**The Taxonomic Distinctness of Benthic Functional Diversity**

Explore whether a is that as more traits are included in analysis, functional diversity approaches taxonomic diversity

**METHODS**

*Data*

*Functional Trait data*

Data from Tyler & Webb compiled

Data selection

The final dataset was limited to species for which information for at least four traits was available and consisted of 202 species with data across a maximum 13 traits (table 1).

*Taxonomic data*

Data on the taxonomic classification of species were sourced from WoRMS, some missing data was supplemented from the biotic database.

***Data Analysis***

Firstly a functional dendrogram was produced. We selected the dendrogram method of representing functional diversity (Petchey & Gaston (Petchey:2002ug}, {Petchey:2006ck}) because it represents a hierarchical partitioning of functional space. As such it shares both conceptual and computational parallels with taxonomic classification and taxonomic measures of diversity {Petchey:2006ck} {Clarke:1998um}.

Briefly, a distance matrix of species dissimilarities is computed across all selected functional traits. Prior to analysis categorical variables were coded as a series of binomial variables, one category for each modality while continuous data were standardised to have a mean of zero and SD of 1. As a result of the mixed nature of trait data (mixture of binomial and continuous variables) the gower distance between each species was calculated ({Petchey:2007dw}{Podani:2006fo}). Binomial variables were treated as asymmetric, ie more importance was placed on species sharing the presence of a trait than sharing the absence of a trait.

A clustering algorithm (UPGMA) was then used to produce the functional dendrogram. The species identity of members of each node were determined and the taxonomic distinctness (Δf+) of the node calculated. (FIGURE?)

Taxonomic distinctness, detailed in {Clarke:1998um} is calculated in a similar way as functional diversity in that taxonomic distance is first calculated between species in an assemblage and then a taxonomic dendrogram is produced. Taxonomic distinctness Δ+, the special case in which only presence of species is used, is then the average path length connecting all pairs of species and represents the average relatedness of two species drawn at random from the assemblage.

The relationship between Δf+ and node height (hf) was then determined by fitting a linear model with Δf+ as the response variable and by calculating the strength of their correlation.

An important difference between the two metrics (FD vs Δ+) is that Δ+ represents a standardized metric in which addition of a closely related species would actually reduce overall Δ+ in contrary to FD in which addition of a species either produces no change or increases overall FD. It has been well established therefore that the variance in Δ+ is strongly related to the number of species in an assemblage, in the case of our analysis, the number of species included in each node. As such we felt it was important to compare any trend in FD dendrogram structure to any trend in node Δ+ underlying the taxonomic dendrogram. In an analogous process to that described for the FD dendograms, Δ+ of each node in the taxonomic dendrogram (Δt+) was also calculated and compared to height of node (ht)

Normal behaviour of cluster TD for a complete set of species going up a taxonomic dendrogram is that TD generally increases as taxonomic aggregation increases. However, as a result of the standardisation of our selected measure of taxonomic distinctness (average taxonomic branch length connecting selected species) taxonomic distinctness of similarly related species will decrease as more species are added despite the consistency in the relatedness of the species making the measure sensitive to number of species included in the analysis and branch size. As such the average trend of cluster ABL was also calculated going up taxonomic trees as the inclusion and exclusion of species is likely to affect it allowing comparison of the FD trend to the underlying taxonomic trend.

To determine the effects of the number of traits (*t*) and species (*s*) on the relationship, the analysis was performed under three conditions. Firstly, both number of traits and species were varied simultaneously according to data availability. Variables were ordered in order to maximize the number of species available in the analysis over all trait combinations as traits were sequentially removed (fig. 1). This was chosen to reflect the trade off between number of species and trait data availability. To tease apart the effect of number of traits and species individually, the same procedure was followed under two further conditions. The number of traits included was varied (between 1 & 10) whilst keeping the number of species constant. This was restricted to a subset of species (*s* = 65) for which data was available for all traits investigated. Similarly, the number of species was varied whilst holding the number of traits constant. The number of traits was held at 8 whilst the number of species included was varied from 15 to 91. The relationship between the number of species and traits included and the strength of the correlation between Δf+ and hf was then examined.

