

PROCEED v6.0

Phenotypic Rates of Change Evolutionary and Ecological Database

TimeSeriesDB — Metadata

General

Release date

2022/08/12

Citation

Gotanda, Kiyoko M.; Gorné, Lucas D. (2022) "Phenotypic Rates of Change Evolutionary and Ecological Database (PROCEED)", version 6, Borealis, <https://doi.org/10.5683/SP3/NXSL3Q>

Subject

Ecology and Evolution

Keywords

Contemporary evolution, haldanes, darwins, quantitative traits, time series.

Current curator

Kiyoko M. Gotanda^{1, 2} <kgotanda@brocku.ca>

Lucas D. Gorné¹ <gorneld@gmail.com>

Affiliations

1. Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada.

2. Département de Biologie, Université de Sherbrooke, Sherbrooke, Quebec, Canada.

Technical

Language: English

Decimal point="."

Not available data (NA)=" " (i.e. empty cell)

Description, Inclusion/Exclusion criteria

This is a database related to the PROCEED RatesDB. TimeSeriesDB is built on a subset of the studies included in the RatesDB. This is the subset of allochronic studies with more than two sampling times along the whole elapsed period studied.

This database includes studies that measure intraspecific change in quantitative (continuous or counting) traits. The maximum elapsed time of the time series is no longer than 500 years (however, currently published papers had cut-off number of years at 200/300 years, up to user to decide to keep those above that previous threshold). The included studies followed a single population through time.

The database includes systems in natural or experimental conditions but in which reproduction was not manipulated (i.e. NO artificial selection is included). It means that, the experimental systems manipulate the environment of the populations being studied but do not select traits deliberately (e.g. an herbivory exclusion experiment,

where the type and load of herbivory are manipulated and the populations evolve under this new condition).

The database does not include systems where the phenotypic change is due to interspecific hybridization, polyploidy, or other chromosomal mutations.

Variables:

uts_ID

Unique identifier for each entry in the TimeSeriesDB. It is a code of the form uts_XXXX, where each X is a character of the ordered sequence: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z. "uts_ID" are correlative along the whole database, such as "uts_00jz" < "uts_00k0".

es_ID

The same identifier as in RatesDB ("PROCEEDv6_RatesDB_README" file). Here, it works as an identifier for each individual time series.

ref_ID

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Unique identifier for each primary source of information (paper, thesis, report, etc).

sys_ID

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Unique identifier for each study system (see "system" for the definition).

released_sys

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

sys_ID of the system originating the new system after removal/extinguishment/stop of the driver/disturbance. For example, the system s445 (Ovis canadensis in Ram Mountain, Alberta, Canada when a hunting moratorium was placed starting in 1996) is the same population of system s122

(*Ovis canadensis* in Ram Mountain, Alberta, Canada when hunting was allowed up to 1995). So, for the system s445 “released_sys”=“s122”. On the contrary, for the system s122 “released_sys”=NA (because s122 was not derived from any other system in the database). Future studies with similar situations (i.e. a single population in periods with different evolutionary drivers) should include, in “released_sys” of the new condition, the sys_ID of the former condition.

released_es

Same as in “RatesDB”. Values linked to the “RatesDB” by the “es_ID”.

es_ID of the case originating the new case after removal/extinguishment/stop of the driver/disturbance in the system. A system being released from the original driver generates a new system. If more than one trait is measured in both systems (the original and the new one), “released_es” links the cases in the new system with the measurements of the same trait in the original systems. For example, in the systems s122 and s445, two traits were measured: mean weight (kg) and mean horn length (cm). In s122, the change in mean weight is the case “es_ID”=“es2798” and the change in mean horn length is the case “es_ID”=“es2799”. So, for the case es9277 (change in mean weight in the system s445), “released_es”=“es2798”. For the case es9278 (change in mean horn length in the system s445), “released_es”=“es2799”. On the contrary, for es2798 and es2799 (system s122), “released_es”=NA (because s122 was not derived from any other system in the database). Future studies with similar situations (i.e. a single population in periods with different evolutionary drivers) should include, in “released_es” of the new condition, the es_ID of the same traits in the former condition.

reference

Same as in “RatesDB”. Values linked to the “RatesDB” by the “es_ID”.

This is the reference of the primary source of information, which is given in the following format (all one line): Last name of the first author and year, Journal (standard abbreviations) Volume: Page numbers or article identifier (e.g. Cox 1997, Can. J. Fish. Aquat. Sci. 54: 1159-1165)

taxa

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

This is a coarse classification, not necessarily a clade (monophyletic group).

