**Methods**

**Measures of host-symbiont phylogenetic congruence**

We considered studies that reconcile host and symbiont phylogeny to test for topological congruence, using knowledge of the associations that exist between symbiont and host tips, where host phylogeny is considered the independent variable and the symbiont phylogeny is considered the dependent variable1. We included reported test statistics from the two most widely applied analytical approaches for estimating host-symbiont phylogenetic congruence (Extended Data Table 1): TreeMap2-4, based on estimates of shared discrete macroevolutionary events, and ParaFit5, based on branch-length comparisons to estimate overall similarity. Both approaches adopt a null hypothesis of independent evolution between host and symbiont phylogenetic trees, but they interrogate the hypothesis using different forms of phylogenetic data. Briefly, in TreeMap congruence is assessed by mapping symbiont phylogeny onto host phylogeny and quantifying apparent host-symbiont cospeciation events (out of a maximum of *n* – 1 for a phylogeny of *n* symbiont taxa), versus other events (i.e. independent speciations, host switches, and sorting events). Host and symbiont phylogenies are considered congruent if the total is significantly greater than the distribution of scores observed when symbiont phylogeny is randomized a large number of times and mapped back onto host phylogeny, with the resultant *p* value indicating the level of significance1-4. In contrast, instead of proposing evolutionary scenarios, ParaFit adopts a distance matrix approach to quantifying congruence between host and symbiont phylogenies5. Host-symbiont data is transformed into a distance matrix using the sum of squared distances to quantify the similarity between host and symbiont trees. Together with an incidence matrix describing host-symbiont associations (i.e. ‘H-P links’), this is then compared to the distribution of values obtained by permutating the dataset to determine significance (i.e. the global test of host-symbiont cophylogeny ‘ParaFitGlobal’)5.

Both ParaFit and TreeMap presume fully resolved and accurate host and parasite phylogenies, and neither method takes phylogenetic uncertainly into account. Currently, many host and symbiont phylogenies are based on single gene analyses, and there is undoubtedly scope for improvements in phylogenetic accuracy by using multi-gene approaches and accounting for phylogenetic uncertainty. However, there is no evidence to suggest that any errors that may be present in individual host or symbiont phylogenies affect assessments of congruence in a systematic manner.

Phylogenetic congruence arising as a result of cospeciation, and phylogenetic congruence arising via other means that can produce a phylogenetic signature similar to cospeciation (such as successive host-switching across host phylogeny), are not differentiated by Parafit or TreeMap, as this requires accurately dated host and parasite phylogenies, of which relatively few pairs are available currently. However, our analyses were not directed at unravelling among alternative scenarios underlying host and symbiont speciation, but rather towards considering broadscale patterns in phylogenetic congruence.

**Data collection**

*Literature search*

A literature search was performed using Google Scholar on 19th March 2019, which identified 368 citations of the original paper for TreeMap2, and 332 citations for the original paper for Parafit5, resulting in a total of 700 citations that were screened to extract metrics for inclusion in our meta-analysis. Articles that did not contain cophylogenetic analyses were immediately excluded. Studies focussing at the population level were also excluded, as these do not represent true cophylogenetic analyses at the macroevolutionary level. Additionally, studies that included less than 4 taxa were excluded from consideration, as these do not provide sufficient power for inclusion in the meta-analysis, while studies that did not report the test statistic for congruence were also necessarily excluded. After exclusions were applied, a sample of 232 separate host-symbiont cophylogenetic studies remained, upon which our meta-analysis is based (Extended Data Table 1).

*Explanation of metrics*

Metrics collected for this study are described below, with reference to Extended Data Table 1.

A short citation of each study was recorded under ‘*authors*’, and the year of publication was recorded in ‘*year*’. Hosts and symbionts were classified broadly according to Linnean taxonomy for ‘*host\_tax\_broad’* and ‘*symbiont\_tax\_broad*’ as either: invertebrate, vertebrate, microbe, or plant. Hosts and symbionts were also classified finely for ‘*host\_tax\_fine’* and ‘*symbiont\_tax\_fine’* as: fungus, virus, protist, bacterium, plant, invertebrate or bird..