**RESULTS**

The relationship between hf (cluster dissimilarity) and Δf+ is funnel shaped with the fitted linear model generally explaining very low levels of variance (generally much lower than <15%) despite occasional significance. The funnel shape demonstrates high levels of variance at lower branch lengths, reminiscent of the relationship between Δf+ and *s* {Clarke:1998um} indicating that the trend is most likely driven mainly by the number of species included in each branch which generally increases moving towards the root of the dendrogram.

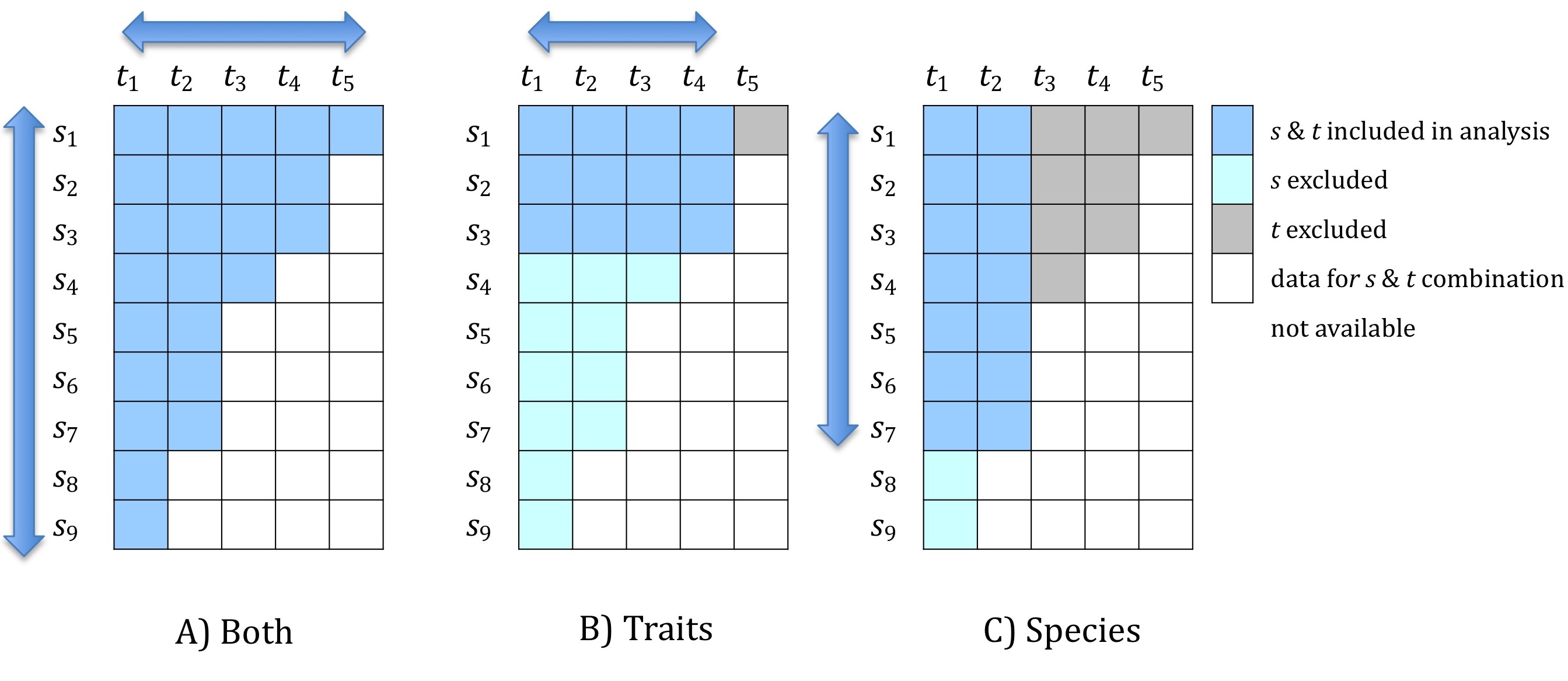
A general trend emerging from variation of trait / species included in the analysis is a general flattening of the slope in the relationship as fewer traits / more species are included in the analysis (fig. 2) coupled with a decrease in the strength of the correlation of branch Δf+ and hf (fig.5a & b). This trend appears to be primarily driven by variation in number of traits included (fig. 3, fig. 5d) as variation in number of species, at least in the range explored here, appears to have minimal effects on both the slope and correlation of the relationship (fig. 4, fig. 5c).