Current categories are: **Annelid, Mollusc, Arthropod, Fish, Amphibian, Mammal, Bird, Reptile, Plant.**

phylum

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Taxonomic clade of rank "phylum" including the studied species, according to the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>) last retrieved May 2022. The supplementary table "PROCEEDv6_taxonomy.csv" contains the complete list of "species" and its "phylum".

class

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Taxonomic clade of rank "class" including the studied species, according to the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>) last retrieved May 2022. The supplementary table "PROCEEDv6_taxonomy.csv" contains the complete list of "species" and its "class".

order

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Taxonomic clade of rank "order" including the studied species, according to the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>) last retrieved May 2022. The supplementary table "PROCEEDv6_taxonomy.csv" contains the complete list of "species" and its "order".

species

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Species name (or subspecific taxon name) of the population(s) being assessed in each case, as was published in the primary source of information.

sp_ncbi

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Species or subspecific taxon name of the population(s) being assessed in each case, according to the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root>) in February 2022. Two species were not found in the NCBI database, *Stipa occidentalis* and *Diaptomus sanguineus*, these names were checked in The Plant List (<http://www.theplantlist.org/>) and WORMS (<https://www.marinespecies.org/index.php>) respectively. The supplementary table "PROCEEDv6_taxonomy.csv" contains the complete list of "species" and its "sp_ncbi".

sex

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Sex of the individuals measured (male, female, asexual, mixed=male+female, NA(empty cell)=sex not reported in the study).

c_value

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

The amount (in picograms) of DNA contained within a haploid nucleus (e.g. a gamete) or one half the amount in a diploid somatic cell. Most of the values were taken from:

- Gregory TR. 2021. Animal Genome Size Database. <http://www.genomesize.com>.
- Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD. 2019. Plant DNA C-values Database (Release 7.1). <https://cvalues.science.kew.org/>

Detailed information of the variable, sources, ploidy level, analytical methods, etc in the supplementary table "PROCEEDv6_species_genome.csv" and "PROCEEDv6_species_genome_README.txt".

WHEN MORE THAN ONE C-VALUE IS AVAILABLE FOR ONE SPECIES

- If there is no difference in ploidy level or chromosome number among records, c-values were averaged.
- Records using Feulgen densitometry (Fe), flow cytometry (FC_any type), DAPI microdensitometry (MDAPI), Gallocyanin chrom alum densitometry (GDC), or Fulgen image analysis densitometry (FIA) methods are preferred over other methods. Thus, records using these methods were chosen for averaging when available.
- In general, if the genome size is available from several subspecies, the value for each subspecies was previously averaged by LG, then the average among them was imputed for the necessary species.
- In general, but especially in plants, when a species has more than one genome size due to differences in ploidy level or chromosome number, information for the specific populations within the system were researched. If no information was found, the available c-values were averaged.

SPECIAL CASES in C-value

- in birds (low c-value variability), if the genome size is available from several congeners (but not for the necessary species), the values of the multiple measurements of c-value (if more than one) for each congener were previously averaged by LG, then the average among congeners was imputed for the necessary species.

system

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

It is a population or a set of populations, of a given species, experiencing a specific driver/disturbance. If the design is synchronic, the system includes both diverging populations (or sets of populations). Each system has its own unique combination of species, disturbance/driver and location. Within a given system, you can have multiple traits – for example, tarsus length and fledging date. We suggest keeping it general. Please note you can have both allochronic and synchronic data within the same system (e.g. Leaver 2012, Biol. J. Linnean. Soc. 107: 494-509). Example systems: "Salmo trutta (brown trout) monitored for 37 y following fish ladder

construction (1966) in River Gudbrandsdalslågen, Norway"; "Troglodytes troglodytes (Eurasian Wren) migration arrival date monitored in its native range in Europe"; "Oncorhynchus gorbuscha (pink salmon) long term trends on northwest coast of US".

genphen

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

It distinguishes studies measuring individuals that developed in nature (**Phenotypic**) from those based on measurements of individuals bred in common-garden conditions or applying other quantitative-genetic methods (**Genetic**). Phenotypic does not mean it is not genetic; we just do not have the information and it does not automatically imply plasticity. Quantitative genetic methods (heritability does not count) are usually based on animal model methods and calculate breeding values which is the genetic contribution to a trait (Wilson et al. 2010). Common garden experiments performed from vegetative (clonal) propagules collected in the field (e.g. tiller, an anatomic unit in grasses) are still "Phenotypic".

out_in

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Population(s) outside or inside its native range. Of your two data points (samples) to calculate the rates – if one of the datapoints is outside the native range, it is considered “out”. Natural range expansions are considered “in”. You can have an introduction within the native range which would be “in”. Invasive species, with range expansions, would be “out”.