We adopted the mode of symbiosis and mode of transmission specified by the authors in each individual study considered for ‘*symbiosis*’ and ‘*mode\_of\_transmission\_broad*’. In cases where either mode of symbiosis or mode of transmission were not directly specified by authors, we consulted the literature for clarification. In a small number of studies restricted to bacterial intracellular symbionts, the mutualism-parasitism distinction was not defined by the authors and either no further information was available, or a symbiont was cited in the literature as being both a mutualist or a parasite, depending on which study was considered. The nature of the relationship between bacterial intracellular symbionts and their hosts is complex, and in some cases they may display both beneficial and detrimental effects simultaneously6. In conflicting cases or where data were absent, we classified bacterial intracellular symbionts as mutualists as the vast majority of available references indicated this. Meanwhile, if authors did not explicitly state mode of transmission for bacterial intracellular symbionts, we assumed a vertical mode of transmission, on the basis of the majority of available references.

The total number of host tips with a link to a symbiont taxon were summed to provide ‘*host\_tips\_linked*’, which in a very few cases was corrected to remove multiple sampling of the same host species, to provide ‘*host\_tips\_linked\_corrected*’. The total number of symbiont tips with a link to a host taxon were summed to provide ‘*symbiont\_tips\_linked*’, while the total number of individual links between hosts and symbionts was recorded as ‘*total\_host\_symbiont\_links*’. If all symbionts in a phylogeny were strict specialists, such that each one had a single link to a single host, ‘*total\_host\_symbiont\_links*’ would simply equal ‘*symbiont\_tips\_linked*’. However, because symbionts are often associated with more than one host, the value of ‘*total\_host\_symbiont\_links*’ was often higher than the total number of symbionts included in a study. Thus, a measure of symbiont generalism was captured using ‘*host\_range\_link\_ratio*’, defined as ‘*total\_host\_symbioint\_links*’ divided by ‘*symbiont\_tips\_linked*’, providing the mean number of host-symbiont links observed per symbiont taxon, with the measure increasing with increasing generalism. An alternative estimate of symbiont host specificity was captured using ‘*host\_range\_taxonomic\_breadth*’, which considers Linnean taxonomic rank, and was calculated by assigning an incremental score to successive host taxonomic ranks per symbiont in turn (i.e. single host species = 1, multiple host species in the same genus = 2, multiple host genera = 3, multiple host families = 4, multiple host orders = 5), summing the total score across all symbionts, and dividing by ‘*symbiont\_tips\_linked*’ (i.e. the total number of symbionts). Consequently, ‘*host\_range\_taxonomic\_breadth*’ increases with symbiont generalism, such that symbiont phylogenies containing symbionts capable of infecting hosts from a wide range of taxonomic ranks are assigned a greater score.

The number of phylogenetic permutations performed by authors during cophylogenetic analyses was recorded as ‘no\_randomizations’. The resultant *p* value from each study was recorded as ‘*p\_value*’, whereby observed *p* values decrease with a decreasing likelihood of observing host-symbiont cophylogeny by chance alone (i.e. as calculated during permutation tests of host-symbiont phylogenies performed by authors during TreeMap or ParaFit analyses).

**Effect size**

We used *p* values obtained from randomization tests that are implemented in ‘TreeMap’ and ‘ParaFit’ as measures of incongruence. These *p* values were converted into *r* and its transformation *Zr*. We can calculate *requivalent* via *t* values with *df* = *N* - 2 from *p* values (one-tailed)10, and then, also obtain *Zr*equivalent from *r*equivalent, as follows:

where *N* is sample size, which is, in our case, the sum of the numbers of host and symbiont species included in a randomisation test.

Importantly, in many studies, host-symbiont phylogenetic permutations (randomisations) performed by authors was, for example, 999 (or 1000), which limited the lowest observable *p* value statistic to *p* ≤ 0.001. When *p* = 0, we used *p* = 0.001 for 999 and 1000, *p* = 0.0001 for 9999 and 10000 and so on (Extended Data Table 1) and this created ceilings for ~32% of the effect sizes included our analysis, i.e. these effect sizes could have been larger. Since the true likelihood of achieving the observed number of co-speciations by chance alone can be considerably lower than, for example, *p* = 0.001 (with randomizations = 1000 and *p* = 0), particularly in cases of perfect host-symbiont congruence involving many taxa, our analyses below thus represent a conservative measure of host-symbiont phylogenetic congruence, and should be considered the lower bound of congruence that exists for the considered studies (sensitivity analyses relevant to this point are discussed below).

