**Effect of Grain Size on Bioregionalisation of Orchids**

This report presents a series of analyses designed to understand why an increase in spatial grain size leads to a greater number of biogeographical regions in the clustering of global orchid phylogenetic beta diversity (mean βsim). Contrary to the prevailing assumption that coarser grains yield fewer, more homogeneous regions, our results reveal the opposite trend.

## Distribution of mean βsim across spatial scales

To interpret why broader grain sizes yielded more biogeographical regions, we first examined whether spatial aggregation increased or decreased floristic dissimilarity between grid cells. We calculated the distribution of mean βsim across the four spatial resolutions (100, 200, 400, and 800 km), using the same distance matrices employed for clustering. This analysis was designed to test whether spatial aggregation reduces compositional differences, which would support fewer, broader regions. As expected, the average “mean βsim” values slightly decreased with grain size (from 0.62 at 100 km to 0.60 at 800 km; Table 1), indicating some spatial homogenisation. However, the overall variance in dissimilarity increased, with larger standard deviations and broader ranges at coarser grains (Table 1). These results suggest that while many grid cells became more similar as grain size increased, others remained highly differentiated, contributing to greater heterogeneity.

Table 1.- Summary statistics of mean βsim values by grain size.

| Scale | No. pairs | Mean | Median | SD | Max | Min |
| --- | --- | --- | --- | --- | --- | --- |
| 100 km | 25,025,006 | 0.6226606 | 0.6644966 | 0.1769543 | 0.9633141 | 0 |
| 200 km | 4,837,800 | 0.6141359 | 0.6617864 | 0.1859851 | 0.9581058 | 0 |
| 400 km | 721,650 | 0.6044446 | 0.6577295 | 0.1964286 | 0.9451540 | 0 |
| 800 km | 99,540 | 0.6011444 | 0.6596660 | 0.2081267 | 0.9468477 | 0 |

## Between-cluster phylogenetic dissimilarity across spatial scales

Based on our hypothesis, we expected that coarser grains would yield more internally homogeneous clusters reflected in higher silhouette values. However, the number of clusters increased rather than decreased with scale, raising questions about their coherence. To investigate this, we calculated silhouette scores both overall and by cluster, at each grain size. The results indicate slightly higher between-cluster dissimilarities at 200 and 800 km, though this pattern is not linear (Table 2). Higher variance at coarser grains suggests more fragmented regionalisation at broad scales (Table 2).

Table 2.- Average mean βsim between clusters per grain size.

| Scale | No. pairs | Mean | Median | SD | Max | Min |
| --- | --- | --- | --- | --- | --- | --- |
| 100 km | 10 | 0.7057584 | 0.7073587 | 0.03638896 | 0.7517470 | 0.6444256 |
| 200 km | 15 | 0.7099921 | 0.7168417 | 0.03893582 | 0.7587710 | 0.6368862 |
| 400 km | 190 | 0.6726459 | 0.6892307 | 0.09709464 | 0.8422587 | 0.4444322 |
| 800 km | 120 | 0.6889031 | 0.6991459 | 0.10476850 | 0.8769634 | 0.4313877 |

## Clustering quality based on silhouette widths

To assess the coherence of cluster assignments, silhouette widths were calculated for each grain size. The average silhouette peaked at 200 km (0.340), followed by 100 km (0.318), 800 km (0.317), with lower values at 400 km (0.202). Cluster-level silhouettes at 800 km showed a wide range, from 0.03 to 0.98 (Table 3), with several large clusters (e.g. clusters 1, 5, and 9) having values > 0.5—indicating good internal consistency (Table 3). These results suggest that many of the additional clusters at 800 km are internally coherent, rather than artefacts. The 200 km grain appears optimal in balancing within-cluster homogeneity and regional distinctiveness, but 800 km still retains biologically meaningful structure.

Table 3.- Silhouette width per cluster per scale.

