



OPEN

DATA DESCRIPTOR

A trait dataset for freshwater mussels of the United States of America

Garrett W. Hopper¹✉, Jamie R. Bucholz², Traci P. DuBose³, Kaelyn J. Fogelman⁴, Sean M. Keogh⁵, Megan E. Kubala², Matthew B. Lodato², David H. Nichols⁶, Irene Sánchez González², John M. Pfeiffer⁶, James A. Stoeckel⁷, Jeffrey D. Lozier² & Carla L. Atkinson²

The United States of America has a diverse collection of freshwater mussels comprising 301 species distributed among 59 genera and two families (Margaritiferidae and Unionidae), each having a unique suite of traits. Mussels are among the most imperilled animals and are critical components of their ecosystems, and successful management, conservation and research requires a cohesive and widely accessible data source. Although trait-based analysis for mussels has increased, only a small proportion of traits reflecting mussel diversity in this region has been collated. Decentralized and non-standardized trait information impedes large-scale analysis. Assembling trait data in a synthetic dataset enables comparison across species and lineages and identification of data gaps. We collated data from the primary literature, books, state and federal reports, theses and dissertations, and museum collections into a centralized dataset covering information on taxonomy, morphology, reproductive ecology and life history, fish hosts, habitats, thermal tolerance, geographic distribution, available genetic information, and conservation status. By collating these traits, we aid researchers in assessing variation in mussel traits and modelling ecosystem change.

Background & Summary

Species traits reflect ecological interactions across broad scales including ontogeny, phenology, and phylogeny and thus when traits are compared across communities, they can be a powerful tool in understanding population dynamics, community ecology, evolutionary biology, and resource management. In the absence of robust occurrence datasets, trait data can serve as an auxiliary source of information because certain traits are often correlated with elevated risks of imperilment. To date, trait-based approaches have been used to study species distribution patterns, abundances, and biotic responses to different types of disturbances^{1–3}. Given the escalating biodiversity crisis, a more complete knowledge of species traits is critical to the improvement of natural resource management outcomes, especially for taxa of conservation concern.

One group of concern, the freshwater mussels (Unionoida), are among the most threatened animals globally^{4–7}. Specifically in North America, more than 25 extinctions have occurred in the last 100 years and ~65% of remaining species considered endangered, threatened, or vulnerable⁸. Many traits have been used to address contemporary questions regarding the management and conservation of freshwater mussels in the United States of America (Margaritiferidae and Unionidae). For instance, mussel traits have been used to explain species distributions^{9–13}, changes in assemblages following habitat alterations^{14–17}, potential competition with non-native species^{18,19}, responses to climate change^{20,21}, local extirpations and extinctions^{22,23}, and their functional effects in ecosystems^{24–27}. In addition, the distribution and abundance of mussel species is hypothesized to be linked to the distribution and abundance of certain freshwater fishes^{9,28} because nearly all mussel species have larvae that are parasites on fishes. Exceptions include direct development, where larvae bypass the parasitic life-stage and an

¹School of Renewable Natural Resources, Louisiana State University and Agricultural Center, Baton Rouge, LA, 70803, USA. ²Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, 35487, USA. ³ORISE Postdoctoral Research Fellow, United States Forest Service, Frankfort, KY, 40601, USA. ⁴Department of Biological and Environmental Sciences, Troy University, Troy, AL, 36082, USA. ⁵Field Museum of Natural History, Chicago, IL, 60605, USA. ⁶National Museum of Natural History, Smithsonian Institution, Washington, D.C., 20560, USA. ⁷College of Agriculture, Auburn University, Auburn, AL, 36849, USA. ✉e-mail: ghopper@agcenter.lsu.edu

amphibian serving as a larval host²⁹. The relationship with fishes has resulted in a large body of research finding fascinating host attraction and attachment strategies that have evolved among species³⁰.

Increasing anthropogenic pressures are changing the structure and function of riverscapes, exacerbating threats to mussel populations and assemblages^{31,32} and elevating the urgency to develop effective tools to identify species responses to environmental change³³. Many mussel traits were considered critical to the management and conservation of freshwater mussels as early as the 1850s and became the focus of intense study as the pearl button industry began and increased dramatically with creation of the United States of America's federally funded Fairport Biological Station (Iowa)³⁴. The contributions of the Fairport Biological station set the stage for others to address habitat requirements, life history traits, and feeding ecology of mussels^{31,35–37}. More recently, conservation agencies such as the United States Fish and Wildlife Service are increasingly incorporating trait data in Species Status Assessments to inform both vulnerability and future adaptability of petitioned species³⁸. Given this rich history and increasing wealth of mussel trait and ecological data, a synthesis and standardization of mussel trait data is needed to inform future studies of mussel ecology, management, and conservation.

Here, we accumulated an extensive functional trait, geographic range, and genetic data availability dataset by reviewing accessible literature, databases, and museum collections. Functional traits represent measurable ecological, life-history, morphological, physiological, and behavioural expressions of individuals without reference to environmental context³⁹. However, we include information that exists outside of this definition that may inform mussel conservation and management. For example, habitat preferences, such as stream size, velocity, and temperature refer to specific environmental contexts and do not fit this definition and may more accurately be referred to as “attributes”. We use the term “trait” in this dataset to reflect any data type associated with a species in a species X trait data matrix⁴⁰.

We compiled this information into a dataset the expectation that our efforts will (1) guide future large scale projects involving trait-based analysis; (2) act as a foundation that research and conservation projects in the future can build on; (3) and prompt discussion about trait integration and expose knowledge gaps. We acknowledge the dataset is incomplete as mussel traits remain understudied. Our intention is to motivate research on the ecology of this less studied and imperilled taxonomic group. The dataset will be useful to researchers engaged in multispecies, regional, and continental -scale analyses of freshwater mussels of the US that could be expanded more broadly by future efforts. Here, we (1) present the dataset and explain the selection process of data for inclusion; (2) interpret the traits and associated data; (3) define the trait states; (4) explain data formats, storage, retrieval, and limitations. We anticipate the identification of knowledge gaps to fuel future research of these understudied traits.

Methods

To address gaps in knowledge surrounding mussel traits, we reviewed numerous sources to collate data. A list of species from within the conterminous US (including Alaska) was generated based on the current checklist assembled by the Freshwater Mollusk Conservation Society (FMCS)⁴¹ and information was sourced primarily from the literature using Google Scholar searches of keywords including trait and scientific names, and by searching existing databases^{35,42–45}. Altogether, we consulted >450 peer-reviewed articles, government reports, book chapters, and databases. We often found significant amounts of information for species documented only in state or regional accounts, such as “Freshwater Mussels of...” books or in non-refereed sources such as theses and dissertations, agency web sites or technical reports. Moreover, regional accounts of species cite each other without consistently acknowledging their sources, making it difficult to identify the original source. For each *scientificName* and each *traitName* we have identified a *traitValue* and its associated reference. Most data were extracted from values listed in tables and qualitative descriptions in text. Details relating to data for a particular *traitName* can be found in the description of each *traitName*. The final dataset of traits is represented as a species by traits matrix that includes traits commonly reported in or inferred from species accounts. All references sourced for the dataset generation are included in the `SHELDreferencesList2023AUG.xlsx`. This file contains the sources for individual entries in the dataset.

Species in the dataset. Our species list (*scientificName*) follows the recent recommendation of the Names of Freshwater Mollusks Subcommittee of Freshwater Mollusk Conservation Society⁴¹. Therefore, traits are reported for 301 species of freshwater mussels found in the US (2 families, 7 tribes, 59 genera). Freshwater mussels are here defined as US freshwater mussels from the order Unionoida.

Shell morphology. Shell size measurements include maximum reported lengths (mm), and mean lengths. When a species mean was reported in multiple studies, we averaged them. Species are considered sculptured if the shell exhibits knobs, pustules, spines, corrugations, or undulations. We classified ((true or false) species sculpturing using photographs from species descriptions in taxonomic keys^{42,43,46,47}. We included museum catalogue numbers of photographed shell material when available.