new_disturbance

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

It is a classification of the environmental change (or driver/disturbance) driving the phenotypic change, based on the categories used by Hendry 2008 (Hendry AP, Farrugia TJ & Kinnison MT. 2008. Human influences on rates of phenotypic change in wild animal populations. *Molecular Ecology* 17(1): 20–29) and Palkovacs 2012 (Palkovacs EP, Kinnison MT, Correa C, Dalton CM & Hendry AP. 2012. Fates beyond traits: Ecological consequences of human-induced trait change. *Evolutionary Applications* 5(2): 183–191). The current categories are:

- **Climate change:** this disturbance can often be difficult to clearly identify. A disturbance is classified as "Climate change" when the primary source of information explicitly considers climate change as a potential effect, or when the driver of the system is a climatic event (e.g. *Brassica rapa* (field mustard) changed flowering time after drought in Southern California, USA). Experimental systems manipulating CO₂ levels were considered "Pollution" (See "Pollution" definition).
- **Hunt_harv:** any case of hunting or harvesting of a species by humans.
- **Introduction:** "when humans transferred a species to a new geographical location, and comparisons were then made between introduced and ancestral populations". Also, range expansion after introduction (i.e. "when humans introduced a species to a new geographical location, and the species then spread on its own accord to occupy multiple sites. Comparisons were here made among the self-colonized populations").
- **Pollution:** an environmental change is classified as "Pollution" when chemical substances (e.g. heavy metals, pesticides, etc) are added or released into the environment contaminating, dirtying, and/or making harmful an area (usually air, water, or soil) where the focal species occur. The environmental change resulting from changing the natural concentration of substances present in the environment (e.g. a change in the level of CO₂ in the air or a change in the level of nutrients in the soil) or a physical parameter (e.g. temperature, pH, etc) is also categorized as "Pollution" even if not harmful to the species being studied.
- **Landscape change:** It is a habitat modification (different from "Pollution" and the introduction of a new species) in the patches occupied by the population(s) being studied. It can be the creation/restoration of new patches/habitats for the population(s) (e.g. a newly formed freshwater pond for an aquatic species, newly restored habitat), changes in the connectivity among patches/habitats (e.g. the construction of a dam in a river or the construction of a fish ladder, level of fragmentation of the habitat), or a change in the land-use/quality of the patches (e.g. forest clearing, livestock/herbivores inclusion/exclusion, newly irrigation system).
- **Response to introductions:** when a species (native or alien) is responding to the introduction of a new species. The newly introduced species can be a competitor, a predator, a parasite, a new host or prey, etc. The comparison is

established among populations (or times, according to the design) of the responding species (not the newly introduced species). When the responding species is native, "out_in"="In"; when it is alien, "out_in"="Out".

- **Other:** It is "in situ natural variation" and "Self induced range or host expansion", i.e. when native populations were not subject to an obvious human impact, most studies in this category involve the long-term monitoring of natural populations, such as Darwin's finches of the Galápagos (Grant & Grant 2002, 2006); also, when new populations were established without any direct human influence, and comparisons were then made among the new populations or among the new and ancestral populations (Examples include birds colonizing new habitats within their native range—Yeh & Price 2004—or new islands outside their native range—Clegg et al. 2002—).

env_change

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

- **o:** ongoing environmental change. The environmental change has been occurring through time and the phenotypic change was measured in a temporal window within this process (but not from the starting point).

- **n:** novel environmental change. The environmental change has a defined starting point and phenotypic change was measured from this point.

This category applies only to allochronic studies, synchronic are always "env_change"="n" (because the proposed ancestral state is placed previous to the drivers happening). When "new_disturbance"="Other", if the driver it is unknown, "env_change"="NA" (empty cell). But in some cases the driver of change is known, like "self induced range expansion", also experimental systems can represent "in situ natural variation" but from a known starting point. In these cases, the value of "env_change" can be chosen.

experimental

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

States if the phenotypic change is driven by experimentally set conditions (**N**=no; **Y**=yes). It is not artificial selection (i.e. selection of specific attributes by choosing

breeders and manipulating mating) but a process of natural selection in an experimental environment.

exp_type

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

When "experimental"="N", "exp_type"=NA (empty cell). When "experimental"="Y", the following categories were applied:

