**Supplemental Figure Legends**

**Supplemental Figure 1:** Summary of geospatial variables. A) Pairwise correlation graph for each geospatial variable with graphs in the lower triangle, correlation values in the upper triangle, and density/histogram plots in the diagonal. B) Principal component analysis (PCA) of the geospatial variables, represented as blue loadings. Points are colored by the sampled river.

**Supplemental Figure 2:** Summary of geochemical variables. A) PCA of geochemical variables, represented as blue loadings. As with Supplemental Figure 1, the points are colored by the sampled river. B) Geochemical variable plotted against stream order. Red asterisks indicate significance via Kruskal-Wallis test. C) Correlation graphs comparing select geospatial variables against geochemical variables to identify potential co-correlation. Spearman correlation statistics are indicated in red in each facet.

**Supplemental Figure 3:** Ecological null modeling overview. A) Density graph of βNTI values for the entire dataset. B) Comparison of pairwise βNTI values against selected geospatial (top) and geochemical (bottom) variables. Spearman-based mantel correlation statistics are indicated in red text for each facet. Full correlations can be found in Supplemental Table 1.

**Supplemental Figure 4:** Comparison of average βNTI calculated across all samples or within stream order against urban coverage (A), forest coverage (B), wetland coverage (C), agricultural coverage (D), shrub coverage (E), and NPOC (F). The generalized additive model fit is represented in blue with a 95% CI in gray, though no fit was significant, so no statistics are provided.

**Supplemental Table Legends**

**Supplemental Table 1**: Statistics for overall and within-stream order βNTI correlations with geospatial and geochemical variables within and across stream orders. This table includes Spearman-based Mantel correlations between βNTI variation and geochemical variables (Sheet 1) or geospatial variables (Sheet 2), but also includes Spearman correlations between average βNTI and geochemical variables (Sheet 3) or geospatial variables (Sheet 4).

**Supplemental Results and Discussion**

The specific pairwise βNTI comparisons were significantly related to many geochemical and geospatial variables. This metric relies on relational information to assess whether the differences between a pair of DOM assemblages developed due to deterministic (i.e., non-random) or stochastic (i.e., random) ecological processes. When βNTI values are significantly high or low (|βNTI| > 2), the differences or similarities between assemblages are larger than we would expect due to random chance due to the effects of ecological selection. When βNTI values are insignificant (|βNTI| < 2), the differences or similarities between assemblages mirror the random expectation indicating that dispersal-based processes or weak ecological processes generated the observed differences. Turning toward correlations between βNTI and various environmental parameters, we observed that relationships formed four rough groups.

Looking at the first group of βNTI relationships, we observe that Geospatial PC1 and associated variables (i.e., average precipitation, average temperature, forest coverage) are significantly related to increased determinism according to both overall and within stream order βNTI (**Supplemental Table 1, Supplemental Figure 3B**). This result reveals that the correlations observed between geospatial variables and the number of molecular formulas in **Figure 2** in the main text likely arose due to deterministic forces. Specifically, this suggests that climate derived (e.g., precipitation) and hydrological (e.g., contact time) variables impose a strong degree of selection on DOM assemblages, though the specific abiotic or biotic process is unclear.

The second group of variables account for geospatial variables not associated with Geospatial PC1, namely other types of land coverage (**Supplemental Table 1, Supplemental Figure 3B**). These variables were also correlated to the number of molecular formulas in **Figure 2** in the main text. These results indicate that land cover variables, in addition to common hydrological and climatic variation, play a role in the development of DOM. Danczak et al., (*20*) revealed that land cover was a key control on the functional diversity of DOM within this basin. These results indicate that the ecological controls extend beyond just interacting with transformations.

The third and fourth groups of variables are both sets of geochemical measurements that differ based upon their relationships across scales. The third group consists of geochemical variables which have greater or similar relationship strengths across stream orders than within stream orders. Specifically, NPOC and TN were the strongest correlations both within and across stream orders; conversely, NO2 was the weakest and the only non-significant relationship within stream orders or throughout the entire basin. These variables point to geochemical parameters that likely impact DOM selection consistently regardless of the scale of comparison. Given the patterns in other analyses, NPOC might be acting as a significant variable in determining the ecological processes impacting riverine DOM. There are a three potential mechanisms behind this pattern: **1)** absolute NPOC concentration stimulates undirected biological activity, therefore the DOM becomes more diverse (e.g., when nutrients are a limiting factor, microbes become less thermodynamically driven) (*17*); **2)** the sources of the NPOC are deterministically changing (e.g., we see that NPOC is related with land cover; different land cover has different plants, therefore carbon deterministically changes); **3)** larger NPOC concentration differences somehow impact the probability that a given molecular formula may be selected in a given DOM assemblage (e.g., a site with high NPOC might have more kinds of DOM). As TN incorporates both organic and inorganic nitrogen, its key role in ecological processes may be hypothesized as **1)** representing a potential nutrient source, **2)** varying along the river continuum as a function land use and hydrologic exchange, and **3)** being connected to organic matter through organic nitrogen degradation and production. TSS could represent a mode of transportation for DOM, i.e., organic matter may bind to solids as they travel through water ways with some subsets of DOM having enhanced adsorption in comparison to others thereby acting as a deterministic filter. Both SO4 and NO3 play roles in microbial metabolisms and may point to differential degradation (e.g., Boye et al., (*72*)).

Finally, the fourth group is composed of only DIC which as a much stronger correlation (Mantel ρ = 0.396, p-value = 0.001) within stream orders than across stream orders (Mantel ρ = 0.273, p-value = 0.001). DIC concentrations are impacted by either non-biological processes (e.g., weathering, anthropogenic) or biological processes (e.g., degradation and CO2 generation, or CO2 fixation). Within our watershed, DIC is significantly related to geospatial PC1 (and associated variables) while being positively related to shrub coverage indicating that DIC is generally higher in forested, higher precipitation regions (**Supplemental Figure 1**). Given this association, we argue that these DIC correlations potentially identify terrestrial-river interactions. This, in turn, is related to DOM diversification because it acts as a proxy for transport of terrestrial DOM. This could further impact DOM diversification by providing an alternative mechanism for carbon utilization (e.g., the creation of new secondary metabolites).

Combined, these results point to a complex interplay of environmental conditions that explain the selection processes occurring at a pairwise scale, whether controlling for within stream order comparisons or looking across the whole dataset. While we are unable to completely reconcile these patterns into a mechanistic model, they provide important context for the conceptual model outlined in the main text. Particularly, these results highlight the specific environmental conditions that might be coordinating to drive increased diversification at mid-order streams and may provide insight in the future with increased data resolution.