Reproductive ecology. Fecundity refers to mean number of eggs produced by a single female in one brood. To our knowledge, no studies strictly evaluated the production of multiple broods per year. Fecundity data were gathered by the original authors by dissecting gravid females during the peak of brooding and counting the number of developing eggs in gills of an individual that contained eggs (i.e., charged gills). Next, we conform to the well-established terminology and refer to two dominant brooding strategies in unionids as long-term species (bradytic) that continue to brood larvae after they are infectious and, short-term species (tachytic) do not⁴⁸. Confusion surrounding brooding strategies in the Margaritiferidae⁴⁸ required us to score all margaritiferid species as “unknown”. We described the marsupium (gills used to brood the larvae) by referring to the outer gills (ectobranchy), or all four (tetrageny). Direct development of larvae has been observed in a few species in the US,

but we did not include this trait because it has been paid little attention. Hermaphroditism describes whether hermaphroditic individuals have been observed in populations (true) or not (false).

Age and growth. Age at maturation refers to the earliest report in years of viable gametes in the species. We coded this trait as 0 if reproductively mature individuals were reported within their first year of life. Longevity refers to the life span in years. Growth rates are reported as a separate spreadsheet and include all values available from the literature (SHELDgrowthRates2023AUG.csv). In most cases, Von Bertalanffy's K was used to estimate growth rates (K) and does not represent true growth rates. Values of K are dimensionless parameter that relates how quickly the growth curve becomes asymptotic. Because methods for calculating growth rates can vary, it may be necessary to view the original sources when making intra- and inter-specific comparisons.

Larval morphology. Larvae description includes measurements, shapes, and possession of hook-like structures based on Hoggarth (1999)⁴⁹. The assignments are mutually exclusive, and each trait is discrete. The original accounts were determined by removing larvae from the marsupia of preserved female unionids in museum collections⁴⁹. Larvae representing 30 genera were described while viewed under a scanning electron microscope (SEM). In the event that SEM images of larvae were available for a species, but a shape was not defined in the article or report we matched the SEM image to previously published illustrations^{46,49}. Definitions of the recorded shape follow those of early efforts⁴⁹. Other larval traits included in the dataset are maximum length and maximum height (µm).

Host data and host infection strategy. We used the Freshwater Mussel Host Database (<https://mollusk.inhs.illinois.edu/57-2/>; accessed January–December 2020) to collate available data on fish host use (primary or secondary) either from laboratory or natural transformations for all unionid species⁴⁴. It was necessary to supplement with references from the literature because not all mussel species are represented in the Freshwater Mussel Host Database. We included only fish species that are recognized by the American Fisheries Society (AFS)⁵⁰ and thus limit the dataset to 247 fish species tested as hosts for at least one unionid mussel species. Next, we created a fish species by mussel species association matrix with 1 indicating the fish species served as a host and 0 indicating the fish is not known to serve as a host and summed each mussel species column. We followed a similar procedure for fish genera and families to count the number of host genera and families, respectively, reported to serve as larval hosts. The matrix can be found in SHELDfishMusselMatrix2023AUG.csv. Finally, we categorized the primary host infection strategy used for each mussel species following previously established classifications³⁴. These include broadcasting, female sacrifice, conglutinate, and mantle lure strategies.

Thermal tolerance and upper lethal limits. We used data compiled during a systematic literature review of lethal thermal tolerance upper lethal limits (ULL) for unionid mussels in US faunal region⁴⁵. All thermal tolerance data is categorized by life stage and species and when available includes measurements of lethal temperature where 5% of the population will experience mortality (LT05), median lethal temperature where 50% of the population will experience mortality (LT50), and critical thermal maximum (CTMax). When one measurement of a test type of ULL has been reported in the literature for a single species and life stage, that value is reported as the mean. If multiple estimates of ULL were found in the literature for a test type, species, and life stage combination, we reported the minimum, maximum, and the mean ULLs for that species in the dataset.

All thermal tolerance data must be interpreted with caution and consideration of the test conditions. Details of confidence intervals, acclimation temperature, test duration, or CTMax rate of change surrounding each of the ULL measurements reported in this dataset can be traced to the original references and a review of thermal tolerances among freshwater mussels⁴⁵.

Habitat, stream characteristics, and geographic distribution. This section consists of trait sub-categories describing species associations with substrate type, lentic and lotic systems collated from the literature. Twenty-five habitat associations are included in this section, including seven substrate classes. In some cases, habitats are described based on similarities to habitats where shell material or freshly dead individuals were collected. Particle size classification are provided to facilitate conversion to numerical values⁵¹. Pfeiffer *et al.*⁵² recently used digitized natural history collections data to estimate various occurrence-derived species attributes; 18 of those attributes are included here.

We also estimated mussel biogeography by recording species presence/absence in recognized freshwater mussel faunal zones⁵³. Specifically, we created shapefiles for each faunal zone by aggregation of Hydrological Unit Code (HUC) at the 10-digit level (HUC10's) in the *sf* R Package⁵⁴ and used georeferenced occurrences to estimate faunal region occupation in the *CoordinateCleaner* R package⁵⁵. We considered a species present in a region if it had >2 occurrences, questionable if 1–2 occurrences, and absent if no occurrences. See https://github.com/seanmkeogh/SHELD_biogeo for R script and associated shapefiles.

Conservation status. We reported conservation status of species (i.e., endangered, threatened, or petitioned/candidate reported on the US Federal Register) searched September⁵⁶, 2022) U.S. Fish and Wildlife Services Species Data Explorer⁵⁶. We also included NatureServe's global (range-wide) conservation status ranks (G-ranks; <https://explorer.natureserve.org/>; searched January 1–October 1, 2022). NatureServe conservation status ranks are a valuable complement to legal status designations assigned by government agencies such as the US Fish and Wildlife Service and the National Marine Fisheries Service in administering the US Endangered Species Act (ESA).

Genetics. To investigate the distribution of genetic data availability and identify data gaps across US freshwater mussels, we collated information on genetic information at the species-level. We focused on the availability of

traditional low-throughput: DNA sequence data from regions of the mitochondrial genome, the nuclear genome, multilocus microsatellite genotyping data, and high-throughput sequencing data. To gather information about available mitochondrial and nuclear DNA sequence data for each species we searched the GENBANK sequence database from the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/genbank/). The principal approach for data discovery was to search the taxonomy database within GENBANK for the genus + species names (including known taxonomic synonyms as needed) and scoring species with data as “YES” if mitochondrial or nuclear sequences were available. If additional data types were present beyond single gene sequences for a species (whole mitochondrial genome, expressed sequence tags, RNA/transcriptome sequences, whole genome) in GENBANK or REFSEQ databases these were also noted. Because no centralized databases exist for microsatellite genotyping data, the availability of microsatellites for each species was collated using literature searches performed using Google Scholar (scholar.google.com). Search terms included “genus” AND “microsatellites”. Studies with microsatellite primer discovery (in the absence of population specific diversity and differentiation estimates) were also included. DOI information for publications was also recorded.

To aggregate information on the availability of high throughput sequencing data, we searched the National Center for Biotechnology Information Sequence Read Archive (SRA) for each species (<https://www.ncbi.nlm.nih.gov/sra>, searched between October 5–10, 2022). For SRA data, values for each species include “NO” if no data were detected, “WGS” if whole genome shotgun sequencing was available “RNA” if RNA sequencing or transcriptome data was available, “RAD” if Restriction Site Associated DNA sequencing (any method) was available, “AHE” if Anchored Hybrid Enrichment or other hybrid capture data was available, and “Microbiome16S” if 16S amplicon sequencing of the bacterial community was available. Multiple values are separated by colons.

Data Records

The dataset is available at Figshare⁴⁰.

