A., & Crowl, T. A. (1999). The role of benthic invertebrate species in freshwater ecosystems: Zoobenthic species
richness across North America. *Freshwater Science*, 33, 1136–1147. https://doi.org/10.1086/678492

Statzner, B., & Bèche, L. A. (2010). Can biological invertebrates traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology*, 55, 80–119. https://doi.org/10.1111/j.1365-2427.2009.02369.x

Stein, E. D., Sengupta, A., Mazor, R. D., McCune, K., Bledsoe, B. P., McCune, K., & Adams, S. (2017). Application of regional flow-ecology relationships to inform watershed management decisions: Application of the ELOHA framework in the San Diego River watershed, California, USA. *Ecohydrology*, 10, e1869. https://doi.org/10.1002/eco.1869

Stoddard, J. L., Herlihy, A. T., Peck, D. V., Hughes, R. M., Whittier, T. R., & Tarquinio, E. (2008). A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society*, 27, 878–891. https://doi.org/10.1899/08-053.1

Suter, G. W., & Cormier, S. M. (2015). Why care about aquatic insects: Uses, benefits, and services. *Integrated Environmental Assessment and Management*, 11, 188–194. https://doi.org/10.1002/ieam.1600

Townsend, C. R., & Hildrew, A. G. (1994). Species traits in relation to a habitat template for river systems. *Freshwater Biology*, 31, 265–275. https://doi.org/10.1111/j.1365-2427.1994.tb01740.x

Twardochleb, L. A., Hiltner, E., Pyne, M., Bills, P., and Zarnetske, P. L. (2020). Freshwater insect occurrences and traits for the contiguous United States, 2001–2018 ver 5. Environmental Data Initiative. https://doi.org/10.1086/678492

Troia, M. J., & McManamay, R. A. (2016). Filling in the GAPS: Evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution*, 6, 4654–4669.

Townsend, C. R., & Hildrew, A. G. (1994). Species traits in relation to a habitat template for river systems. *Freshwater Biology*, 31, 265–275. https://doi.org/10.1111/j.1365-2427.1994.tb01740.x

Twardochleb, L. A., Hiltner, E., Pyne, M., Bills, P., and Zarnetske, P. L. (2020). Freshwater insect occurrences and traits for the contiguous United States, 2001–2018 ver 5. Environmental Data Initiative. https://doi.org/10.1086/678492

Wiens, J. J., & Donoghue, M. J. (2004). Historical biogeography, ecology, and species richness. *Trends in Ecology and Evolution*, 19, 639–644. https://doi.org/10.1016/j.tree.2004.09.011

BIOSKETCH

Laura Twardochleb is a Senior Environmental Scientist at California Department of Water Resources studying the effects of adaptive management on estuarine food webs. She earned her PhD in Fisheries and Wildlife and Ecology, Evolutionary Biology and Behavior at Michigan State University, where she investigated global change effects on freshwater ecology at multiple spatial and organizational scales. She holds an MS in Aquatic and Fishery Sciences from the University of Washington.

How to cite this article: Twardochleb L, Hiltner E, Pyne M, Zarnetske P. Freshwater insects CONUS: A database of freshwater insect occurrences and traits for the contiguous United States. *Global Ecol Biogeogr*. 2021;30:826–841. https://doi.org/10.1111/geb.13257

APPENDIX

DATA SOURCES FOR 2014 TRAIT CONTRIBUTION

Brigham, A. R., Brigham, W. V., & Sniffka, A. (1982). *Aquatic insects and Oligochaetes of North and South Carolina*. Midwest Aquatic Enterprises.

Edmunds, G. F., Jr., Jensen, S. L., & Berner, L. (1976). The mayflies of north and central America. University of Minnesota Press.

Epler, J. H. (2006). *Identification manual for the aquatic and semi-aquatic Heteroptera of Florida*. State of Florida, Department of Environmental Protection, Division of Water Resource Management.

Epler, J. H. (2010). *The water beetles of Florida*. State of Florida, Department of Environmental Protection, Division of Environmental Assessment and Restoration.

Hilsenhoff, W. L. (1995). *Aquatic insects of Wisconsin*. Natural History Museums Council, University of Wisconsin-Madison, Wisconsin County Extension Office.

McAlpine, J. F., Peterson, B. V., Shewell, G. E., Teskey, J. R., Vockeroth, J. R. & Wood, D. M. (1981). *Manual of Nearctic Diptera* (Vols. 1 and 2). Biosystematics Research Institute.

McAlpine, J. F., & Wood, D. M. (1989). *Manual of Nearctic Diptera* (Vol. 3). Biosystematics Research Centre.

McCafferty, W. P. (1998). *Aquatic entomology: The fishermen's and ecologists' illustrated guide to insects and their relatives* (revised ed.). Jones and Bartlett.

Menke, A. S. (1979). *The semiaquatic and aquatic Hemiptera of California (Heteroptera: Hemipera)*. Bulletin of the California insect survey (Vol. 21). University of California Press.

Merritt, R. W., Cummins, K. W., & Berg, M. B. (2008). *An introduction to the aquatic insects of North America* (4th ed.). Kendall/Hunt Publishing.

Needham, J. G., Westfall, M., & May, M. L. (2000). *Dragonflies of North America*. Scientific Publishers.

Smith, D. G. (2001). *Pennak's freshwater invertebrates of the United States: Porifera to Crustacea*. John Wiley & Sons.

Stewart, K. W., & Stark, B. P. (2002). *Nymphs of North American stonefly genera* (Plecoptera) (2nd ed.). The Caddis Press.

Tachet, H., Richoux, P., Bournaud, M. & Usseglio-Polatera, P. (2010). *Invertébrés d'eau douce. Systématique, biologie, écologie*. CNRS Editions.

Thor, J. H., & Covich, A. P. (2009). *Ecology and classification of North American freshwater invertebrates* (3rd ed.). Academic Press.

Usinger, R. L. (1963). *Aquatic insects of California: with keys to North American genera and California species*. University of California Press.

Ward, J. V., Kondratieff, B. C., & Zuellig, R. E. (2002). *An illustrated guide to the mountain stream insects of Colorado* (2nd ed.). University Press of Colorado.

Westfall, M. J., & May, M. L. (1996). *Damselflies of North America*. Scientific Publishers.

Wiggins, G. B. (1996). *Larvae of the North American caddisfly genera* (Trichoptera) (2nd ed.). University of Toronto Press.