**Description of the empirical datasets of the manuscript entitled “Scaling of biological rates with body size as a backbone in the assembly of metacommunity biodiversity”**

The analyses performed in this study are based on two datasets called: (1) “Pond\_xy”; (2) “Dataset”. The file “Pond\_xy” provides information about the spatial position (X-Y coordinates) of the 18 temporal ponds that comprise the empirical metacommunity and from which biological samples were collected. From these coordinates, the geographic distance between each pair of ponds is estimated. Then, it is used to simulate the assembly of the metacommunity with the coalescent-based metacommunity model. Geographic distance between pair of ponds affects the probability that a new recruit comes from a neighbour community.

The file “Dataset” provides the biological information collected in the empirical metacommunity. It is used to estimate the alpha, beta and gamma diversity, and mean body size, of each taxon included in the empirical analysis. Each row of the file corresponds to the information of an animal individual (i.e., macroinvertebrates, killifish and anurans). The columns of the file indicate the pond ID from which the individual was taken (column “Pond”), the ID of the sample unit (column “Sample”), the species (or morphospecies) name of the individual (Column “Species”), the individuals body size (column “Biovol”), the taxonomic order, family and genus of the individual (columns “Order”, Family” and “Genus” respectively), and a categorization of the diet of the individual (column “Diet”).