Interestingly an opposite trend is observed in the slope of the relationship between Δt+ and ht where the slope appears to become steeper and the change being mainly driven by variation in *s* (fig. 2, 3, 4).

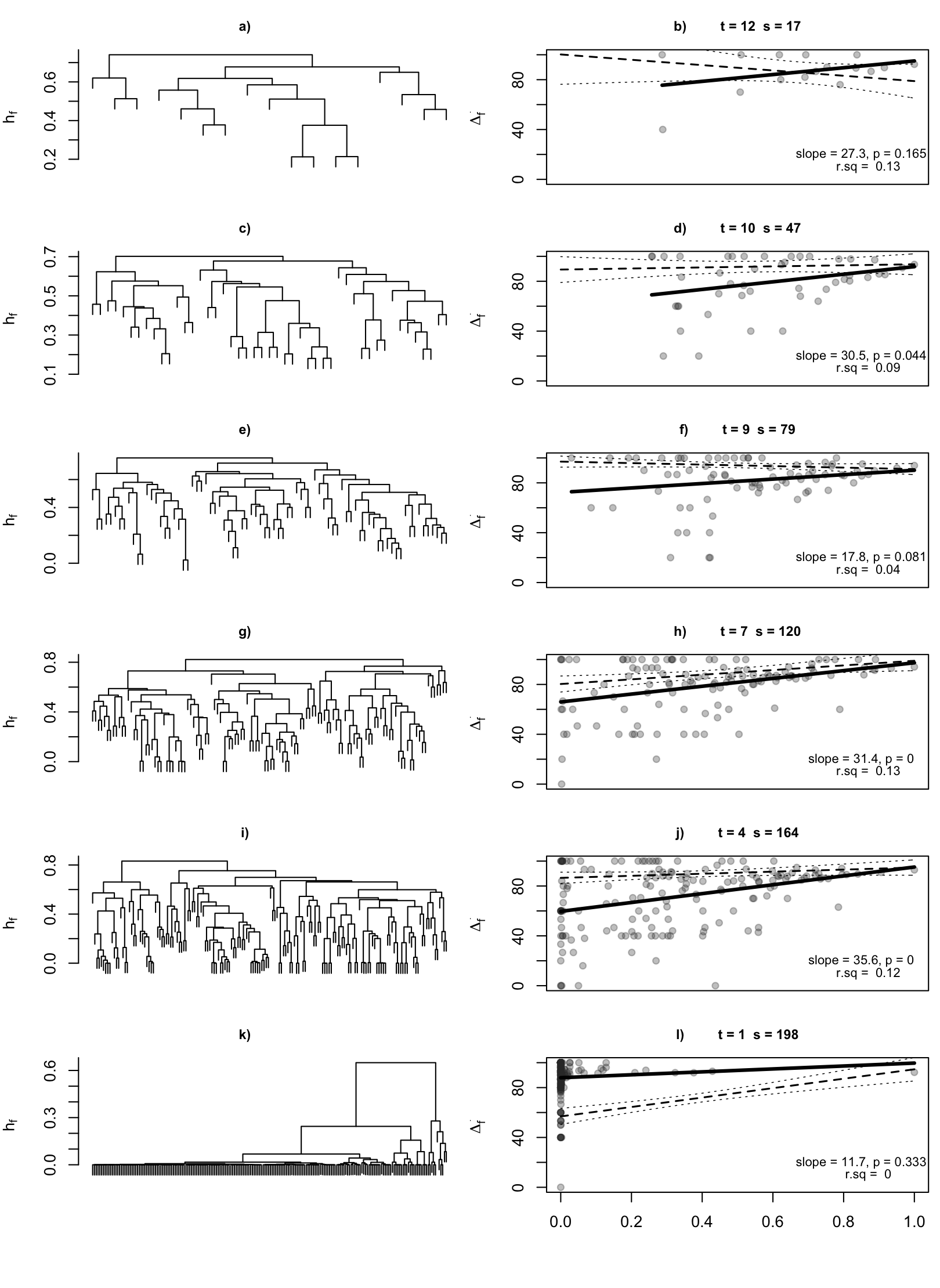
Much of the functional dendrogram structure appears to be minimally affected by both addition of species and / or traits indicating relative stability in the functional classification without large scale reorganization of the tree under the varying analysis conditions. The glaring exception is when only a single trait – body size – was used for functional classification (fig 2k, 4k). This produced a dendrogram of a few long branches with the main partitioning of species being very shallow and occurring near or at the leaves of the tree. It is likely that should an analysis could be highly to sensitive to inconsistencies between metric available to describe trait for each species (ie. maximum, average etc) and differences in sources of information and should therefore be avoided.

