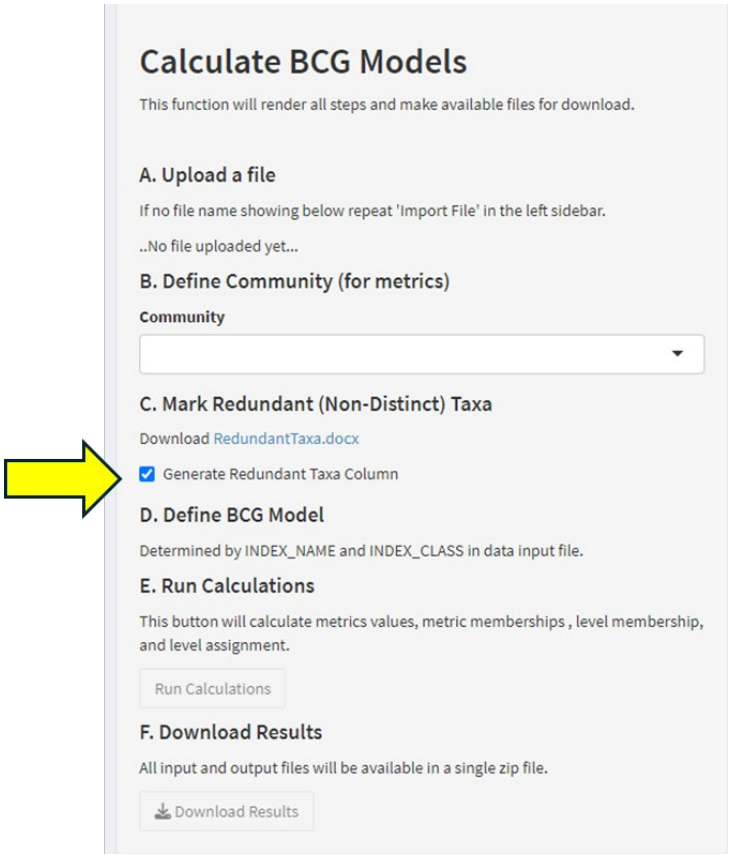
Redundant taxa (7/19/2024)

Default setting



The default setting (box is checked) marks redundant (non-distinct) taxa, using the same criteria that were used for index development (see Appendix A for more details). The intent is to avoid double-counting taxa that may be redundant (or non-distinct, depending on your preferred terminology). For example, if organisms identified as Dytiscidae (family-level) and Oreodytes (genus-level) are both present in a sample, organisms identified as Dytiscidae could be the same taxon as Oreodytes.

