ONLINE SUPPLEMENT 16.5

ADDITIONAL GUIDANCE ON MULTIDIMENSIONAL SCALING (MDS) – INTERCITY DISTANCE EXAMPLE

When first learning to understand and apply MDS, it is often helpful to begin with a non-biological example. We therefore present a tutorial using U.S. cities and the distances between them. Consider the locations of 10 U.S. cities: Miami (FL), Atlanta (GA), Richmond (VA), Newark (NJ), Chicago (IL), Austin (TX), Denver (CO), Reno (NV), San Francisco (CA), and Portland (OR). Measure straight-line or 'as the crow flies' distances between each pair of cities and you will obtain a matrix of inter-city distances very similar to the values shown in the upper-right, gray cells of Figure S16.5A.[1] Now imagine that each city is a push-pin to insert into a large corkboard, with inter-city distances to be simulated by strings that are proportional in length to the measured distances. Beginning with any two cities, it is simple to insert their respective pins so that the correct inter-city distance is maintained; just stretch the corresponding string between the two pins. Then add a third city/pin to the board, ensuring that all three of the inter-city distances are preserved. As more pins are added, it becomes increasingly difficult to maintain the correct inter-city distances, but work at it long enough and a solution with each of the inter-city strings stretched tight between the 10 pins will be obtained. This process of iteratively moving points around to preserve the *a priori* specified distances between them is, in essence, MDS.

A computer-generated,[2] two-dimensional[3] MDS plot of the 10 cities is presented in Figure S16.5B. Is this a good representation of the inter-city distances? The very small stress value shown in Figure S16.5B confirms that it is. (Remember that stress is our index of the 'goodness-of-fit' between empirical data (distances between cities in this example) and the

[1] But remain cognizant of the fact that these inter-city distances are fundamentally different than the dissimilarity values shown in the main text among fish assemblages; the inter-city distances are true Euclidian distances, measured within a two-dimensional Cartesian coordinate system, while the fish assemblage dissimilarities are abstract indices of distance, derived from information theory.

[2] We used SPSS (version 22) software.

[3] Two dimensions (i.e., X and Y axes) are sufficient in this example, but MDS can incorporate many more axes. Of course, plotting and interpreting more than three axes is difficult.

corresponding distances within the MDS plot; see main text). However, a more intuitive demonstration of the agreement between the measured inter-city distances and the MDS results is obtained by superimposing a rescaled map of the U.S. on the MDS plot, as shown in Figure S16.5C. The superimposed state boundaries confirm that the MDS plot is an excellent representation of the measured inter-city distances, with each point in the MDS plot located close to its true location. However, the map also shows that the X and Y axes of the MDS ordination plot must be interpreted with caution. The U.S. map has clearly been flipped along the major horizontal axis and must be rotated clockwise by *c*. 12 degrees to fit the map to the MDS plot. This demonstrates a critically important point: the axes generated in MDS are not, of themselves, biologically meaningful or interpretable. Only the *relative locations* of points (i.e., samples) within the MDS plot are meaningful. Flip or rotate the ordination plot in any way that you like and it will not affect the basic interpretation of the plot.

If we wish to take a more conservative approach, we can apply nonmetric MDS using the ranked inter-city distances shown in the lower-left, black cells of Figure S16.5A. The smallest inter-city distance (Reno to San Francisco) receives rank 1, progressing through to the largest distance (rank 45) between Miami and Portland. The resulting nonmetric MDS plot is shown in Figure S16.5D. Although the relative positions of the cities are preserved, distortion has increased: Richmond has migrated to the West Virginia panhandle while Portland, Reno, and San Francisco have all converged on the Mount Shasta region in northern California. These modest distortions are indicated by the new, higher stress value (0.054). Even so, the overall representation of cities is not grossly distorted and the nonmetric MDS plot would likely be useful for preliminary or exploratory analyses of the city data.

A

	Mia	Atl	Rich	New	Chi	Aus	Den	Reno	San	Port
Mia	0	977	1,334	1,758	1,916	1,790	2,776	3,970	4,164	4,351
Atl	7	0	746	1,184	942	1,318	1,952	3,206	3,440	3,492
Rich	14	4	0	456	998	2,029	2,398	3,659	3,923	3,819
New	19	10	2	0	1,133	2,607	2,607	3,839	4,119	3,916
Chi	21	6	8	9	0	1,564	1,476	2,707	2,986	2,824
Aus	20	13	23	28	17	0	1,235	2,246	2,406	2,739
Den	31	22	25	28	15	11	0	1,266	1,524	1,576
Reno	42	34	37	39	29	24	12	0	305	700
San	44	35	41	43	33	26	16	1	0	866
Port	45	36	38	40	32	30	18	3	5	0

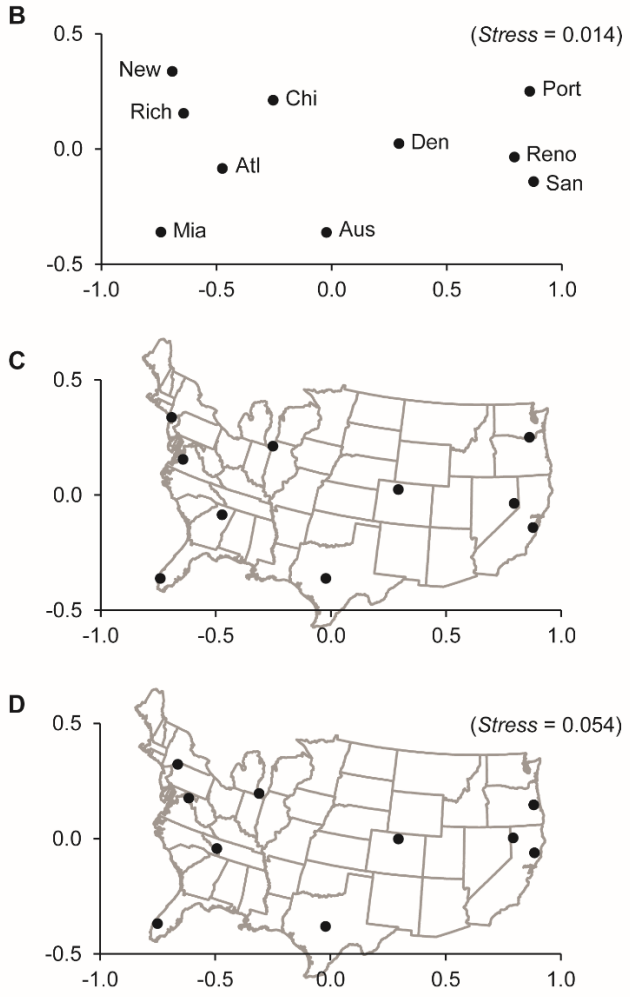

Figure S16.5 MDS results for 10 U.S. cities: Miami, FL (Mia); Atlanta, GA (Atl); Richmond, VA (Rich); Newark, NJ (New); Chicago, IL (Chi); Austin, TX (Aus); Denver, CO (Den); Reno, NV (Reno); San Francisco, CA (San); and Portland, OR (Port). Panel A lists linear, Euclidian distances between cities (in km) at upper-right (gray cells) and ranked distances at lower-left (black cells). Panel B illustrates a two-dimensional metric MDS plot of inter-city distances. Panel C is identical to Panel B, with a map of U.S. states superimposed. Panel D shows a nonmetric MDS plot for the 10 cities.