**Comprehensive explanations of terms employed in our host-symbiont cophylogeny study**

Quickguide:

- Use **host\_tips\_linked\_corrected** for the number of host taxa.

- Use **symbiont\_tips\_linked** for the number of symbiont taxa.

- **Host\_range\_link\_ratio** and **Host\_range\_taxonomic\_breadth** are simple estimates of the overall degree of host range (generalism) captured in each study.

(i) **Host\_tips\_linked** and **Symbiont\_tips\_linked**

* Sometimes the phylogenies presented in a study include unlinked orphan taxa (i.e. those not linked to a symbiotic partner). No information is available to associate these orphan taxa with a host (in the case of orphan symbionts) or with a symbiont (in the case of orphan hosts).
* When we began this study, all taxa presented in author’s phylogenies were recorded in our data table. These taxa are all relevant to host-symbiont cophylogeny, but they are impossible to interpret, as it is unclear if: (a) the author’s did not know their symbiotic association, (b) the author’s did not sample their partner, or, (c) if their partner is now extinct and they are ligitimate orphans.
* More to the point, the cophylogeny methods we include do not cope with these orphan taxa. So, while authors do sometimes include them in their phylogenies, they are not present in their actual final cophylogenetic analyses (The field of cophylogenetic methods need updating with the development of more comprehensive analysis methods. In addition, each study case is fundamentally difficult to address and can only be properly resolved with painstaking lab and field studies. Hopefully our meta-analysis cuts through both these problems to elucidate the general underlying patterns).
* Since we only want to utilise information on taxa included in author’s analyses, we use the terms ‘host\_tips\_linked’ and ‘symbiont\_tips\_linked’ to be absolutely unambingous that we only consider host and symbiont taxa that are linked with one another. That is, we do not include orphan taxa that are presented in a tree, but not included in the analysis.
* Frequently authors are not very transparent about their analyses or they do not report what they actually did. But with reference to provided tanglegrams (phylogenies including links between symbionts and their hosts), we can deduce what taxa were included in their analyses.
* So, in conclusion:

**-Host\_tips\_linked** - can simply be regarded as the number of hosts – but see the important proviso below regarding ‘host\_tips\_linked\_corrected’.

**-Symbiont\_tips\_linked** - can simply be regarded as the number of symbionts.

**(ii) Host\_tips\_linked\_corrected**

* To complicate matters slightly, in a few cases, authors include mulitple within species tips in what is supposed to be a species-level analysis.
* In such cases, we decided that we should adjust down the number of ‘host\_tips\_linked’, since we are interested in relationships at species-level and above, as within-species patterns are much more complicated and reticulate, they are not covered by Fahrenholtz’ rule, and they are not even strictly relevant in phylogenetic analyses.
* Furthermore, if we were to include within species replication, this would artificially increase the number of ‘cospeciation events’ estimated.
* Thus, ‘host\_tips\_linked\_corrected’ should be the central measure of the number of hosts included in a study, in our meta-analysis. In practice, relatively few studies are affected by this issue, but it’s good to be clear about what was done.

**(iii) Total\_host\_symbioint\_links**

* This is the total number of the host-symbiont links present in a study (removing redundant within-species links), and is simply required for the calculation of ‘host\_range\_link\_ratio’, see below.
* If all symbionts were strict specialists, this would simply equal the number of symbionts included in the study. However, because symbionts are often associated with more than one host, this value is often higher than the total number of symbionts included in the study.

**(iv) Host\_range\_link\_ratio**

* This is just **Total\_host\_symbioint\_links** divided by **Symbiont\_tips\_linked**.
* This provides a simple estimate of symbiont host specificity (i.e. total number of links between hosts and symbionts / total number of symbionts included in the study)

**(v) Host\_range\_taxonomic\_breadth**

* This is an alternative estimation of symbiont host specificity to that captured by ‘host\_range\_link\_ratio’, that works by considering Linnean taxonomic rank.
* It is calculated by first summing the number of host taxonomic ranks linked to each symbiont (i.e. single host species = 1, multiple host species in the same genera = 2, multiple host genera = 3, multiple host familes = 4, multiple host orders = 5), and dividing by the total number of symbionts included in the study (i.e. 'symbiont\_tips\_linked').