Testing taxonomic distinctness against a master list

Wide-ranging biogeographic studies, and particularly historic data, are often restricted to simple species lists. Even where quantitative information exists, it is rarely from sampling protocols that have been standardised with respect to sampling effort over the whole data. Where sampling is so exhaustive that the asymptote of the species-area curve is approached, then it may be valid to compare diversity status by the length of these lists (species richness S), but this is not often the case (in marine science, certainly). As is well known, S is heavily sampling effort dependent so, if sampling effort is variable and unknown, any valid statements about richness appear problematic. However, the two relatedness measures discussed earlier, average taxonomic distinctness (AvTD, $\Delta^{\scriptscriptstyle +}$) and variation in taxonomic distinctness (VarTD, $\Lambda^{\scriptscriptstyle +}$), can not only be calculated from simple species lists, with the added knowledge of their Linnaean (or other) classification, but also possess a robustness to the varying number of species S in the lists. To be more precise, in different-sized sublists generated by random sampling from a larger list (simulating the action of sampling with variable effort) their mean values are unchanged. This suggests that it is valid to compare $\Delta^{\scriptscriptstyle +}$ (or $\Lambda^{\scriptscriptstyle +}$) over historic time or biogeographic space scales, under conditions of variable sampling effort. (Note that the indices are *average* not *total* measures, and orthogonal to species richness – along a third PC diversity axis, would be one way of thinking of it – and therefore an *addition* to S, rather than a substitute for it, in cases where sampling effort is controlled and S can be validly compared.)

Furthermore, a test can be constructed for the null hypothesis that a species list from one locality (or time) has the same taxonomic distinctness structure as a ‘master’ list (e.g. of all species in that biogeographic region) from which it is drawn. This is again by simple randomisation: given there are s species observed in a particular sample, make repeated drawings at random of s species from the master list and compute $\Delta^{\scriptscriptstyle +}$ for each drawing, building up a histogram and a 95% probability range of values of $\Delta^{\scriptscriptstyle +}$ expected under the null hypothesis, with which the true $\Delta^{\scriptscriptstyle +}$ can be compared. Values below the lower probability limit suggest a biodiversity that is ‘below expectation’. This can be carried out for a range of sublist sizes and the limits plotted against *s*, to give a 95% *funnel plot* of expected values (the funnel arises from uncertainty being greater for smaller sublists). This can be repeated for VarTD ($\Lambda^{\scriptscriptstyle +}$), giving a second set of histograms and funnel. Together, the true $\Delta^{\scriptscriptstyle +}$ and $\Lambda^{\scriptscriptstyle +}$, and the simulated values obtained by drawing their number of species from the master list, can be plotted on a single (*x*,*y*) scatter plot. Probability regions (‘egg-shaped’ contours, called *ellipse plots* since they are back-transformed ellipses) covering 95% of the simulated values can be drawn for a range of sample sizes, and the true ($\Delta^{\scriptscriptstyle +}$, $\Lambda^{\scriptscriptstyle +}$) compared with their appropriate contour.

Revision #1
Created 5 November 2024 20:39:34 by Arden

Updated 5 November 2024 20:51:16 by Arden