First file. Identity: SHELDSpeciesTraitMatrix2023AUG.xlsx

Size: 215 KB

Format and storage mode: xlsx

Contents: This file contains the raw data with all *trait* Values and is considered the final dataset

Variable information:

scientificNameAuthority: Scientific name used under the Linnaean classification system with taxonomic authority of the scientific name and date published^{41,57,58}

scientificName: Scientific name used under the Linnaean classification system. The contents of this column can be used as a key between the spreadsheets included in this dataset.

family: Family recognized by FMCS founded on Turgeon (1998), updated by Williams *et al.* (2017) and then updated during FMCS 2021 as recommend by the Names of Freshwater Mollusk Committee.

subFamily: Subfamily identification recognized by FMCS founded on Turgeon (1998), updated by Williams *et al.* (2017) and then updated as of FMCS 2021 as recommend by the Names of Freshwater Mollusk Committee.

tribe: Tribe identification per Pfeiffer *et al.* (2019). Name recognized by FMCS founded on Turgeon (1998), updated by Williams *et al.* (2017) and then updated as of FMCS 2021 as recommend by the Names of Freshwater Mollusk Committee.

genus: Generic name recognized by FMCS founded on Turgeon (1998), updated by Williams *et al.* (2017) and then updated as of FMCS 2021 as recommend by the Names of Freshwater Mollusk Committee. These refer to those used by the Linnaean classification system.

species: Specific epithet recognized by FMCS founded on Turgeon (1998), updated by Williams *et al.* (2017) and then as of FMCS 2021 as recommend by the Names of Freshwater Mollusk Committee. These refer to those used by the Linnaean classification system. These names were used as search terms in some cases during the data retrieval phase.

commonName: Common names recognized by FMCS founded on Williams *et al.* (2017) and updated as of FMCS 2021 as recommend by the Names of Freshwater Mollusk Committee. These names were used as search terms in some cases during the data retrieval phase.

tsn: Taxonomic Serial Number from Integrated Taxonomic Information System. Represent a persistent identifier to facilitate tracing name changes in the future. May not reflect all name changes accepted by FMCS 2021. NA- no data of this classification available for the species.

natureServeElement: Unique identifier for NatureServe queries.

meanLength: Mean of shell lengths. Measured across the anterior-posterior plane of the shell. NA- no data of this classification available for the species.

maxLength: Maximum shell length (length description) in millimeters. Some records may estimate asymptotic length. We assumed that the difference between maximum observed length and asymptotic length is negligible. Measured across the anterior-posterior plane of the shell. NA- no data of this classification available for the species.

shellSculpture: Species are considered sculptured (true) if the shell exhibits knobs, pustules, spines, corrugations, or undulations. Mussels without the aforementioned sculpturing were scored “false” in the dataset This binary classification (true or false) was based on photographed material found on MolluscaBase, Musselp⁴³, taxonomic keys, and species descriptions^{46,47,59}.

larvalDescription:

Descriptions that reflect two discrete larval characteristics. The term “hooked” is applied indiscriminately to independently derived structures that function similarly, to grab onto the host fish. Hookless refers to the absence of such a structure. Here, the larvae are hooked or hookless combined with terms describing their shape based on Hoggarth⁶⁰.

hookless_ligulate: Without hook. “Axe-head shape”, much higher than long. Dorsal margin short and straight. Anterior and posterior margins are parallel to about 80% from dorsal to ventral becoming evenly curved to point of maximum lateral inflation distant from the dorsal margin. Ventral margin only slightly curved.

hookless_subligulate: Without hook. Much higher than long. Dorsal margin slightly curved. Anterior and posterior margins straight to slightly incurved dorsally. Margins become straight again at about 40% from dorsal margin, but slightly divergent. Maximum lateral inflation about 80% from dorsal margin.

hookless_subspatulate: Without hook. Dorsal margin straight. Lateral margins straight, divergent dorsally, ventrally parallel; ventral margin gently curved.

hookless_subelliptical: Without hook. Dorsal margin straight. Lateral margins gently curved but unequal. Maximum inflation of posterior margin at about 70% from dorsal to ventral. Maximum inflation of the anterior margin at about 40% from dorsal margin. Ventral margin is narrowly rounded.

hookless_elongate_oval: Without hook. Dorsal margin slightly curved outward. Lateral and ventral valve margins more or less round in outline, with maximum inflation of both side margins at about 50% from dorsal to ventral.

hookless_subrotund: Without hook. Dorsal margin slightly curved outward. Lateral and ventral valve margins not quite round in outline. Maximum inflation of side margins about 50% from dorsal to ventral.

hookless_subtriangular: Without hook. Dorsal margin is straight; posterior margin gently and evenly curved; anterior margin curved broadly to is maximum inflation at ~70% from dorsal to ventral. Ventral terminus is broadly pointed.

hookless_fabelliform: Without hook. “bean-shaped”, much longer than high having a straight dorsal margin, narrowly rounded anterior and posterior margins, broadly curved ventral margin. Anterior and posterior margins equal, with their points of maximum inflation at about 30% from dorsal margin.

hooked_ligulate: Possessing a hook. “Ax-head shape”, much higher than long. Dorsal margin short and straight. Anterior and posterior margins are parallel to about 80% from dorsal to ventral becoming evenly curved to point of maximum lateral inflation distant from the dorsal margin. Ventral margin only slightly curved.

hooked_subelliptical: Possessing a hook. Dorsal margin straight. Lateral margins gently curved but unequal. Maximum inflation of posterior margin at about 70% from dorsal to ventral. Maximum inflation of the anterior margin at about 40% from dorsal margin. Ventral margin is narrowly rounded.

hooked_subtriangular: Possessing a hook. Dorsal margin is straight; posterior margin gently and evenly curved; anterior margin curved broadly to is maximum inflation at ~70% from dorsal to ventral. Ventral terminus is broadly pointed.

hooked_pyriiform: Possessing a hook. Dorsal margin is straight. Posterior margin broadly curving, with maximum inflation at about 30% from dorsal to ventral. Anterior margin more broadly curved than posterior margin. Maximum anterior inflation at about 40% from dorsal margin. Lateral margins slightly incurved ventrally, producing a narrowly rounded ventral terminus located 40% from posterior to anterior.

hooked_depressed_pyriiform: Possessing a hook. Posterior margin is broadly curved; anterior margin is rounded. Lateral margins meet at a narrowly rounded, nipple-like ventral terminus located about 50% from posterior to anterior.

NA: no data of this classification available for the species.

larvalHeight: Mean height of larvae in micrometers (μm) measured perpendicular to the hinge. Max height is given if no mean is reported. NA- no data of this classification available for the species.

larvalLength: Mean length of larvae in micrometers (μm) measured parallel to the hinge. NA- no data of this classification available for the species.

matureAge: Mean, median, or modal age at maturity in years for females. Where different ages at maturity were gathered for distinct populations, the best supported records were averaged. Species that are known to mature prior to age 1 are defined as 0. Male maturity age was accepted where female data were not available. NA- no data of this classification available for the species.

maxAge: Longevity in years based on life in the wild wherever available. Where not indicated, the record was assumed to be from the wild. If wild records were not known, records for captive individuals were considered. NA- no data of this classification available for the species.

growthRate: Binary indicator (true or false) of the availability of growth rate data for each species. Literature sources often evaluated growth rates inconsistently. We dealt with this by including a separate sheet of growth rates, SHELDgrowthRates2023AUG.csv. That sheet details growth rates, sample sizes, location of the study, and the data source.

fecundity: Refers to mean number of eggs produced by a single female in one brood. There were not studies strictly evaluating the production of multiple clutches per year. Fecundity data were gathered by the original authors by dissecting gravid females during the peak of brooding and counting the number of developing eggs in all charged gills of an individual. NA- no data of this classification available for the species.

brood: long_term indicates long (bradyctictic)- or short_term indicates short term (tachyctictic) brooding strategy; long term brooders hold larvae from fall (September to December) to spring (March to June) or summer (June to September), whereas short term brooders hold larvae from winter to spring or summer. NA- no data of this classification available for the species.