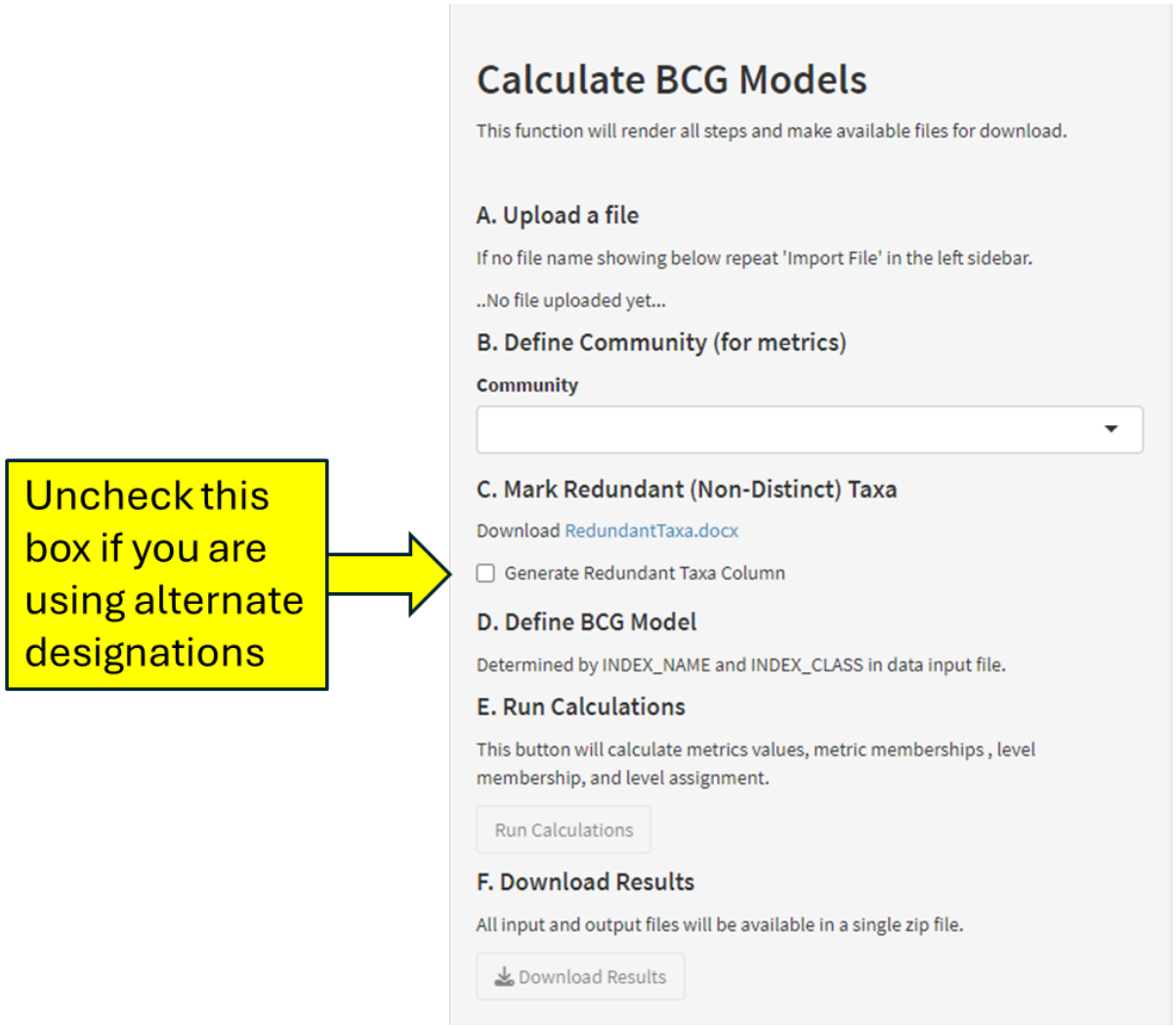
Redundant taxa are excluded from richness metric calculations but are included in composition (percent individuals) and tolerance metrics. This is different from NonTarget designations, which apply to organisms that are terrestrial or otherwise nonrelevant. NonTarget taxa are excluded from all metric calculations.

When you calculate the index with the box checked, a new TRUE/FALSE field is added to the input file called 'EXCLUDE' (TRUE = redundant; FALSE = distinct). This is done by the BioMonTools R package (https://leppott.github.io/BioMonTools/). The R tool identifies and marks redundant taxa on a sample-by-sample basis.

What if you want to use your own redundant taxa designations?

1. Create an input file with all the required fields (see Prepare Data – Outside App web page).
2. Add a column with your redundant taxa designations. Name the column 'EXCLUDE' and convert to TRUE/FALSE if needed (TRUE = redundant; FALSE = distinct).
3. Uncheck the box (see screenshot below)
4. Click the 'calculate metric values' button

Uncheck this box if you are using alternate designations



APPENDIX A

Redundant Taxa Decision Criteria

When calculating metrics there are occasions when certain taxa are not included in taxa richness metrics but the individuals are included for all other metrics. This is done to avoid double counting taxa that may have been identified to a coarser level when taxa of a finer level are present in the same sample.

These taxa have been referred to by many names – e.g., Excluded Taxa, non-distinct taxa, redundant taxa or ambiguous taxa. This document will use the term Excluded.

When you run the markExcluded function in BioMonTools (https://github.com/leppott/BioMonTools), non-distinct taxa are excluded based on the following steps:

1. Calculate and find all taxa names that appear in a sample at each taxonomic rank more than once (for an example, see Figure A1). These are the potential "parents" to be excluded.