- **01**: A new environmental condition is imposed on a natural population, in the field, in situ.
- **03**: A new environmental condition is imposed on a natural population, in the field, ex situ (transplant).
- **05**: A new environmental condition is imposed on a natural population, in a common garden/greenhouse experiment.
- **07**: A new environmental condition is imposed on an artificially generated population in the field. Here, the artificial population is a set of genotypes from the natural population, evenly represented by (full or half) sibling families.
- **09**: A new environmental condition is imposed on an artificially generated population in a common garden experiment. Here, the artificial population is a set of genotypes from the natural population, evenly represented by (full or half) sibling families.
- **11**: A new environmental condition is imposed on an artificial population (it was obtained by randomly crossing F2 individuals from a set of interbreeds between contrasting genotypes selected from a natural population) in a common garden experiment.

trait_description

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Trait being quantified in the words of the authors of the paper. If the name of the trait is not self descriptive, please include the definition from the paper (e.g. shell slenderness ($\log_{10}[\text{width}] / \log_{10}[\text{length}]$)). If units are given, please include. Each trait has its own row.

trait_type

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Types of traits based on Kingsolver and Diamond 2011 but with some additional categories. Current categories are:

- **phenology**: the timing of cyclic and seasonal life-history events (e.g. flowering date, laying date, and hatching date). Do not confound with "otherLH". For example, the time spent in a given stage of the life cycle (e.g. larval stage) or the time to reach a given stage of the life cycle (e.g. time to maturity, time to first flower) are life-history traits, not phenology. "new.data.type" for "phenology" traits is typically a date while for "otherLH" it is time.

- **behaviour**: "Broadly speaking, animal behavior includes all the ways animals interact with other members of their species, with organisms of other species, and with their environment. Behavior can also be defined more narrowly as a change in the activity of an organism in response to a stimulus, an external or internal cue or combination of cues." (<https://www.khanacademy.org/science/ap-biology/ecology-ap/responses-to-the-environment/a/intro-to-animal-behavior>)

- **physio**: traits related to physiology or a biochemical parameter (e.g. enzymatic activity). Example physiology traits include metabolism, bite force, leaf toughness, decomposition rate, stoichiometry (e.g. C:N), concentration or amount (yield) of a given chemical compound in the body (or part of the body), diet (trophic level, diet, food type), severity or level of affectation by a disease/pathogen/parasite/herbivore (e.g. Fungal infection % of plants, % leaf consumed), dose/concentration/time exposed to a given substance/condition (e.g. pesticide, drought, light, temperature, etc) needed to produce some effect (e.g. death, flowering, etc.), performance of a standard parasite/herbivore/competitor feeding/competing on/with the focal population (e.g. Pupal weight of the herbivorous *Trichoplusia ni* feeding on native vs introduced populations of *Eschscholzia californica*; or Leaf feeder (*Galerucella pusilla*, L1-Imago) survival, allelopathic effects of *Centaurea maculosa*, *Achillea* biomass ratio with-competition to without-competition).

- **growth**: a measurement of the change in size (or a proxy of size) in a time period (where the starting point is not the birth, germination or equivalent), or a growth rate (irrespective of the starting point). It can be a relative or an absolute

measurement of growth or growth rate. Irrespective of the environmental condition (pollution, stress, competition, herbivory/parasites, fertigated/non-fertigated, etc.), the measurements of growth or growth rate are classified as "growth". If the trait is a change in growth (or another performance measurement) see "response".

- **size**: Total body mass, or an overall aspect/measurement of body size as a whole (e.g. total length, plant height, thorax width, plant diameter, basal area, number of tillers (grasses)). Size of parts of the body (e.g. tarsus length, intracranial length, length of the longest leaf, etc.) are not included here but in "othermorphology". In plants, measurements of the whole aboveground size (e.g. aboveground biomass, plant height, plant diameter or basal area, etc.), as well as measurements of the whole belowground size (i.e. root mass) are included as "size". Other morphological features of roots (e.g. length, diameter, etc.) are classified as "othermorphology". The "stem diameter" of tree species is a measurement of "size" but in herbaceous plants, it is "othermorphology". "size" is not necessarily the final or average body size in a population, the measurements of size at a given age/stage (but not at birth/hatch or equivalent) are also classified as "size" (e.g. Body mass lambs (kg)). Note that, similar situations can be classified differently according to the operational decision of the authors of the primary source of information, e.g.: in a common garden experiment for a short period of time if the authors report final biomass, the trait is classified as "size", but if the authors report growth rate (biomass over time) it is classified as "growth". If the trait is a change in size (or another performance measurement) see "response".

- **othermorphology**: all remaining morphological traits. This includes the size of parts of the body (e.g. tarsus length, intercranial-length, etc.) or plant (e.g. leaf size). The size of a part of the body at a given age/stage is also "othermorphology". It is independent of the environmental condition of the measurement, i.e. a measurement of tarsus length (or leaf size, etc) is always "othermorphology" independently of if it was measured in a controlled environment or in a polluted environment (or in another condition), and so on.