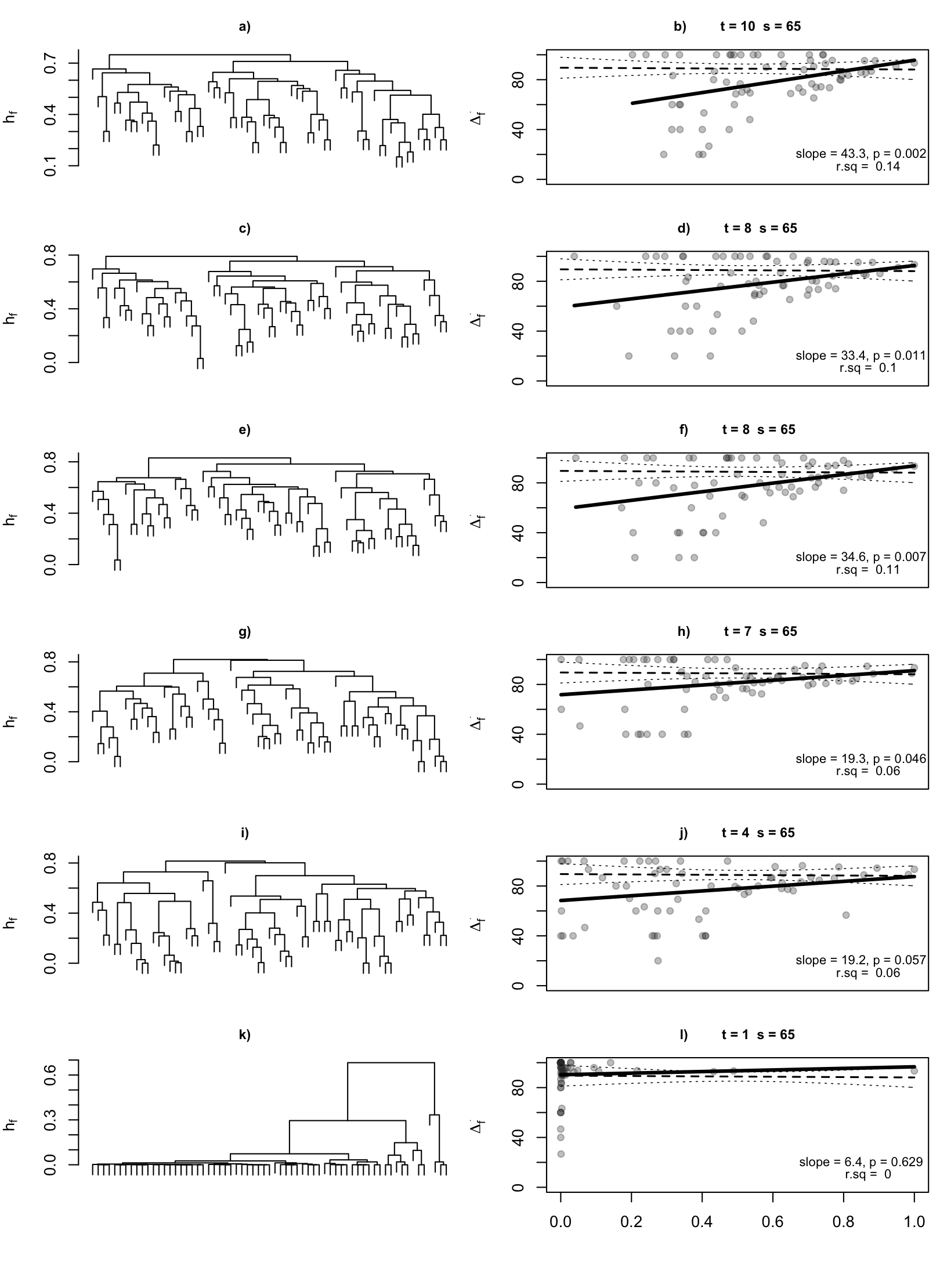
Interestingly a large proportion of nodes (70-80%) across the vast majority of analyses are completely taxonomically distinct ie. they are composed of species of completely different phyla.**Table 1** Details of Trait data. The ordering from top to bottom reflects the sequence of inclusion of variables. Data availability indicates the number of species for which trait data were available and complete cases indicates the number of species for which data for the trait and all previous traits in the sequence are available.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Trait** | **Data availabi- lity** | **Complete cases** | **Type** | **Range / modalities** |
| ***Body Size*** | 204 | 198 | Continuous | 0.4 - 90 cm |
| ***Movement Method*** | 195 | 188 | Categorical | burrower, crawler, floater, motile, permanent attachment, sessile, swimmer, temporary attachment, tube-dwelling |
| ***Feeding Method*** | 195 | 179 | Categorical | active suspension feeder, browser, deposit feeder, filter feeder, grazer, passive suspension feeder, predator, scavenger, surface deposit feeder, suspension feeder |
| ***Sociability*** | 183 | 164 | Categorical | colonial, gregarious, solitary, sometimes gregarious |
| ***Habitat*** | 133 | 120 | Categorical | demersal, epibenthic, epifaunal, epilithic, epiphytic, epizoic, infaunal |
| ***Diet*** | 120 | 101 | Categorical | algae, benthic organisms, detritus, macroalgae, meiobenthic organisms, micro-organisms, microalgae, omnivore, phytoplankton, plankton, sediment particles, suspended particles/material, zooplankton |
| ***Migration*** | 142 | 91 | Categorical | irregular/single migration, no evidence, non-migratory, regular |
| ***Reproductive Frequency*** | 116 | 79 | Categorical | annual, biannual, semelparous |
| ***Larval Feeding Strategy*** | 113 | 65 | Categorical | lecithotrophic, planktotrophic |
| ***Lifespan*** | 146 | 47 | Continuous | 1.5 - 60.5 |
| ***Reproductive Period*** | 111 | 46 | Categorical | episodic, protracted |
| ***Developmental Mechanism*** | 105 | 17 | Categorical | direct development, larval development, benthic larval development, pelagic larval development, oviparous, fission |
| ***Reproductive Timing*** | 91 | 15 | Categorical | 1 to 12 |