| Cluster | Mean silhouette | No. cells | Scale |
| --- | --- | --- | --- |
| 1 | 0.47788968 | 305 | 100 km |
| 2 | 0.25993145 | 1,199 | 100 km |
| 3 | 0.26118420 | 631 | 100 km |
| 4 | 0.18676454 | 664 | 100 km |
| 5 | 0.38328002 | 2,204 | 100 km |
| 1 | 0.40938753 | 955 | 200 km |
| 2 | 0.19438429 | 335 | 200 km |
| 3 | 0.52511184 | 109 | 200 km |
| 4 | 0.90136960 | 10 | 200 km |
| 5 | 0.27780851 | 473 | 200 km |
| 6 | 0.29505833 | 318 | 200 km |
| 1 | 0.22410013 | 102 | 400 km |
| 2 | 0.47498364 | 13 | 400 km |
| 3 | 0.43618485 | 47 | 400 km |
| 4 | 0.47610936 | 15 | 400 km |
| 5 | -0.08233796 | 68 | 400 km |
| 6 | 0.04652772 | 35 | 400 km |
| 7 | 0.03535123 | 145 | 400 km |
| 8 | 0.00000000 | 1 | 400 km |
| 9 | 0.06815943 | 4 | 400 km |
| 10 | 0.00000000 | 1 | 400 km |
| 11 | 0.00000000 | 1 | 400 km |
| 12 | 0.54257122 | 14 | 400 km |
| 13 | 0.00000000 | 1 | 400 km |
| 14 | 0.33846003 | 23 | 400 km |
| 15 | 0.19436452 | 309 | 400 km |
| 16 | 0.00000000 | 1 | 400 km |
| 17 | 0.00000000 | 1 | 400 km |
| 18 | 0.48732189 | 57 | 400 km |
| 19 | 0.94464777 | 7 | 400 km |
| 20 | 0.80326387 | 5 | 400 km |
| 1 | 0.81632419 | 5 | 800 km |
| 2 | 0.28760516 | 16 | 800 km |
| 3 | 0.16152327 | 38 | 800 km |
| 4 | 0.97948507 | 3 | 800 km |
| 5 | 0.81816683 | 5 | 800 km |
| 6 | 0.21719988 | 37 | 800 km |
| 7 | 0.56526846 | 6 | 800 km |
| 8 | 0.14452354 | 21 | 800 km |
| 9 | 0.69284178 | 6 | 800 km |
| 10 | 0.18649962 | 12 | 800 km |
| 11 | 0.23168702 | 11 | 800 km |
| 12 | 0.19385301 | 14 | 800 km |
| 13 | 0.51824109 | 5 | 800 km |
| 14 | 0.00000000 | 1 | 800 km |
| 15 | 0.34209717 | 120 | 800 km |
| 16 | 0.52784675 | 16 | 800 km |

## Hierarchical depth of clusters (dendrogram analysis)

To determine whether the additional regions at coarser scales arose from low clustering thresholds (i.e., shallow dendrogram cuts), we evaluated the height values at which clusters formed in UPGMA dendrograms across grain sizes. At a given number of clusters (e.g. k = 6 or k = 10), the cut heights were consistently lower at coarser grains: for instance, the height at k = 10 was 0.596 for 100 km but only 0.493 for 800 km (Table 4). This indicates that regions at broader grains were defined at shallower levels of the dendrogram, supporting the idea that finer-scale structure becomes subdivided or re-partitioned differently with spatial aggregation.

Table 4.- Dendrogram cut heights by cluster number (k) and grain size.

| k | 100 km | 200 km | 400 km | 800 km |
| --- | --- | --- | --- | --- |
| 2 | 0.7426225 | 0.7531611 | 0.7551834 | 0.7593216 |
| 3 | 0.6990274 | 0.6977364 | 0.6906664 | 0.6859332 |
| 4 | 0.6841377 | 0.6958563 | 0.6824983 | 0.6809667 |
| 5 | 0.6444256 | 0.6721241 | 0.6505469 | 0.6320099 |
| 6 | 0.6353237 | 0.6368862 | 0.6399523 | 0.6144414 |
| 7 | 0.6239641 | 0.5908666 | 0.6029722 | 0.5406557 |
| 8 | 0.6114461 | 0.5851656 | 0.5401137 | 0.5262909 |
| 9 | 0.6029831 | 0.5738017 | 0.5365228 | 0.5160759 |
| 10 | 0.5962214 | 0.5724405 | 0.5098907 | 0.4927011 |
| 11 | 0.5898633 | 0.5505813 | 0.5073786 | 0.4688776 |
| 12 | 0.5883935 | 0.5501284 | 0.4941845 | 0.4667727 |
| 13 | 0.5868238 | 0.5435192 | 0.4913742 | 0.4620846 |
| 14 | 0.5852509 | 0.5326988 | 0.4797130 | 0.4619324 |
| 15 | 0.5776132 | 0.5242182 | 0.4566411 | 0.4346734 |
| 16 | 0.5775845 | 0.5219593 | 0.4556515 | 0.4313877 |
| 17 | 0.5769221 | 0.5166514 | 0.4556095 | 0.4121717 |
| 18 | 0.5753434 | 0.5133109 | 0.4521599 | 0.4094120 |
| 19 | 0.5747858 | 0.5122057 | 0.4468969 | 0.4066251 |
| 20 | 0.5707070 | 0.5101258 | 0.4444322 | 0.4032259 |

## Final remarks

Across all analyses, we found that larger grain sizes did not reduce the number of regions or increase floristic homogeneity, as initially hypothesised. Instead, broader grains revealed additional, internally coherent regions that often corresponded to subdivisions of finer-grain clusters. These results suggest that spatial aggregation does not dilute biogeographical signal but may instead highlight hierarchical structuring of orchid diversity. Consequently, coarse grains offer complementary perspectives on floristic organisation rather than obscured or artefactual views.