- **otherLH**: all remaining life history traits that are not classified as phenology or growth (e.g., offspring size, # of offspring, fecundity, survival, parental investment, reproductive effort/allocation, time to reach maturity or any stage in the life cycle). It is independent on the environmental condition of the measurement, i.e. a measurement of survival (or fecundity, etc.) is always "otherLH" independently of is

was measured in a controlled environment or in a polluted environment (or in another condition), and so on. If the trait is a change in survival (or another performance measurement) see "response".

- **response**: response to a number of abiotic and biotic factors. It is the change, relative change, or ratio of a measurement of performance (size, growth, survival, fecundity, etc) between two conditions (e.g. Size reduction—difference in size between pH 7 and 4, relative to size in pH 7—; Leaf Cpi—[no.leaves_woc - no.leaves_wc]/no.leaves_woc—; Mycorrhizal response [inoculated_biomass-Uninoculated_biomass]/Uninoculated_biomass).

new.data.type

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

The current categories are:

- **linear**: a measurement of length with a dimensionality of 1. A ratio between lengths (or a length over another variable) is not a "linear" variable because its dimensionality is not 1. A ratio between linear dimensions (e.g. length over width, or height over diameter), or a linear dimension over another variable (e.g. specific root length (SRL) [mm/mg]) is not "linear" because its dimensionality is not 1 (e.g. $SRL = [1D]/[3D]$, dimensionality -2).
- **area (2D)**: a measurement of area with a dimensionality of 2. A ratio between areas, or an area over another variable (e.g. specific leaf area (SLA) is not "area (2D)" because its dimensionality is not 2 (e.g. $SLA = [2D]/[3D]$, dimensionality -1).
- **cube (3D)**: a measurement of volume or mass or amount of matter with a dimensionality of 3. A ratio between volumes/mass, or a volume/mass over another variable (e.g. "Stem mass ratio (stem mass / whole plant mass)" or "Rhizome mass over length") is not "cube (3D)" because its dimensionality is not 3 (e.g. stem mass / whole plant mass = $[3D]/[3D]$ is a dimensionless proportion; Rhizome mass over length = $[3D]/[D]$, dimensionality is 2 but it is not an "area (2D)" because it depends on the shape of the organ and the density of the tissue, so it is "other").
- **count**: an amount of discrete units (e.g. number of leaves, clutch size, fecundity when number of offspring or propagules, etc.).

- **proportion**: the ratio between a part over the total (including that part). It can be expressed as a proportion (0-1), as a percentage (0-100). Some concentrations are also proportions, concentrations expressed as % M/M, % V/V, ppm, mg/g, ug/g (or similar). When the ratio is not a part over the total, this is not a "proportion" (e.g. root to shoot biomass, reproductive over somatic/vegetative biomass), this is a dimensionless ratio (i.e. "ad_ratio").

SPECIAL CASES OF PROPORTION:

Molar concentration or umol/g were also classified as "proportion". These are not exactly proportions because the numerator and denominator are not in the same units but are still a measurement of the ratio between a part over the total.

A ratio between a given amount in a stressing/disturbing condition (e.g. biomass in a toxic environment or in presence of parasite/herbivore or competitors) over a control condition expected to express the maximum potential (e.g. biomass in a non-toxic environment or in absence of parasite/herbivore or competitors), was also considered a "proportion". Similarly, the relative difference between these conditions was also classified as "proportion", because it is a proportional change (e.g. tolerance to herbivory: (damaged plant biomass - undamaged plant biomass) / undamaged plant biomass). The difference is that in the first situation "transf_data"="raw" and in the last one "transf_data"="resid".

- **time**: the time needed for some process or event.

- **date**: a temporal landmark for a given event in a reference system with an arbitrary starting point (e.g. number of days from January 1st). A time in the day (e.g. hours from midnight) is the same type of trait.

- **temperature**: temperature in the scale and units reported in the primary source of information (e.g. °C, °K, °F).

- **rate**: a measurement of speed, an amount over time (e.g., growth rate, metabolic rate (cal/s), etc.). "rate" is a diverse category because the numerator can be "count" (e.g. eggs/day), "linear" (e.g. Growth (mm/day)), "proportion" (Grow(%/d)), "cube (3D)" (e.g. growth rate (pupal mass divided by preadult developmental period) OR Resting Metabolic Rate (mg O₂ h⁻¹)), "ad_ratio" (e.g. RGR=ln[m₂/m₁]/time), "other" ratios (e.g. Stomatal Conductance (mmol m⁻² s⁻¹)), etc. However, even when a trait is called "rate" in the primary source of

information is not necessarily a rate according to the current classification (e.g. "pregnancy rate", "survival rate", "germination rate" are "proportion"; "Unit leaf rate ($\ln [\text{mass_t1} / \text{mass_t0}]$)" is an "ad_ratio"; etc.).