**Meta-analysis and meta-regression**

All statistical analyses were conducted using R version 3.6.111. We used multilevel (random-effects) meta-analytic and meta-regression models12, because multiple effect sizes were obtained from some studies (i.e., study IDs were included as a random factor in the models to account for non-independence). All meta-analytic and meta-regression models were implemented using the function, *rma.mv* in the R package, *metafor* version 2.1-013. We used multilevel meta-analytic and meta-regression models and associated heterogeneity measures (*I*2) [REF]. Notably, we calculated not only 95% confidence intervals of all the estimates associated our meta-analysis but also 95% prediction intervals. The prediction interval indicates the probability of obtaining an effect size value from a new study within the interval [REF]. All model specifications, model selection procedures and associated coding are found in our electronic supplementary materials (ESM) where we also provide statistical justification for merging datasets for both ‘TreeMap’ and ‘ParaFit’.

**Publication bias and sensitivity analysis**

We used the three types of publication analysis tests, but all of these tests were a multilevel model version of the original tests: 1) contour-enhanced funnel plots[REF] of residuals [REF], 2) a type of Egger regression [REF]., and 3) a regression-based time-lag bias test[REF]. These analyses indicated our results are unlikely to be affected by publication bias and time-lag bias. Instead, our estimates of category-wise effects or contrasts between categories (e.g. parasites *versus* mutualists) are likely to be underestimated (for the details, see ESM) because of the ceiling set by the number of randomizations.

Therefore, we conducted two sensitivity analyses to show that our underestimation of effect sizes did not our main conclusions: 1) linear mixed models testing if there was no difference in the number of randomizations among categories of moderators and 2) generalized linear mixed models with the binomial family with the logit link showing that categories with high effect sizes were affected by the ceiling effect (both fitted by the lme4 package, version 1.1-21). The second analysis also indicated that the higher the effect size estimate for a particular category (e.g. the mutualism group or the vertically transmitted symbionts group), the more underestimated, the estimate (for all associated results, see ESM).

**References**

1 Page, R. D. Component analysis: a valiant failure? *Cladistics* **6**, 119-136 (1990).

2 Page, R. D. Parallel phylogenies: reconstructing the history of host‐parasite assemblages. *Cladistics* **10**, 155-173 (1994).

3 Charleston, M. A. Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Math Biosci* **149**, 191-223 (1998).

4 Charleston, M. A. & Perkins, S. L. Lizards, malaria, and jungles in the Caribbean. *Tangled Trees: Phylogeny, Cospeciation and Coevolution*, 65-92 (2003).

5 Legendre, P., Desdevises, Y. & Bazin, E. A statistical test for host–parasite coevolution. *Systematic biology* **51**, 217-234 (2002).

6 Zug, R. & Hammerstein, P. Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts. *Biol Rev Camb Philos Soc* **90**, 89-111, doi:10.1111/brv.12098 (2015).

7 Saikkonen, K., Faeth, S. H., Helander, M. & Sullivan, T. Fungal endophytes: a continuum of interactions with host plants. *Annual review of Ecology and Systematics* **29**, 319-343 (1998).

8 Palmer, T. M. *et al.* Breakdown of an ant-plant mutualism follows the loss of large herbivores from an African savanna. *Science* **319**, 192-195, doi:10.1126/science.1151579 (2008).

9 Cheney, K. L. & Cote, I. M. Mutualism or parasitism? The variable outcome of cleaning symbioses. *Biol Lett* **1**, 162-165, doi:10.1098/rsbl.2004.0288 (2005).

10 Rosenthal, R. & Rubin, D. B. Requivalent: A simple effect size indicator. *Psychological methods* **8**, 492 (2003).

11 Team, R. C. R: A Language and Environment for Statistical Computing. (2018).

12 Nakagawa, S. & Santos, E. S. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology* **26**, 1253-1274 (2012).

13 Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J Stat Softw* **36**, 1-48 (2010).