marsupialGills: Indicates that brooding of larvae occurs in outer gills only (ectobrancy) or all four gills (tetrageny). This trait is a distinct characteristic of some genera. NA- no data of this classification available for the species.

hermaphrodite: Indicates reports of hermaphroditic populations (true). If no hermaphroditism has been reported the species is assumed to have male and female sexes (false). If this trait has not been explicitly reported or investigated, we report as NA.

hostInfectionStrategy: Category of primary host infection strategy. For example, some species display mantle lures, but also release secondary conglutinates for a time. Categories are defined as:

broadcast-Broadcasting involves larvae release whereby host fish are encountered by chance; no apparent adaptations to attract hosts; Passive entanglement and release of free larvae are not separated.

sacrifice-Species that specialize on molluscivorous host fish; high fecundity, small body sizes, high male:female ratios, and high larvae infection of molluscivorous fishes support this; May also broadcast into sediment to increase chance of infection via suction feeding by host fish.

conglutinate-Release larvae in discrete clusters, that can resemble fish prey items that facilitate host infection; Can be further categorized by functional attraction of specialists. 1) Pelagic, 2) Demersal, 3) Mucoid, 4.) superconglutinate.

mantle lure-Lures displayed by gravid females that mimic a variety of fish prey items; Can be further categorized large (Tribe: Lampsilini) and cryptic lures (Tribe: Lampsilini) or mantle magazines (Tribe: Quadrulini).

kinetic lure-The conglutinates consist of milky-white, rod-shaped (3–7 mm by 0.3–0.4 mm), composed of homogeneous, adhesive mucous within which 1–15 larvae are embedded. The conglutinates swell upon release from the hypertonic gill solution into hypotonic ambient conditions eliciting a writhing and quivering action until equilibrium is reached. Swelling forced larvae to the exterior of the conglutinate where they will gape their valves while remaining tethered to the conglutinate by a larval thread.

NA-no data for host infection strategies specific to this species.

nHostSpecies: Number of fish species reported to serve as larval hosts (primary or secondary) either from laboratory or natural infestations⁴⁴ NA- no data available for the species.

nHostGenera: Number of fish genera reported to serve as larval hosts (primary or secondary) either from laboratory or natural infestations⁴⁴ NA- no data available for the species.

nHostFamily: Number of fish families reported to serve as larval hosts (primary or secondary) either from laboratory or natural infestations⁴⁴. NA- no data available for the species.

mud: Indicates associations with (true) substrate type or not (false) Mud substrate; particle size <2 mm. NA- no data available for the species.

sand: Indicates associations with (true) substrate type or not (false) Sand substrate; particle size <2 mm. NA- no data available for the species.

claySilt: Indicates associations with (true) substrate type or not (false) Clay or silt substrate; particle size <2 mm. NA- no data available for the species.

gravel: Indicates associations with (true) substrate type or not (false) Gravel substrate; particle size = 2–16 mm. NA- no data available for the species.

pebble: Indicates associations with (true) substrate type or not (false). Pebble substrate; particle size = 16–64 mm. NA – no data available for the species.

cobble: Indicates associations with (true) substrate type or not (false) Cobble substrate; particle size = 65–256 mm. NA- no data available for the species.

boulder: Indicates associations with (true) substrate type or not (false) Boulder substrate; particle size >256. NA- no data available for the species.

bedrock: Indicates associations with (true) substrate type or not (false) Bedrock substrate; smooth surface. NA- no data available for the species.

macrophyte: Indicates associations with (true) aquatic plants or not (false). Na-no data available for the species.

lotic: Indicates associations with (true) habitat type or not (false). Lotic preference. NA- no data available for the species.

lentic: Indicates associations with (true) habitat type or not (false). Lentic preference; includes flood plain lakes and reservoirs. NA- no data available for the species.

largeRiver: Indicates associations with (true) habitat type or not (false). Medium to large rivers as referred in source. NA- no data available for the species.

smallRiver: Indicates associations with (true) habitat type or not (false). Streams to small rivers as referred in source. NA- no data available for the species.

creek: Indicates associations with (true) habitat type or not (false). Creek is used in the species habitat description in source. NA- no data available for the species.

lacustrine: Indicates associations with (true) habitat type or not (false). Preferences for natural lake systems; excludes reservoirs. NA- no data available for the species.

slowCurrent: Indicates associations with (true) habitat type or not (false). Slow current speed as described in source. NA- no data available for the species.

moderateCurrent: Indicates associations with (true) habitat type or not (false). Moderate current speed as described in source. NA- no data available for the species.

fastCurrent: Indicates associations with (true) habitat type or not (false). Fast current speed as described in source. NA- no data available for the species.

larvalMinLT05: Minimum LT05, or lethal temperature where 5% of the population will experience mortality, reported for larvae when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. NA- no data available for the species.

larvalMaxLT50: Maximum reported LT05 reported for larvae when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species,

which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

larvalMeanLT05: Mean of all reported LT05s reported for larvae when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

larvalMinLT50: Minimum reported LT50 reported for larvae when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

larvalMaxLT50: Maximum reported LT50 reported for larvae when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

larvalMeanLT50: Mean of all reported LT50s reported for larvae when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

juvenileMinLT05: Minimum reported LT05 reported for juveniles when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

juvenileMaxLT05: Maximum reported LT05 reported for juveniles when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

juvenileMeanLT05: Mean of all reported LT05s reported for juveniles when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

juvenileMinLT50: Minimum reported LT50 reported for juveniles when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

juvenileMaxLT50: Maximum reported LT50 reported for juveniles when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

juvenileMeanLT50: Mean of all reported LT50s reported for juveniles when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

adultMinLT05: Minimum reported LT05 reported for adults when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

adultMaxLT05: Maximum reported LT05 reported for adults when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

adultMeanLT05: Mean of all reported LT05s reported for adults when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

adultMinLT50: Minimum reported LT50 reported for adults, when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

adultMaxLT50: Maximum reported LT50 reported for adults when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

adultMeanLT50: Mean of all reported LT50s reported for adults when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT50 is the lethal temperature where 50% of the population will experience mortality.

adultCTMax: Mean of all reported critical thermal maximum values for adults of each species. NA- no data available for the species.

federalStatus: Listing under the United States of America Endangered Species Act. Categories are defined as: endangered- any species which is in danger of extinction throughout all or a significant portion of its range. threatened-any species which is likely to become an endangered species within the foreseeable future throughout all or a significant portion of its range.

petitioned_or_candidate- species for which the U.S. Fish and Wildlife Service (FWS) has sufficient information on their biological status and threats to propose them as endangered or threatened under the

Endangered Species Act (ESA), but for which development of a proposed listing regulation is precluded by other higher priority listing activities.

not_listed- Species is not currently listed under the Endangered Species Act (ESA). Some species may be not-listed because there has not been a petition for listing yet, which could be due to insufficient information or oversight. Some species may be not-listed because they have been delisted. Species can be delisted if the species has recovered to the point that is no longer needs protection under ESA. The original information warranting listing is demonstrated as inaccurate or new information suggests the species is not in need of protection. Extinct species can be scored “not_listed” because once they are declared extinct they are “delisted” (=not_listed) and many species went extinct prior to listing (=not_listed).

conservationStatus: A binary classification system that column used to account for where the USFWS designations incompletely captured decline.

listed- if its ESA Listing Status was Endangered, Threatened, or Extinct according to the USFWS Species Data Explorer (<https://ecos.fws.gov/ecp/report>) accessed on September 13, 2022. Many species that are widely considered to be extinct³⁴ were also categorized this way.

Non-listed-species included all other ESA Listing Status categories (i.e., Proposed Endangered, Proposed Threatened, Resolved Taxon, Species of Concern, Status Undefined, and Under Review) and all unassessed species.

gRank: Global Rank as defined on NatureServe. Categories are defined as:

GX-(Presumed Extinct) — Not located despite intensive searches and virtually no likelihood of rediscovery.

