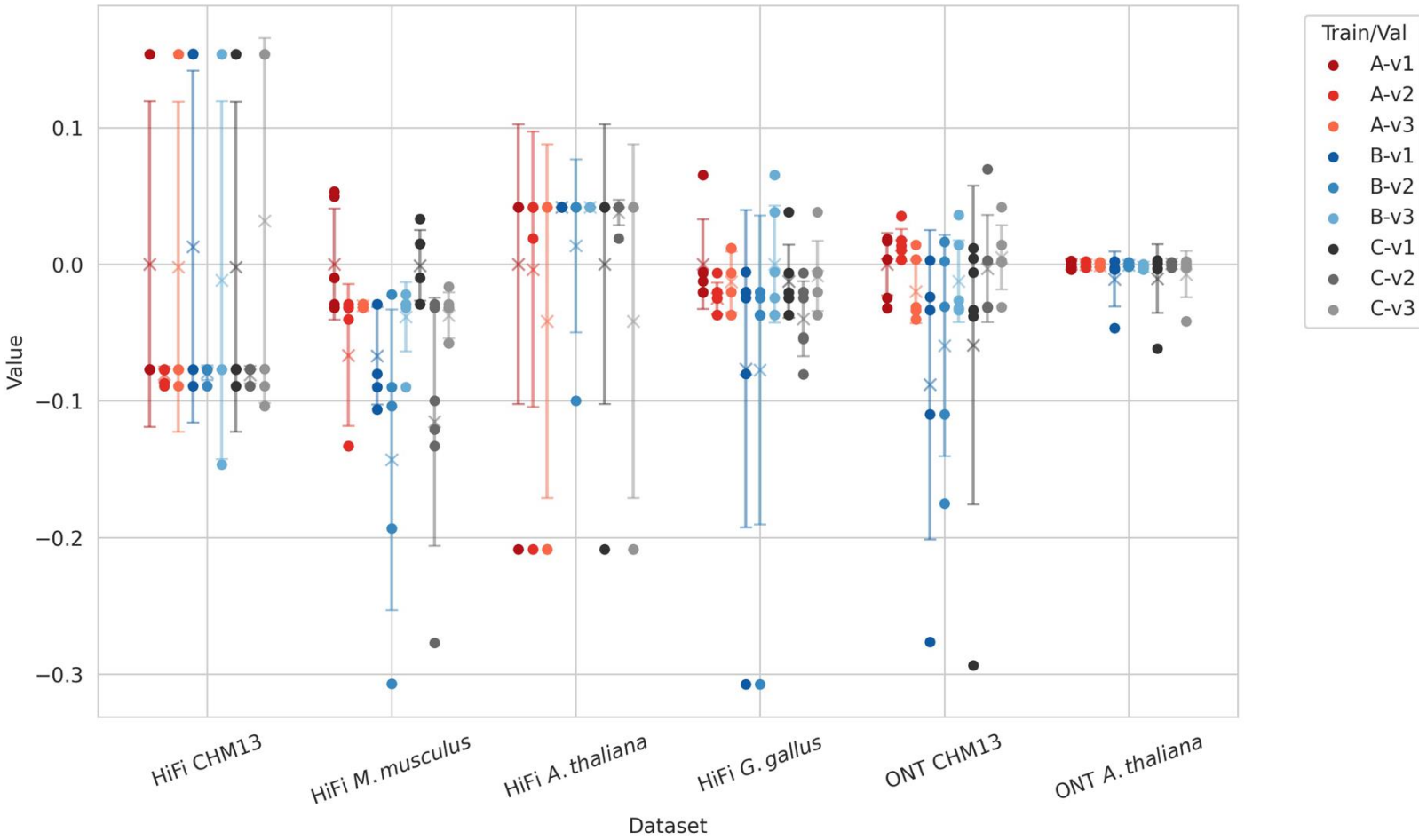

Supplemental Figure S9: Influence of different training datasets on the contiguity of the assemblies. For each choice of training and validation sets, a model was trained six times with six different random seeds. The result of each run is represented with a dot, while the mean is represented with a cross. The error bars show one standard deviation. The value shown is the mean NGA50, normalized to the mean value of the A-v1 training/validation set. Choices A-v1, B-v2, etc. are all described in Supplemental Note S4.