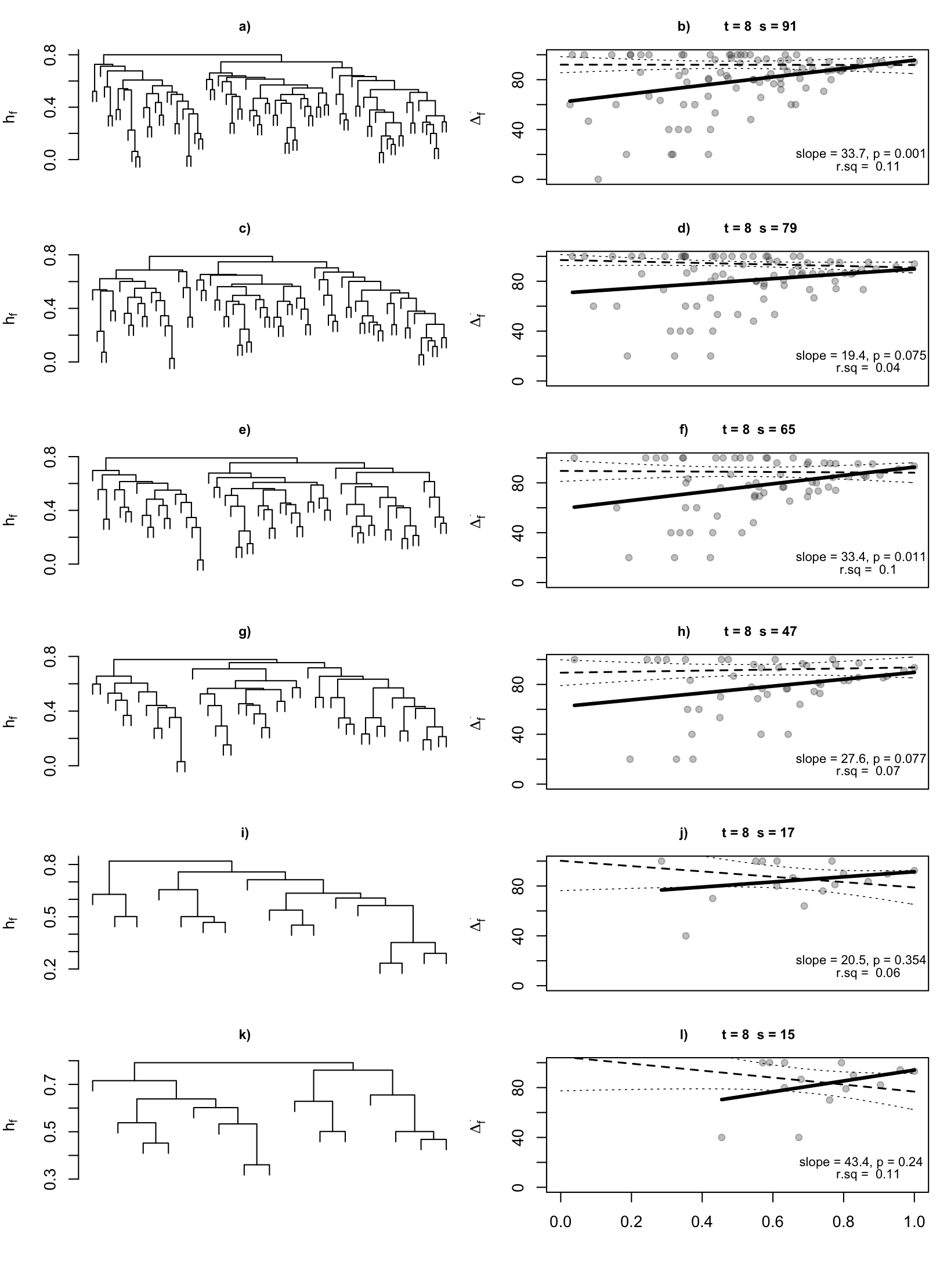


**Figure 1** Representation of ordering of species (*s*) and traits (*t* ) and selection procedure for each condition. A) Both species and trait number included in analysis is varied at each iteration resulting in all data being used in complete analysis B) Number of species included in analysis is held constant while trait inclusion is varied C) Number of traits is held constant while species are varied. Sequence of trait / species exclusion is fixed and follows ordering imposed by data availability.