- **ad_ratio**: the ratio between two independent amounts with the same dimensionality is a dimensionless ratio. It is not a proportion because the numerator is not a component or fraction of the denominator (e.g. shoot:root ratio, C:N ratio, reproductive over vegetative biomass).

- **index**: ordinal variables and their derivations. However, when a trait is called "index" in the primary source of information it is not necessarily an index according to the current classification. For example, "index of ear size" ($[\text{EL} \times \text{EW}]^{0.5}$) is a called index but it is quantitative variable proportional to the ear size, its dimensionality is 1. We classified this trait as "other". Another example is "Gonadosomatic index (ova mass over body mass)" this is a "proportion" even when called index.

- **other**: other ratios and other quantitative variables not meeting the previous descriptions.

"new.data.type" is particularly relevant when computing and comparing darwins (D) among different data types. The rate of change darwins is the relative (to mean) rate of change of a given variable (X) at a given time (dt) (Haldane 1949 Suggestions as to quantitative measurement of rates of evolution. *Evolution* 3(1): 51–56). Darwins is:

$$(1) \quad D = \frac{d\left(\frac{dX}{X}\right)}{d(t)}$$

by integrating $1/X$ on dX we obtain:

$$(2) \quad D = \frac{d(\ln(X))}{d(t)}$$

In a discrete interval of time (t), we get:

$$(3) \quad D = \frac{\ln(x_2) - \ln(x_1)}{t}$$

The dimensionality of the trait affects darwins. As an example, if we have three traits total length (l), body mass (m), and mass of gonads over total body mass

(p). The relation between these three traits will depend on the allometric relationships, but it would be approximately as follows:

$$m = a \cdot l^3$$

$$p = \frac{b \cdot l^3}{a \cdot l^3} = \frac{b}{a}$$

So, the rate of change in darwins for these traits would be:

$$D(l) = \frac{\ln(l_2) - \ln(l_1)}{t}$$

$$\begin{aligned} D(m) &= \frac{\ln(m_2) - \ln(m_1)}{t} \\ &= \frac{\ln(a \cdot l_2^3) - \ln(a \cdot l_1^3)}{t} \\ &= \frac{\ln(a) + \ln(l_2^3) - \ln(a) - \ln(l_1^3)}{t} \\ &= \frac{3 \cdot \ln(l_2) - 3 \cdot \ln(l_1)}{t} \\ &= \frac{3 \cdot (\ln(l_2) - \ln(l_1))}{t} \\ &= 3 \cdot D(l) \end{aligned}$$

If the allometric relationships remain constant:

$$\begin{aligned} D(p) &= \frac{\ln(p_2) - \ln(p_1)}{t} \\ &= \frac{\ln\left(\frac{b}{a}\right) - \ln\left(\frac{b}{a}\right)}{t} \\ &= 0 \end{aligned}$$

As was shown in the previous example, the type of variable, by itself, strongly affects the estimation of evolutionary change in darwins.

transf_data

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

If the trait measurement has been transformed and the means and standard deviations were computed on basis of the transformed data (e.g. body mass

expressed as $\log(g)$ instead of grams, so the mean is the average of the logarithm of the body mass). The current categories are:

- **raw**: mean and sd computed from raw data (no transformation).
- **ord**: the trait is the score in the axis of an ordination analysis (i.e. PCA, CVA).
- **arcsin**: arc sin (\sin^{-1}) transformation.
- **arcsin.sqr**: angular transformation, i.e. $2 \cdot \sin^{-1}(\sqrt{x})$.
- **resid**: residuals of regression or other statistical model ($y-\hat{y}$), centered variables ($y-\bar{y}$), absolute (Y_t-Y_c) or relative difference ($[Y_t-Y_c]/Y_c = Y_t/Y_c - 1$) between a treatment and a control.
- **ln**: mean and sd computed from natural log-transformed data. We also include data like $\ln(x)/\ln(y)$ and $\ln[x/y]/t$
- **log10**: mean and sd computed from log10-transformed data. We also include data like $\log_{10}(x)/\log_{10}(y)$ and $\log_{10}[x/y]$.

data.scale

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

- **ratio**: constant interval with a precise zero.
- **interval**: constant interval with an arbitrary zero (e.g. temperature, time of day, rescaled data: Principal components or discriminate functions).
- NA (**empty cells**): undefined scales and ordinal scales (i.e. "new.data.type"="index") because ordinal scales do not have a constant interval.