GH-(Possibly Extinct) Known from only historical occurrences but still some hope of rediscovery.

G1-(Critically imperiled) At very high risk of extinction or elimination due to very restricted range, very few populations or occurrences, very steep declines, very severe threats, or other factors.

G2-(Imperiled) At high risk of extinction or elimination due to restricted range, few populations or occurrences, steep declines, severe threats, or other factors.

G3-At moderate risk of extinction or elimination due to a fairly restricted range, relatively few populations or occurrences, recent and widespread declines, threats, or other factors. (Vulnerable).

G4-(Apparently secure) At fairly low risk of extinction or elimination due to an extensive range and/or many populations or occurrences, but with possible cause for some concern as a result of local recent declines, threats, or other factors.

G5-(Secure) At very low risk of extinction or elimination due to a very extensive range, abundant population or occurrences, and little to no concern from declines or threats.

GU-(Unrankable) Currently unrankable due to lack of information or due to substantially conflicting information about status or trends.

GNR-(Unranked) Global rank not yet assessed.

mitochondrialsequences: Indicates presence (true) or absence (false) of mitochondrial sequences on GENBANK. NA- no data of this classification available for the species.

nuclearSequencesInclRNAexcMSATclones: Indicates presence (true) or absence (false) of nuclear sequences on GENBANK.

otherGenbankDataTypes: If additional data types were present beyond single gene sequences for a species (whole mitochondrial genome, expressed sequence tags, RNA/transcriptome sequences, whole genome) in GENBANK or REFSEQ databases these were also noted. NA- no data of this classification available for the species.

microsatellites: Indicated presence (true) or absences (false) of microsatellite data.

sra: Indicated presence or absences for high throughput sequencing data in the National Center for Biotechnology Information Sequence Read Archive; NA = no data; WGS = whole genome shotgun sequencing; RNA = RNA sequencing or transcriptome data were available; RAD = restriction site associated DNA sequencing by any method; AHE = Anchored hybrid enrichment or other hybrid capture data; Microbiome 16S = availability of 16S amplicon sequencing of bacterial community.

noccs: Number of georeferenced occurrences from 45 museum collections in the United States that were in the species range, were not duplicated, were geolocated within the same state listed on the lot, and were within the National Hydrography Dataset. These were used to calculate geographic trait values. NA- no data available for the species.

nhuc8: Number of HUC8s that contained occurrences described in noccs. NA- no data available for the species.

aoHUC8sqkm: Area of occupancy, measured as the total area of the HUC8s that included at least one occurrence point. NA- no data available for the species.

totalAOOwithyear: Area of occupancy, measured identically to above but only for records that had temporal (year) data. NA- no data available for the species.

minLongitude: Minimum decimal degree longitude of any occurrence of the species. NA- no data available for the species.

minLatitude: Minimum decimal degree latitude of any occurrence of the species. NA- no data available for the species.

maxLongitude: Maximum decimal degree longitude of any occurrence of the species. NA- no data available for the species.

maxLatitude: Maximum decimal degree latitude of any occurrence of the species. NA- no data available for the species.

midLongitude: Decimal degree longitude of the center of the convex polygon that includes all occurrences of the species in Albers projection. NA- no data available for the species.

midLatitude: Decimal degree latitude of the center of the convex polygon that includes all occurrences of the species in Albers projection. NA- no data available for the species.

modeStreamOrder: Mode Strahler stream order of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. NA- no data available for the species.

iqrStreamOrder: Interquartile range of Strahler stream order of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. NA- no data available for the species.

medianStreamSlope: Median stream slope (slope of flowline (meters/meters) based on smoothed elevations) of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. NA- no data available for the species.

iqrStreamSlope: Interquartile range of stream slope (slope of flowline (meters/meters) based on smoothed elevations) of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. NA- no data available for the species.

medianQAMA: Median of mean annual discharge (flow from runoff (cubic feet per second)) of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. See NHDPlusV2 documents⁶¹ for definition of QAMA. NA- no data available for the species.

iqrQAMA: Interquartile range of mean annual discharge (flow from runoff (cubic feet per second)) of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. See NHDPlusV2 documents⁶¹ for definition of QAMA. NA- no data available for the species.

medianVAMA: Median annual velocity (for QAMA (feet per second)) of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. See NHDPlusV2 documents for definition of VAMA. NA- no data available for the species.

iqrVAMA: Interquartile range of the mean annual velocity (for QAMA (feet per second)) of occurrences of the species that were snapped to NHDPlusV2 flowlines with stream order >1. See NHDPlusV2 documents for definition of VAMA. NA- no data available for the species.

nStreamCharacteristicLots: Number of occurrences used to estimate the stream characteristic traits. Only occurrences within the NHDPlusV2 dataset (which does not include Alaska) and with non-negative flow velocity, discharge, and stream slope were used and therefore this value may be less than nocc. NA- no data available for the species.

mississippiEmbayment: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). All biogeographic definitions follow Haag (2009)⁵³. Encompasses the lower Mississippi River and all tributaries below the mouth of the Ohio River, including most the Lower Red River system and the Atchafalaya basin, and the Mermentau, with flow into the Gulf of Mexico just east of the Atchafalaya basin.

upperMississippi: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes the entire Mississippi River system upstream of the Ohio River, excluding the Missouri River system except for the southern tributaries of the lower Missouri River.

ohioan: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes the Ohio River and all of its tributaries except the upper two-thirds of the Cumberland and Tennessee River systems.

tennesseeCumberland: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes the upper two-thirds of the Tennessee and Cumberland River systems. The downstream boundary of the provinces is located downstream of Muscle Shoals in the Tennessee River in northwestern AL and in the Cumberland River near Clarksville, TN. Boundaries are set based on the occurrences of endemic species.

interiorHighlands: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Encompasses two geographically discontinuous regions: 1.) upper White and upper St. Francis systems with the Ozark Plateaus province plus the adjacent Verdigris, Neosho, and Illinois River systems. 2) The upper Ouachita, Kiamichi, and Little river systems, and the Poteau River.

greatPlains: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes all river systems east of the Rocky Mountains, from the upper Red River of Texas and Oklahoma, north to the Nelson-Churchill basins in central Canada.

stLawrenceGreatLakes: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes all five Great Lakes and Lake St. Clair and their watersheds. Also includes St. Lawrence and Ottawa river systems and rivers flowing into James Bay.

westernGulf: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). All rivers flowing into the Gulf of Mexico from the Brazos south to the Rio Grande.

sabineTrinity: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Encompasses rivers of the central Gulf Coast.

pontchartrainPearlPascagoula: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes the Pearl and Pascagoula River systems and all streams that flow into Lakes Pontchartrain and Maurepas.

mobileBasin: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). All rivers flowing into Mobile Bay in the Gulf of Mexico.

escambiaChoctawhatchee: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes the Escambia, Yellow, and Choctawhatchee rivers.

apalachicolan: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Includes the Apalachicola, Ochlockonee, and Suwannee rivers, and Econfina Creek that flows into the Gulf of Mexico.

peninsularFlorida: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Extends from Waccasassa River on the Florida Gulf Coast, south and around the Florida Peninsula, north to and including St. Mary's River.

southernAtlantic: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Satilla River north to the James River.

northernAtlantic: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). Extends from the York River system north of Chesapeake Bay to Newfoundland.

pacific: We considered a species present (present) in this region if it had >2 occurrences, questionable if 1–2 occurrences (questionable), and absent if no occurrences (absent). All rivers in the United States of America flowing into the Pacific Ocean; Headwaters of the of the upper Missouri River above Great Falls.

