Appendix S1: Pre-processing `popler` data

Before uploading datasets into the online `popler` database, we combined datasets, transformed datasets from wide to long form, converted non-ASCII characters, and modified ambiguous study site names.

The variables of many datasets were contained in two or more separate files, which we combined in a single file. When the original dataset provided data in wide form, we transformed it into long form. In wide form datasets, abundance data associated with different species was stored in separate columns. `popler` stores these datasets in long form, whereby each row of abundance data is related to a specific taxonomic unit in the table containing taxonomic information (Fig. 1B). We converted all data in ASCII format, because the encoding of the database is the UTF-8. We often re-defined study site names to unambiguously associate them with one of the 26 LTER sites. Many site names are alphanumeric codes (e.g. "U1") which can overlap across several LTER sites. Hence, we changed site names following a standard formula (namely, from "U1" to "site_sbc_U1", where "sbc" refers to the Santa Barbara coastal LTER site).

In a handful of cases, we removed single data rows from the original dataset. These data rows were associated with two types of typos in the original dataset. First, some abundance observations were not associated with a time of observation. We removed this data because `popler` can only accommodate population information associated with a time of observation. Second, a handful of abundance data points were clear typos (e.g. the letter "l" instead of a numeric value). We substituted these data points with a missing value. We uploaded these pre-processed datasets in the `popler` database through a Graphic User Interface developed in Python using libraries panda and pyqt5.

Table S1: Taxonomic variables contained in the popler table on original taxonomic information.

Variable
sppcode
kingdom
subkingdom
infrakingdom
superdivision
division
subdivision
superphylum
phylum
subphylum
class
subclass
order
family
genus
species
common_name

Table S2: Metadata variables used to describe the datasets stored in `popler`.

Variable	Description
`proj_metadat_key`	Unique ID
`lter_project_key`	ID of LTER site
`lter_project_key`	ID of LTER site
`title`	Title of study
`samplingunits`	Unit of measure (if any) referred to population data.
`datatype`	Data type: count, biomass, cover, density, and individual. These correspond to the tables in Fig. ??A.
`structured_type_n`	If individual data, this shows what type of structure is stored. A study can contain up to $n = 4$ types of structure.
`structured_type_n_units`	Unit of measure (if any) referred to structure data.
`studystartyr`	Start year of the study
`studyendyr`	End year of the study
`duration_years`	Duration of the study in years
`samplefreq`	Frequency of population census
`studytype`	Whether study is observational or experimental
`community`	Whether study includes single taxon (`community = F`) or multiple taxa (`community = T`)
`spatial_replication_level_n_extent`	Extent of spatial replication level number n. A dataset can have up to to 5 replication levels.
`spatial_replication_level_n_extent_units`	Unit of spatial extent of the n spatial replication level.
`spatial_replication_level_n_label`	Label of the spatial replication level (e.g. transect, plot, quadrat, ect.). The label of spatial replication level 1 is "site".
`spatial_replication_level_n_number_of_unique_reps`	The number of unique replicates for the nth level of spatial replication.
`treatment_type_n`	The type of treatment (e.g. resource manipulation). A study can contain up to n = 3 treatments.
`control_group`	If study is experimental, this shows the field(s) that identify the control replicate.
`derived`	Is population size data raw, or is it derived (e.g. it is aggregated)?
`authors`	Author(s) of the original dataset
`authors_contact`	Email address(es) of the author(s) associated with the original dataset.
`metalink`	url of the original dataset
`knbid`	Knowledge Network for Biocomplexity identifier.

Table S3: LTER identification acronyms and their meaning as used in the `popler` database.

Variable	LTER name
AND	Andrew Forest LTER
ARC	Arctic LTER
BES	Baltimore Ecosystem Study
BNZ	Bonanza Creek LTER
CAP	Central Arizona - Phoneix LTER
CCE	California Current Ecosystem LTER
CDR	Cedar Creek Ecosystem Science Reserve LTER
CWT	Coweeta LTER
FCE	Florida Coastal Everglades LTER
GCE	Georgia Coastal Ecosystems LTER
HBR	Hubbard Brook LTER
HFR	Harvard Forest LTER
JRN	Jornada Basin LTER
KBS	Kellogg Biological Station LTER
KNZ	Konza Prairie LTER
LNO	LTER Network Office
LUQ	Luquillo LTER
MCM	McMurdo Dry Valleys LTER
MCR	Moorea Coral Reef LTER
NCO	LTER Network Communications Office
NTL	North Temperate Lakes LTER
NWT	Niwot Ridge LTER
PAL	Palmer Antarctica LTER
PIE	Plum Island Ecosystems LTER
SBC	Santa Barbara Coastal LTER
SEV	Sevilleta LTER
SGS	Shortgrass Steppe LTER
VCR	Virginia Coastal Reserve LTER