Darwins cannot use interval data, but Haldanes can (because it's rescaled units, essentially removing an absolute zero point). Ratio data are actually interval data, but with a clearly defined zero.

gtime

Same as in "RatesDB". Values linked to the "RatesDB" by the "es_ID".

Generation time expressed in years. There are several definitions of the generation time. For our purposes, we define the generation time as the time (in years) in which a population replaces itself.

- For species with non-overlapping generations or species whose individuals reproduce only once in life (e.g. monocarpic plants), the average age at maturity is appropriate.
- For species with non-overlapping generations or species whose individuals reproduce only once in life, with the average age at maturity shorter than one year but experiencing a period of non reproduction in the year (e.g. diapause in mosquitoes, plants with dormant seeds), the generation time was computed as the reciprocal of the number of generations per year (1/GPY). If the generation time is less than one year, the number will be a decimal (# days/365 or # months/12).
- For species with overlapping generations, individuals reproducing multiple times in life and age-structured populations, the age at which members of a given cohort are expected to reproduce is an accurate proxy. It can be computed from a life table as:

$$GT = \frac{\sum_{i=1}^n x_i \cdot l_{(x_i)} \cdot m_{(x_i)}}{\sum_{i=1}^n l_{(x_i)} \cdot m_{(x_i)}}$$

Where x_i is the age (or the mean for an age class), $l_{(x_i)}$ is the proportion of survivors (from the age=0) to a given age (x_i), and $m_{(x_i)}$ is the average fecundity of the individuals at a given age (x_i).

A similar measurement is the average difference in age between parent and offspring, but it requires a stable age structure in the population.

- For species with overlapping generations, individuals reproducing multiple times in life and stage-structured populations, the time it takes for the population to grow by a factor of its net reproductive rate (R_0) is an accurate proxy. It can be computed as:

$$GT = \frac{\ln(R_0)}{r}$$

such as:

$$e^{r \cdot GT} = R_0$$

Where $r=\ln(\lambda)$, and λ is the finite population growth rate. These parameters can be computed from the transition matrix.

Ultimately, the generation time is used to compute the number of elapsed generations between Sample 1 and Sample 2. Ideally, we want the specific generation time for the population (because the generation time can vary among populations). When the authors from the primary source of information provide an estimation of the generation time, or number of generations per year or total number of generations, this information was used. If this is unavailable, the following steps can be taken:

- Query the author directly (see contacting authors).
- Web of Science or Google Scholar search looking for the generation time.
- For marine species, consult the "The Marine Life Information Network" (<https://www.marlin.ac.uk/>). Then, search for your species, in the "Life history" table you might find the "generation time".
- An option for fish is to consult the FishBase (<https://www.fishbase.se/search.php>).
- The databases COMPADRE (<https://compadre-db.org/>) and COMADRE (<https://compadre-db.org/Data/Comadre>) provide population transition matrices for plants and animal species respectively. From these matrices, the generation time can be computed as the time it takes for the population to grow by a factor of its net reproductive rate by using the R package "popbio" (Stubben C & Milligan B. 2007. Estimating and analyzing demographic models using the popbio package in R. Journal of Statistical Software 22(11): 1-23). Generation time from Caswell eq. 5.73:

$$GT = \frac{\ln(R_0)}{\ln(\lambda)}$$

Example using “popbio” in R:

```
#Fmtr: fecundity matrix
#Umtr: survival matrix
#Amtr: (full) transition matrix

#R0:
```

```

Rmtr <- Fmtr*%fundamental.matrix(Umtr)$N
R0 <- eigen(Rmtr)$value[1]

#λ:
lmd <- pop.projection(Amtr, rep(100, nrow(Amtr)), 100)$lambda

#GT:
log(R0)/log(lmd)

```

- If the generation time, a table of life, a transition matrix, or the demographic information to build a table of life are not available, you can look for detailed information on the life cycle and the time and probability for each stage and transition. Based on this information you can produce an educated guess of the actual generation time.
- Look at the Encyclopedia of Life website (<http://eol.org/>). Then, (i) search for your species, (ii) click on read full entry, (iii) look for the box on reproduction – there might be information on age at maturity. For animals with a long gestation period (e.g. bears), sum this gestation period to the age at maturity. Then, the generation time is the earliest age when they give birth, not when they become sexually mature. Even this is (most likely) an underestimation of the actual generation time.
- The supplementary table "PROCEEDv6_species_GT.csv" summarizes all the information about generation time for each species, compiled in the current version of PROCEED.

date__year

This is the calendar year of each observation.

time__years

This is the relative year of each time series. Each time series starts from "time__year"=0. "time__years" can be a fraction of a year if the observations were made in intervals shorter than one year (e.g. 0, 0.5, 1, 1.5, 2 in an experimental system).

time__generations

This is the relative time of each time series in generations, according to the "time_years" and the "gtime" (from RatesDB linked by the "es_ID"). Each time series starts from "time_generations"=0.

n

sample size.

mean

mean of the phenotypic trait for a given sample in a time series.

sd

the standard deviation, around the mean of the phenotypic trait, for a given sample in a time series.

obs

observations and comments.