Second file. Identity: SHELDmusselFishData2023AUG.csv

Size: 152.4KB

Format and storage mode: CSV

Contents: Long format data set with known fish hosts for each mussel species. The first column is the mussel scientificName and the last three columns (A–C) are hierarchical Linnaean classifications for fish host families, genera, and species with each row representing a fish host–mussel association.

Variable information:

musselScientificName: Scientific name used under the Linnaean classification system^{41,57,58}.

fishFamily: A list of the family used under the Linnaean classification system of fishes known to serve as larval hosts (primary or secondary) either from laboratory or natural transformations. Names are current with the American Fisheries Society⁵³.

fishGenus: A list of the genera used under the Linnaean classification system of fishes known to serve as larval hosts (primary or secondary) either from laboratory or natural transformations. Names are current with the American Fisheries Society⁵³.

fishSpecies: A list of the specific epithet used under the Linnaean classification system of fishes known to serve as larval hosts (primary or secondary) either from laboratory or natural transformations. Names are current with the American Fisheries Society⁵³.

Third file. Identity: SHELDgrowthRates2023AUG

Size: 31.6 KB

Format and storage mode: CSV

Contents: Contains growth rates for mussel species found in the United States of America. Each row represents a population of a species such that one species may have several rows if growth rates were estimated for several populations.

Variable information:

scientificName: Scientific name used under the Linnaean classification system^{41,53,57}.

growthRate: K value estimated from the von Bertalanffy equation by the original source

sampleSize: number of individuals in the data set.

sexIdentified: the sex of the animals used in the study (male or female).

methodUsed: The method used to estimate growth rate. Typically, vonBertalanffy equations were used. There is one case where absolute growth rate was measured.

location: the geographic location the animals were collected from in the United States of America as reported by the original source. We provide a specific location and the two-letter abbreviation for the state separated by a comma.

reference: A list of references the data for each growth rate was sourced from.

Fourth file. Identity: SHELDreferenceList2023AUG.xlsx

Size: 359.2 KB

Format and storage mode: xlsx

Contents: A complete list of mussel scientificNameAuthority with references for traitName found in columns J to BS in SHELDspeciesTraitMatrix2023AUG.xlsx

Variable information:

traitName: references to the traits in the complete dataset SHELDspeciesTraitMatrix2023AUG.xlsx. The trait levels included here are those we collated data for from the search. Traits are listed below with a brief definition. Complete definitions for each trait and trait subcategory can be found above in the variable information for the first file.

tsn: Taxonomic Serial Number from Integrated Taxonomic Information System. Represent a persistent identifier to facilitate tracing name changes in the future. May not reflect all name changes accepted by FMCS 2021. NA- no data of this classification available for the species.

natureServeElement: Unique identifier for NatureServe queries.

meanLength: Mean of shell lengths. Measured across the anterior-posterior plane of the shell.

maxLength: Maximum shell length (length description) in millimeters. Some records may estimate asymptotic length. We assumed that the difference between maximum observed length and asymptotic length is negligible. Measured across the anterior-posterior plane of the shell.

shellSculpture: Species are considered sculptured if the shell exhibits knobs, pustules, spines, corrugations, or undulations. This binary classification was based on photographs of material from museum lots found on MolluscaBase⁴², Musselp⁴³, and species accounts^{46,47,59}.

larvalDescription: Descriptions that reflect two discrete larval characteristics. The term “hooked” is applied indiscriminately to independently derived structures that function similarly, to grab onto the host fish. Hookless refers to the absence of such a structure. Here, the larvae are hooked or hookless combined with terms describing their shape based on Hoggarth⁴⁹.

larvalHeight: Mean height of larvae in micrometers (μm) measured perpendicular to the hinge. Max height is given if no mean is reported.

larvalLength: Mean length of larvae in micrometers (μm) measured parallel to the hinge.

matureAge: Mean, median, or modal age at maturity in years for females. Where different ages at maturity were gathered for distinct populations, the best supported records were averaged. Species that are known to mature prior to age one are defined as 0. Male maturity age was accepted where female data were not available. NA- no data of this classification available for the species.

maxAge: Longevity in years based on life in the wild wherever available. Where not indicated, the record was assumed to be from the wild. If wild records were not known, records for captive individuals were considered.

growthRate: Binary indicator of the availability of growth rate data for each species. Literature sources often evaluated growth rates inconsistently. We dealt with this by including a separate sheet of growth rates, SHELDgrowthRates2023AUG.csv. That sheet details growth rates, sample sizes, location of the study, and the data source.

fecundity: Refers to mean number of eggs produced by a single female in one brood. There were not studies strictly evaluating the production of multiple clutches per year.

brood: long_term indicates long (bradyctictic)- or short_term indicates short term (tachyctictic) brooding strategy; long term brooders hold larvae from fall to spring or summer, whereas short term brooders hold larvae from winter to spring or summer. NA- no data of this classification available for the species.

marsupialGills: Indicates that brooding of larvae occurs in outer gills only (ectobranchy) or all four gills (tetragony).

hermaphrodite: Indicates reports of hermaphroditic populations.

hostInfectionStrategy: Category of primary host infection strategy. Categories are defined above in the variable information for the first file.

nHostSpecies: Number fish species reported to serve as larval hosts (primary or secondary) either from laboratory or natural infestations⁴⁴.

nHostGenera: Number fish genera reported to serve as larval hosts (primary or secondary) either from laboratory or natural infestations⁴⁴.

nHostFamily: Number of fish families reported to serve as larval hosts (primary or secondary) either from laboratory or natural infestations⁴⁴.

habitat: Combines the trait sub-categories in SHELDspeciesTraitMatrix2023AUG.xlsx. describing species associations with substrate type, lentic and lotic systems collated from the literature. Trait subcategories include sand, claySilt, gravel, cobble, boulder, bedrock, lotic, lentic, largeRiver, smallRiver, creek, lacustrine, slowCurrent, moderateCurrent, and fastCurrent. These subcategories are defined above in the variable information for the first file.

larvalMinLT05: Minimum LT05, or lethal temperature where 5% of the population will experience mortality, reported for larvae when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. NA- no data available for the species.

larvalMaxLT50: Maximum reported LT05 reported for larvae when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

larvalMeanLT05: Mean of all reported LT05s reported for larvae when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

larvalMinLT50: Minimum reported LT50 reported for larvae when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

larvalMaxLT50: Maximum reported LT50 reported for larvae when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

larvalMeanLT50: Mean of all reported LT50s reported for larvae when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

juvenileMinLT05: Minimum reported LT05 reported for juveniles when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

juvenileMaxLT05: Maximum reported LT05 reported for juveniles when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

juvenileMeanLT05: Mean of all reported LT05s reported for juveniles when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

juvenileMinLT50: Minimum reported LT50 reported for juveniles when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

juvenileMaxLT50: Maximum reported LT50 reported for juveniles when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

juvenileMeanLT50: Mean of all reported LT50s reported for juveniles when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

adultMinLT05: Minimum reported LT05 reported for adults when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

adultMaxLT05: Maximum reported LT05 reported for adults when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

adultMeanLT05: Mean of all reported LT05s reported for adults when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT05 is the lethal temperature where 5% of the population will experience mortality. NA- no data available for the species.

adultMinLT50: Minimum reported LT50 reported for adults, when multiple estimates of LT05 are available for that species. If no minimum is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

adultMaxLT50: Maximum reported LT50 reported for adults when multiple estimates of LT05 are available for that species. If no max is reported, that means there is only one estimate of LT05 reported for that species, which is shown in the mean. LT50 is the lethal temperature where 50% of the population will experience mortality. NA- no data available for the species.

adultMeanLT50: Mean of all reported LT50s reported for adults when multiple estimates of LT05 are available for that species OR the single LT05 estimate that has been reported for that species. LT50 is the lethal temperature where 50% of the population will experience mortality.

adultCTMax: Mean of all reported critical thermal maximum values for adults of each species. NA- no data available for the species.

federalStatus: Listing under the United States of America Endangered Species Act.

