**Excluded Taxa Decision Criteria**

When calculating metrics for benthic macroinvertebrates 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 more coarse 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, NonUnique Taxa, or Ambiguous Taxa. This document will use the term Excluded. This is done so that taxa to be removed from taxa richness calculations have to “opt in”. That is, it is assumed all taxa will be counted. Only those taxa marked as “Excluded” will be removed.

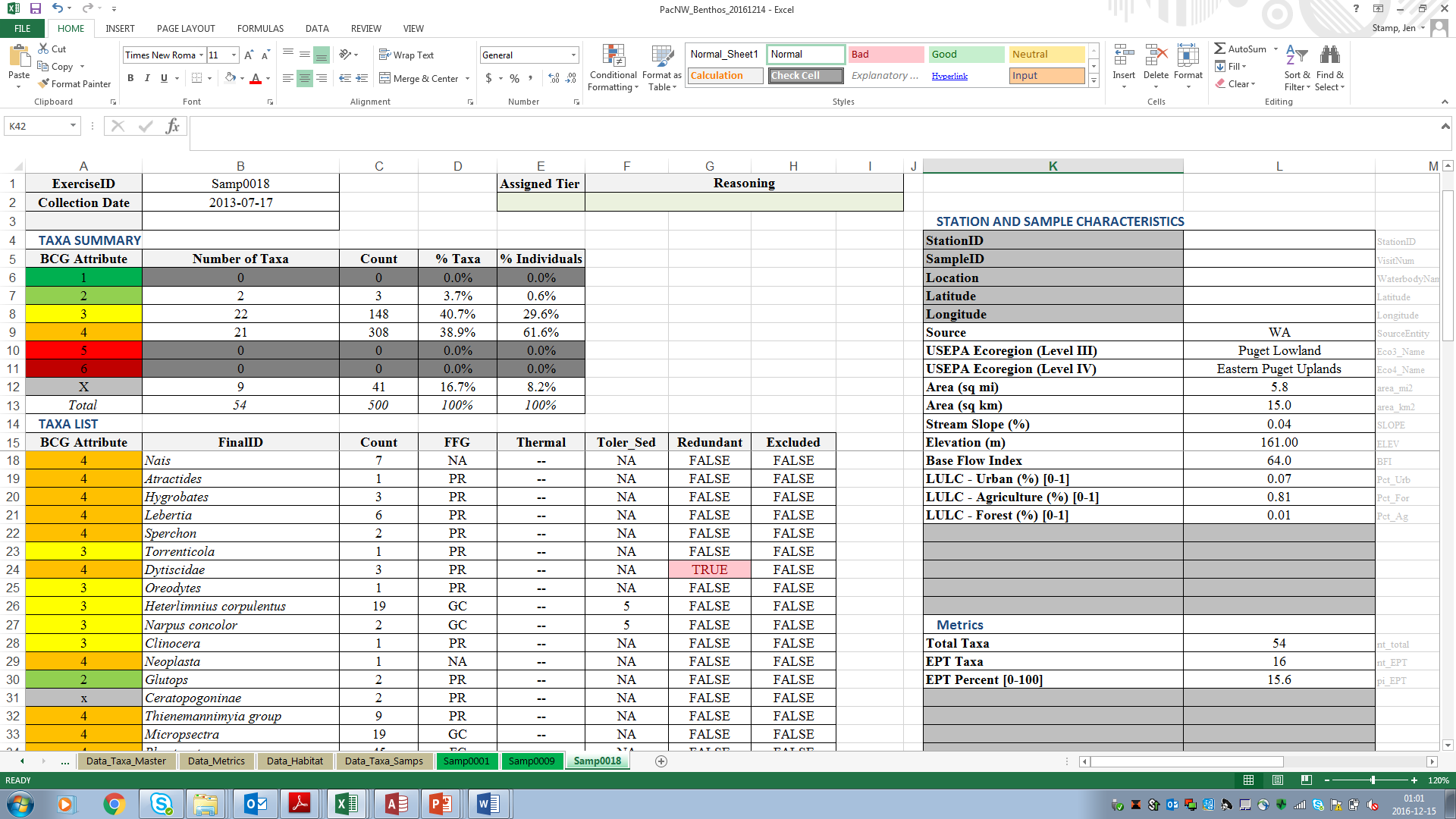
When the MassDEP IBIs were calibrated, redundant taxa were excluded from richness metric calculations based on the following steps:

1. Calculate and find all taxa names that appear in a sample at each taxonomic rank more than once (for an example, see Figure 1). 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 were marked as "Excluded"

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



**Figure 1**. 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 was followed. 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.
   1. 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.
   1. All non-count and zero individual taxa have been removed.
   2. Unique sample ID code in a single column.
   3. A column with a final identification that is narrative not numeric. That is, Baetidae is ok but the IT IS number is not.
   4. Phylogenetic rank/level columns.
      1. This can be applied from a master taxa table but needs to be included in this table. One column per rank.
      2. 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.
   1. 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.