Explanation of columns in the database “SampleTREE_TraitCompiled.csv”

In the database, “NA” denote missing data.

1) **kingdom, phylum, class, order, family, genus, species, species_name, author**: Taxonomic information

2) **taxonID**: Taxon ID in GBIF

3) **fossil_status**: Was the extracted species a fossil. All “resampled” have been resampled

4) **uniqueness_family** [Numeric; integer]: Number of species in the same family, based on GBIF backbone taxonomy

5) **uniqueness_genus** [Numeric; integer] Number of congeneric species, based on GBIF backbone taxonomy

6) **size_m, size_f, size_avg** [Numeric; in millimeters]: Size of male and female and average male & female.

7) **model_organism** [Binary. 1 = Yes; 0 = No]: Is the species an established scientific model organism beyond ecology/evolution (e.g. *Neurospora, Arabidopsis, Drosophila)*? (Subjective assessment)

8) **harmful_to_human** [Binary. 1 = Yes; 0 = No]: Is the species harmful to human in some way; e.g. pest, notorious alien species, potentially dangerous (lion, venomous spiders, etc.) (Subjective assessment)

9) **human_use** [Binary. 1 = Yes; 0 = No]: Any commercial and or cultural use? e.g. used as pets, as food, in the zoo, for pharmaceutics. (Subjective assessment) IMPORTANT: for “Zoo” you can use this resource:

https://zooinstitutes.com/#

10) **domain** [Categorical, One of = freshwater, marine, terrestrial]: If multiple domains apply, these are separated with a semicolon, e.g. “freshwater ; marine”

11) **reproductive_habitat**. [Categorical. One of = multiple (4+ habitats), forest, shrubland, grassland, river, desert, lake, other wetlands, rocky areas, subterranean habitats, marine oceanic, marine coastal, deep ocean floor, artificial, zoo/epiphytic or parasite; unknown] These general habitats are readapted from IUCN (https://www.iucnredlist.org/resources/habitat-classification-scheme). If multiple habitats apply, these are separated them with a semicolon, e.g. “shrubland ; grassland”.

12) **trophic_role** [Categorical. One of = producer, primary consumer, secondary consumer, tertiary consumers, decomposer, parasite, unknown]: If a species changes trophic role during the life cycle, separate them with a semicolon, e.g. “parasite ; secondary consumer”.

13) **common_name** [Binary. 1 = Yes; 0 = No]: Does the species has a popular name in English?

14) **colorful** [Binary. 1 = Yes; 0 = No]: Is the species bright-fully coloured? (Subjective assessment)

15) **color_blu** [Binary. 1 = Yes; 0 = No]: Is the species bright blue coloured[1] / has violet/blue/light blue markings? (Subjective assessment)

16) **color_red** [Binary. 1 = Yes; 0 = No]: Is the species bright red/purple coloured / has red/purple markings? (Subjective assessment)

17) **mimetism** [Binary. 1 = Yes; 0 = No]: Does the species present striking mimetic pattern (disruptive, aposematic, batesian)? We did not consider cryptic (Subjective assessment)

18) **photo_google** [Binary. 1 = Yes; 0 = No]: Are there ‘popular’ photo of the species on Google? Photo of the species in the natural habitat or in captivity (do not consider photo published in scientific articles or museal specimens).

19) **CITES** [Categorical. 1 = Yes; 0 = No]: Is the species listed in CITES?

20) **IUCN** [Categorical]: One of = Not evaluated (NE); Extinct (EX); Critically Endangered (CR); Endangered (EN); Vulnerable (VU); Near Threatened (NT); Least Concern (LC); Data Deficient (DD)] Extinction risk *sensu* IUCN as of October 2020.

21) **mean_divergence_time_Mya** , **median_divergence_time_Mya** [Numeric]: The mean and median divergence time (MYA) from *Homo sapiens* obtained from TimeTREE database (http://www.timetree.org/)

22) **n_occurrences** [Numeric; integer]: Number of unique species occurrences in GBIF

23) **n_geo_occurrences** [Numeric; integer]: Number of unique species occurrences with coordinates

24) **n_sampled_occurrences** [Numeric; integer]: Number of sampled occurrence for estimations of range size

25) **centroid_lat, centroid_long** [Numeric; decimal degrees]: Coordinates of the distribution centroid in WGS84 decimal degree

26) **range_size** [Numeric]: Extent of the range size

27) **higherGeography ; Verbatim Habitat** [Text] Additional info on habitat and distribution extracted automatically form IUCN and GBIF

28) **biogeography** [Text]: species biogeographic region

29) **total_wiki_pgviews** [Numeric]: Number of wikipedia page views for the species

30) **wiki_langs** [Numeric]: Number of wikipedia page languages for the species

31) **wiki_mean_month_pgviews** [Numeric]: Average views of the species’s wikipedia pages monthly

32) **Total_wos** [Numeric]: Number of papers focusing on the species in the Web of Science

33) **Assignment** [Text]: Author who extracted traits

33) **Notes** [Text]: Any additional information