conservationStatus: A binary classification system that column used to account for where the USFWS designations may incompletely captured decline.

gRank: Global Rank as defined on NatureServe.

mitochondrialSequences: Indicates presence (YES) or absence (NO) of mitochondrial sequences on GENBANK. NA- no data of this classification available for the species.

nuclearSequencesInclRNAexclMSATclones: Indicates presence (YES) or absence (NO) of nuclear sequences on GENBANK.

otherGenbankDataTypes: If additional data types were present beyond single gene sequences for a species (whole mitochondrial genome, expressed sequence tags, RNA/transcriptome sequences, whole genome) in GENBANK or REFSEQ databases these were also noted NA- no data of this classification available for the species.

microsatellites: Indicated presence (YES) or absences (NO) of microsatellite data.

sra: Indicated presence or absences for high throughput sequencing data in the National Center for Biotechnology Information Sequence Read Archive.

referenceID1: Citation or museum catalogue number for the traitName identified. NA - no data available for the species. To derive traitValues for some traitNames, it was necessary to use multiple sources. In such cases, an additional column or two will follow and will be identified as **referenceID2** or **referenceID3**.

referenceID2: Citation or museum catalogue number for the traitName identified. NA - no data available for the species.

referenceID3: Citation or museum catalogue number for the traitName identified. NA - no data available for the species.

Fifth file. Identity: SHELDspeciesStateMatrix.csv

Size: 37.0 KB

Format and storage: CSV

Contents: Association matrix of political boundaries for states in the United States of America and each mussel species. The first column is the name of the state political boundary. The remaining column headers (B-KP) correspond to the scientificName column SHELDspeciesTraitMatrix2023AUG.xlsx listed as the First file. The number one (1) is used to identify the occurrence of each species within the political boundary as indicated by NatureServe.

Variable information:

state: the name of each state political boundary in the United States of America.

Technical Validation

Initial quality assurance and quality control procedures included having two separate authors examine data for each reference prior to including data in the complete dataset SHELDspeciesTraitMatrix2023AUG.xlsx, the host data set SHELDfishMusselMatrix2023AUG.csv, SHELDgrowthRates2023AUG.csv, and SHELDspeciesStateMatrix2023AUG.csv. Upon completion of data compilation, we checked each file for formatting errors. We checked that all text columns contained only relevant text. We checked that all numerical trait columns contained only correctly formatted numeric data input in correct units or NA when appropriate. Next, we completed an outlier analysis of numerical traits to identify potential errors for the complete dataset SHELDspeciesTraitMatrix2023AUG.xlsx. We performed three levels of outlier analysis for every numerical trait: 1) outliers among all species, 2) outliers among members of the same genus, 3) outliers among members of each phylogenetic tribe. We flagged data points as outliers if they were greater than 2.5 standard deviations of the mean of the given trait. If we flagged a data point, we rechecked the original reference. If the data point matched the value in the original reference, we made no change. If the data point did not match the value in the original reference, we either updated it to the correct value or deleted it if the value did not exist in that reference. Overall, 1.7% of the raw continuous data were flagged as outliers across all species, 4.9% were flagged as outliers at the genus level, and 2.2% were flagged as outliers at the tribe level. Of the identified outliers, 6.40% were ultimately corrected or deleted.

Code availability

R code and associated shapefiles for biogeographic determinations can be found at https://github.com/seanmkeogh/SHELD_biogeo.

Received: 23 February 2023; Accepted: 11 October 2023;

Published online: 27 October 2023

References

- McGill, B. J., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends in Ecology & Evolution* **21**, 178–185 (2006).
- Ackerly, D. D. & Cornwell, W. K. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. *Ecol Lett* **10**, 135–145 (2007).
- Messier, J., McGill, B. J. & Lechowicz, M. J. How do traits vary across ecological scales? A case for trait-based ecology. *Ecol Lett* **13**, 838–848 (2010).
- Böhm, M. *et al.* The conservation status of the world's freshwater molluscs. *Hydrobiologia* **848**, 3231–3254 (2021).
- Ferreira-Rodríguez, N. *et al.* Research priorities for freshwater mussel conservation assessment. *Biol Conserv* **231**, 77–87 (2019).
- Lopes-Lima, M. *et al.* Conservation of freshwater bivalves at the global scale: diversity, threats and research needs. *Hydrobiologia* **810**, 1–14 (2018).
- Lopes-Lima, M. *et al.* Conservation status of freshwater mussels in Europe: state of the art and future challenges. *Biological Reviews* **92**, 572–607 (2016).
- Haag, W. R. & Williams, J. D. Biodiversity on the brink: an assessment of conservation strategies for North American freshwater mussels. *Hydrobiologia* **735**, 45–60 (2013).
- Schwalb, A. N., Morris, T. J., Mandrak, N. E. & Cottenie, K. Distribution of unionid freshwater mussels depends on the distribution of host fishes on a regional scale. *Divers Distrib* **19**, 446–454 (2012).
- Daniel, W. M. & Brown, K. M. The role of life history and behavior in explaining unionid mussel distributions. *Hydrobiologia* **734**, 57–68 (2014).
- Mitchell, Z. A., Burlakova, L. E., Karatayev, A. Y. & Schwalb, A. N. Changes in community composition of riverine mussels after a severe drought depend on local conditions: a comparative study in four tributaries of a subtropical river. *Hydrobiologia* **848**, 3015–3029 (2019).
- Mitchell, Z. A., McGuire, J., Abel, J., Hernandez, B. A. & Schwalb, A. N. Move on or take the heat: Can life history strategies of freshwater mussels predict their physiological and behavioural responses to drought and dewatering? *Freshw Biol* **63**, 1579–1591 (2018).
- Randklev, C. R., Hart, M. A., Khan, J. M., Tsakiris, E. T. & Robertson, C. R. Hydraulic requirements of freshwater mussels (Unionidae) and a conceptual framework for how they respond to high flows. *Ecosphere* **10** (2019).
- Khan, J. M., Dudding, J., Hart, M., Tsakiris, E. & Randklev, C. R. Linking life history strategies and historical baseline information shows effects of altered flow regimes and impoundments on freshwater mussel assemblages. *Freshw Biol* **65**, 1950–1961 (2020).
- Khan, J. M. *et al.* Linking flow and upper thermal limits of freshwater mussels to inform environmental flow benchmarks. *Freshw Biol* **65**, 2037–2052 (2020).
- Atkinson, C. L., Julian, J. P. & Vaughn, C. C. Species and function lost: Role of drought in structuring stream communities. *Biol Conserv* **176**, 30–38 (2014).