2. Check if any of the potential "parents" equal a final ID in their respective samples.

3. If you get a match these are marked as "Excluded"

All Excluded decisions are sample-specific and the rules should be reapplied if sample contents change. Also, if the level of effort or operational taxonomic units change, the Excluded taxa designations should be recalculated.

TAXA LIST

BCG Attribute	FinalID	Count	FFG	Thermal	Toler_Sed	Redundant	Excluded
4	*Nais*	7	NA	--	NA	FALSE	FALSE
4	*Atractides*	1	PR	--	NA	FALSE	FALSE
4	*Hygrobates*	3	PR	--	NA	FALSE	FALSE
4	*Lebertia*	6	PR	--	NA	FALSE	FALSE
4	*Sperchon*	2	PR	--	NA	FALSE	FALSE
3	*Torrenticola*	1	PR	--	NA	FALSE	FALSE
4	*Dytiscidae*	3	PR	--	NA	TRUE	FALSE
3	*Oreodytes*	1	PR	--	NA	FALSE	FALSE
3	*Heterlimnius corpulentus*	19	GC	--	5	FALSE	FALSE
3	*Narpus concolor*	2	GC	--	5	FALSE	FALSE
3	*Clinocera*	1	PR	--	NA	FALSE	FALSE
4	*Neoplasta*	1	NA	--	NA	FALSE	FALSE
2	*Glutops*	2	PR	--	NA	FALSE	FALSE
x	*Ceratopogoninae*	2	PR	--	NA	FALSE	FALSE
4	*Thienemannimyia group*	9	PR	--	NA	FALSE	FALSE
4	*Micropsectra*	19	GC	--	NA	FALSE	FALSE

... Data_Taxa_Master | Data_Metrics | Data_Habitat | Data_Taxa_Samps | Samp0001 | Samp0009 | Samp0018

Figure A1. Example - Dytiscidae (family-level) is excluded from the richness metrics in this sample because these organisms could be the same taxon as Oreodytes (genus-level). The exclusion rule is applied on a sample by sample basis.

Below is a more detailed description of the process that the markExcluded function follows. Before starting, it is necessary to have a complete and correct master taxa list (all phylogenetic information and ranks).

Terminology

- Target Rank = intended level of taxonomy for identification, e.g., genus. Typically, specified in the project's SOP but can be adjusted during the OTU process.
- Parent or Parent Taxon = a taxon that occurs in the data in addition to other taxa in the same group that are identified to a more specific level. For example, the family Baetidae may occur in the data in addition to genera within the family Baetidae. In this case the name Baetidae is a parent to the other taxa within the family. Parents do not have to be only a single rank above the child taxon. That is, the class and order ranks are parents of any family ranks within them.
- Child or Children Taxa = a taxa or taxon that occurs in the data in addition to individuals identified to a coarser level. For example, the genera Baetis and Procloeon may occur in addition to the family Baetidae (of which the 2 genera listed are a member). In this case Baetis and Procloen are children of Baetidae.

Rule Development

For each sample:

1. Determine "potential" taxa for exclusion based on rank (or level) names appearing more than once in a sample.
    a. This is done for all ranks present; phylum, class, order, family, tribe, genus, species.
2. Check if any "potential" taxa are equal to a final (unique) ID in the same sample.
3. Stage is combined with taxa names if used in the dataset.

Requirements

1. A sample taxa table or data frame.
    a. All non-count and zero individual taxa have been removed.
    b. Unique sample ID code in a single column.
    c. A column with a final identification that is narrative not numeric. That is, Baetidae is ok but the ITIS number is not.
    d. Phylogenetic rank/level columns.
        i. This can be applied from a master taxa table but needs to be included in this table. One column per rank.
        ii. Names need to be consistently spelled.

Procedures

1. Find all potential Parents (those with a rank coarser than the target rank). This is done by creating a list of taxa rank names that appear more than once in a sample. This is done for each taxonomic rank.
2. The above list is compared to the final identifications for each sample.
    a. Special consideration is made for ranks of finer detail than genus. That is, names that are a combination of more than one field.
3. Any matches are marked as "Excluded".

There is still a need for manual review / QC check of the final list of Excluded designations.