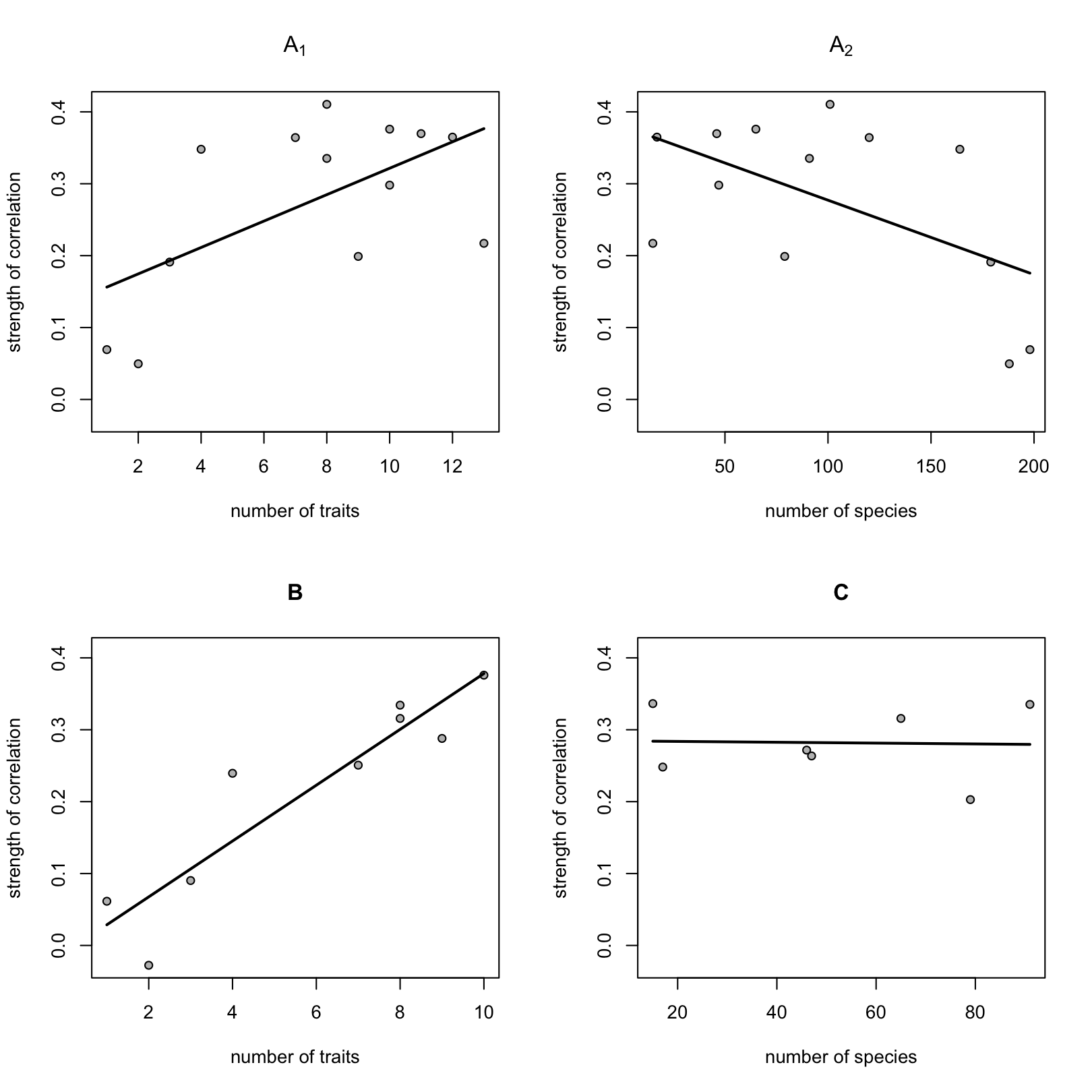
**Figure 2** Results from analysis A in which both number of traits (*t* ) and species (*s*) included was varied for selected *t* and *s* combinations. Left panels show FD dendrograms and right panels show corresponding relationship between Δf+ and hf. Solid line indicates linear model fit while dashed line indicates model fit and 95% confidence intervals of relationship between Δt+ and ht. hf in right panels is standardized to values between 0 & 1.



**Figure 3** Results from analysis B in which species were held constant (*s* = 65) while number of traits (*t* ) included was varied. Left panels show FD dendrograms and right panels show corresponding relationship between Δf+ and hf. Solid line indicates linear model fit while dashed line indicates model fit and 95% confidence intervals of relationship between Δt+ and ht. hf in right panels is standardized to values between 0 & 1.



**Figure 4** Results from analysis C in which traits were held constant (*t* = 8) while number of species (*s* ) included was varied. Left panels show FD dendrograms and right panels show corresponding relationship between Δf+ and hf. Solid line indicates linear model fit while dashed line indicates model fit and 95% confidence intervals of relationship between Δt+ and ht. hf in right panels is standardized to values between 0 & 1.



**Figure 5** Strength of correlation between Δf+ and hf+ A1) as a function of *t* when both *t* and *s* were varied, A2) as a function of *s* when both *t* and *s* were varied, B) as a function of *t* when only *t* was varied and C) as a function of *s* when only *s* was varied.

**Table 2** Linear regression coefficientsof relationships depicted in figure 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **A1 (*t*)** | **A2 (*s*)** | **B (*t*)** | **C (*s*)** |
| ***slope*** | 0.018 | -0.001 | 0.039 | 0.000 |
| ***p*** | \* | \* | \*\*\* | NS |
| ***r.sq*** | 0.36 | 0.31 | 0.85 | 0.00 |