Code to link TimeSeriesDB to related values in RatesDB, compute "time_years", "time_generations" and input "uts_ID".

```
#####
setwd("...")
rates <- read.csv("PROCEEDv6_RatesDB.csv", na.string="")
tser <- read.csv("PROCEEDv6_TimeSeriesDB.csv", na.string="")

#updating TimeSeriesDB with current information in RatesDB
i=1
for (i in 1:nrow(tser)){
  tser$ref_ID[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"ref_ID"]
  tser$sys_ID[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"sys_ID"]
  tser$released_sys[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"released_sys"]
  tser$released_es[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"released_es"]
  tser$reference[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"reference"]
}
```

```

tser$taxa[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"taxa"]
tser$phylum[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"phylum"]
tser$class[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"class"]
tser$order[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"order"]
tser$species[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"species"]
tser$sp_ncbi[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"sp_ncbi"]
tser$sex[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"sex"]
tser$c_value[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"c_value"]
tser$system[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"system"]
tser$genphen[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"genphen"]
tser$out_in[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"out_in"]
tser$new_disturbance[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"new_disturbance"]
tser$env_change[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"env_change"]
tser$experimental[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"experimental"]
tser$exp_type[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"exp_type"]
tser$trait_description[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"trait_description"]
tser$trait_type[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"trait_type"]
tser$new.data.type[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"new.data.type"]
tser$transf_data[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"transf_data"]
tser$data.scale[i] <-
rates[which(tser$es_ID[i]==rates$es_ID),"data.scale"]
tser$gtime[i] <- rates[which(tser$es_ID[i]==rates$es_ID),"gtime"]
}

#Adding "time_years" and "time_generations" variables
starts <- which(!duplicated(tser$es_ID))

time_years <- vector(mode="numeric", length=nrow(tser))

i=1
for(i in 1:length(starts)){
  if(i<length(starts)){
    for(j in starts[i]:(starts[i+1]-1)){
      if(!is.na(tser$time_years[j])){
        time_years[j] <- tser$time_years[j]
      }
    }
  }
}

```

```

        next}
        time_years[j] <- tser$date_year[j]-
tser$date_year[starts[i]]
    }
}
if(i==length(starts)){
    for(j in starts[i]:nrow(tser)){
        if(!is.na(tser$time_years[j])){
            time_years[j] <- tser$time_years[j]
            next}
        time_years[j] <- tser$date_year[j]-
tser$date_year[starts[i]]
    }
}
}

time_generations <- time_years / tser$gtime
time_generations[which(!is.na(tser$time_generations))] <-
tser$time_generations[which(!is.na(tser$time_generations))]

tser$time_years <- time_years
tser$time_generations <- time_generations

#Adding "uts_ID"
digits <- c("0","1","2","3","4","5","6","7","8","9",
            "a","b","c","d","e","f","g","h","i","j","k","l",
            "m","n","o","p","q","r","s","t","u","v","w","x","y","z")
c1 <- 0
c2 <- 0
c3 <- 0
c4 <- 0

uts_ID_end <- "uts_zzzz"
uts_ID <- vector(mode="character", length=length(digits)^4)

i=1
repeat{
    i1 <- (as.integer((i-1)/length(digits)^3)+1)
    i2 <- (as.integer((i-1)/length(digits)^2)+1) -
length(digits)*as.integer((i-1)/length(digits)^3)
    i3 <- (as.integer((i-1)/length(digits)^1)+1) -
length(digits)*as.integer((i-1)/length(digits)^2)

```



```

        i4 <- (as.integer((i-1)/length(digits)^0)+1) -
length(digits)*as.integer((i-1)/length(digits)^1)

        c1 <- digits[i1]
        c2 <- digits[i2]
        c3 <- digits[i3]
        c4 <- digits[i4]

        uts_ID[i] <- paste("uts_",c1,c2,c3,c4, sep="")
if(uts_ID[i]==uts_ID_end) {break}
if(i==nrow(tser)) {break}
i=i+1
}

tser$uts_ID <- uts_ID[1:nrow(tser)]

write.csv(tser, "PROCEEDv6_TimeSeriesDB_2.csv", na="")
#####

```