17. Hamstead, B. A., Hartfield, P. D., Jones, R. L. & Gangloff, M. M. Changes to freshwater mussel assemblages after 25 years of impoundment and river habitat fragmentation. *Aquat Conserv* **29**, 2162–2175 (2019).
18. Haag, W. R., Berg, D. J., Garton, D. W. & Farris, J. L. Reduced Survival and Fitness in Native Bivalves in Response to Fouling by the Introduced Zebra Mussel (*Dreissena polymorpha*) in Western Lake Erie. *Canadian Journal of Fisheries and Aquatic Sciences* **50**, 13–19 (1993).
19. Atkinson, C. L., Opsahl, S. P., Covich, A. P., Golladay, S. W. & Conner, L. M. Stable isotopic signatures, tissue stoichiometry, and nutrient cycling (C and N) of native and invasive freshwater bivalves. *J North Am Benthol Soc* **29**, 496–505 (2010).
20. Vaughn, C. C. Biodiversity Losses and Ecosystem Function in Freshwaters: Emerging Conclusions and Research Directions. *Bioscience* **60**, 25–35 (2010).
21. Galbraith, H. S. & Vaughn, C. C. Effects of reservoir management on abundance, condition, parasitism and reproductive traits of downstream mussels. *River Res Appl* **27**, 193–201 (2011).
22. Vaughn, C. C. Life history traits and abundance can predict local colonisation and extinction rates of freshwater mussels. *Freshw Biol* **57**, 982–992 (2012).
23. Schwalb, A. N., Morris, T. J. & Cottenie, K. Dispersal abilities of riverine freshwater mussels influence metacommunity structure. *Freshw Biol* **60**, 911–921 (2015).
24. Vaughn, C. C., Gido, K. B. & Spooner, D. E. Ecosystem Processes Performed by Unionid Mussels in Stream Mesocosms: Species Roles and Effects of Abundance. *Hydrobiologia* **527**, 35–47 (2004).
25. Spooner, D. E. & Vaughn, C. C. Species' traits and environmental gradients interact to govern primary production in freshwater mussel communities. *Oikos* **121**, 403–416 (2011).
26. Spooner, D. E. & Vaughn, C. C. A trait-based approach to species' roles in stream ecosystems: climate change, community structure, and material cycling. *Oecologia* **158**, 307–317 (2008).
27. Atkinson, C. L. & Vaughn, C. C. Biogeochemical hotspots: temporal and spatial scaling of the impact of freshwater mussels on ecosystem function. *Freshw Biol* **60**, 563–574 (2014).
28. Vaughn, C. C. & Taylor, C. M. Macroecology of a host-parasite relationship. *Ecography* **23**, 11–20 (2000).
29. Howard, A. D. A river mussel parasitic on a salamander. *Natural History Miscellaneous* 1–6 (1951).
30. Barnhart, M. C., Haag, W. R. & Roston, W. N. Adaptations to host infection and larval parasitism in Unionoida. *J North Am Benthol Soc* **27**, 370–394 (2008).
31. Strayer, D. *Freshwater Mussel Ecology*. <https://doi.org/10.1525/california/9780520255265.001.0001> (University of California Press, 2008).
32. Haag, W. R. Reassessing Enigmatic Mussel Declines in the United States. *Freshwater Mollusk Biology and Conservation* **22**, 43 (2019).
33. Strayer, D. L. *et al.* Changing Perspectives on Pearly Mussels, North America's Most Imperiled Animals. *Bioscience* **54**, 429 (2004).
34. Haag, W. R. *North American Freshwater Mussels*. <https://doi.org/10.1017/cbo9781139048217> (Cambridge University Press, 2012).
35. Moore, A. P., Galic, N., Brain, R. A., Hornbach, D. J. & Forbes, V. E. Validation of freshwater mussel life-history strategies: A database and multivariate analysis of freshwater mussel life-history traits. *Aquat Conserv* **31**, 3386–3402 (2021).
36. Haag, W. R. The role of fecundity and reproductive effort in defining life-history strategies of North American freshwater mussels. *Biological Reviews* **88**, 745–766 (2013).
37. Haag, W. R. & Rypel, A. L. Growth and longevity in freshwater mussels: evolutionary and conservation implications. *Biological Reviews* **86**, 225–247 (2010).
38. Thurman, L. L. *et al.* Persist in place or shift in space? Evaluating the adaptive capacity of species to climate change. *Front Ecol Environ* **18**, 520–528 (2020).
39. Violle, C. *et al.* Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
40. Hopper, G. W. *et al.* A trait dataset for freshwater mussels of the United States of America. *Figshare*. <https://doi.org/10.6084/m9.figshare.24115998> (2023).
41. FMCS. *The 2021 checklist of freshwater bivalves (Mollusca: Bivalvia: Unionida) of the United States and Canada* Considered and approved by the Bivalve Names Subcommittee December 2020. (2021).
42. MolluscaBase eds. MolluscaBase. Accessed at <https://www.molluscabase.org> on 2020-10-20 (2023).
43. Graf, D. L. & Cummings, K. S. The MUSSEL project Web Site: MUSSELp. <http://www.mussel-project.net/>. Accessed [20 October 2021]. (2020).
44. Illinois Natural History Survey & Ohio State University Museum of Biological Diversity. The Freshwater Mussel Host Database. (2017).
45. Fogelman, K. J. *et al.* A review of lethal thermal tolerance among freshwater mussels (Bivalvia: Unionida) within the North American faunal region. *Environmental Reviews* **31**, 278–297 (2023).
46. Williams, J. D., Butler, R. S., Warren, G. L. & Johnson, N. A. *Freshwater mussels of Florida*. vol. 53 (American Library Association, 2015).
47. Parmalee, P. W. & Bogan, A. E. *The Freshwater Mussels of Tennessee*. (The University of Tennessee Press, 1998).
48. Daniel, L. G. & Diarmaid, Ó. Foighil. The Evolution of Brooding Characters Among The Freshwater Pearly Mussels (Bivalvia: Unionoidea) of North America. *Journal of Molluscan Studies* **66**(2) 157–170, <https://doi.org/10.1093/mollus/66.2.157> (May 2000).
49. Hoggarth, M. A. Descriptions of some of the glochidia of the Unionidae (Mollusca: Bivalvia). *Malacologia* **41**, 1–118 (1999).
50. *Common and Scientific Names of Fishes from the United States, Canada, and Mexico, 8th edition*. <https://doi.org/10.47886/9781934874691> (American Fisheries Society, 2023).
51. Bain, M. B., Finn, J. T. & Booke, H. E. Quantifying Stream Substrate for Habitat Analysis. *Studies. N Am J Fish Manag* **5**, 499–500 (1985).
52. Pfeiffer, J. M., Dubose, T. P. & Keogh, S. M. Synthesis of natural history collections data reveals patterns of US freshwater mussel diversity and decline. *BioRx* (2023).
53. Haag, W. R. A hierarchical classification of freshwater mussel diversity in North America. *J Biogeogr* **37**, 12–26 (2009).
54. Pebesma, E. & Bivand, R. *Spatial Data Science*. <https://doi.org/10.1201/9780429459016> (Chapman and Hall/CRC, 2023).
55. Zizka, A. *et al.* CoordinateCleaner: Standardized cleaning of occurrence records from biological collection databases. *Methods Ecol Evol* **10**, 744–751 (2019).
56. USFWS Species Data Explorer, Available from <https://ecos.fws.gov/ecp/report/adhocDocumentation?catalogId=species&reportId=species> (2022).
57. Turgeon, D. D. *et al.* *Common and scientific names of aquatic invertebrates from the United States and Canada: Mollusks*. (American Fisheries Society, 1998).
58. Williams, J. D. *et al.* A Revised List of the Freshwater Mussels (Mollusca: Bivalvia: Unionida) of the United States and Canada. *Freshwater Mollusk Biology and Conservation* **20**, 33 (2017).
59. Williams, J. D., Bogan, A. E. & Garner, J. T. *Freshwater Mussels of Alabama and the Mobile Basin in Georgia, Mississippi, and Tennessee*. (University of Alabama Press, 2008).
60. Hoggarth, M. A. & Gaunt, A. S. Mechanics of glochidial attachment (Mollusca: Bivalvia: Unionidae). *J Morphol* **198**, 71–81 (1988).
61. Moore, R. B. *et al.* User's guide for the national hydrography dataset plus (NHDPlus) high resolution. *Open-File Report* <https://pubs.usgs.gov/publication/ofr20191096> 10.3133/ofr20191096 (2019).

Acknowledgements

We thank the scientific community and natural history enthusiasts that published trait data for freshwater mussels. Without their efforts, this data resource would not have been possible.

Author contributions

This project was coordinated by G.W.H. and C.L.A. Access to museum collections and estimated range and stream characteristic data was coordinated by J.M.P., T.P.D. and S.M.K., T.P.D. collated range and stream characteristic data. S.M.K. collated biogeographic data. K.J.F. and J.A.S. collated the thermal trait data. The remaining attributes of the dataset were organized by J.R.B., M.E.K., J.D.L., M.B.L., D.H.N. and I.S.G.

Competing interests

The authors declare no competing of interest.

Additional information

Correspondence and requests for materials should be addressed to G.W.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023