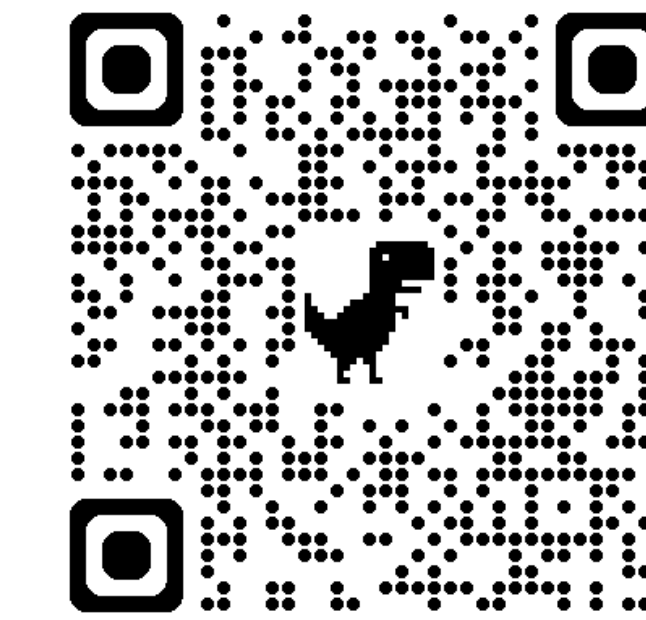
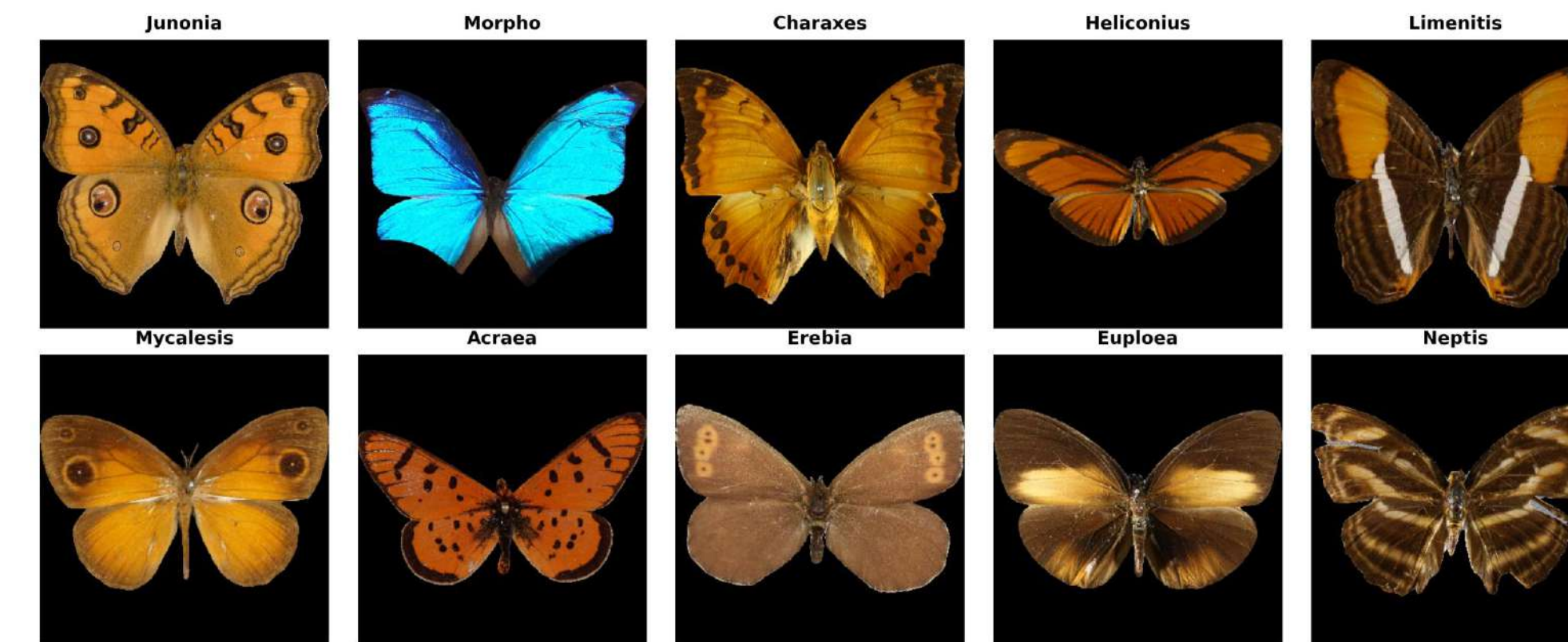




## 1. Motivation : Collection-Scale Bottleneck?



**Digitization Acceleration** : Biological image archives are expanding far faster than manual expert curation can scale.



**Repository Noise** : Historical collections contain unresolved boundaries, labeling errors, and subtle population structures.



**Supervised Trap**: Supervised classifiers force images into predefined species labels, inheriting training label noise and failing on novel or variable morphs.

*Proposed Approach: Visual clustering offers a practical triage tool for large digitized natural history collections*

## 2. Dataset: Why Butterflies? A Rich Vision Benchmark

### Biological significance

- Wing patterns encode mimicry, polymorphism, geography, seasonality, and divergence.
- Species labels and visual phenotypes are related but non-equivalent.
- Mismatches can guide expert review and collection triage.

### Computer vision significance

- Fine-grained differences and within-species variation make recognition hard.
- Open-set structure exposes limits of fixed-label classification.
- Long-tailed sampling reflects real biodiversity archives.
- Supports clustering, retrieval, anomaly detection, and



*Butterflies provide a rare testbed where visual similarity, taxonomy, and biological process interact—but do not always agree.*

## 3. Benchmark: Species-Capped 1-50 Sampling

230,000+

SOURCE IMAGES

189,541

DORSAL ARCHIVE

11,821

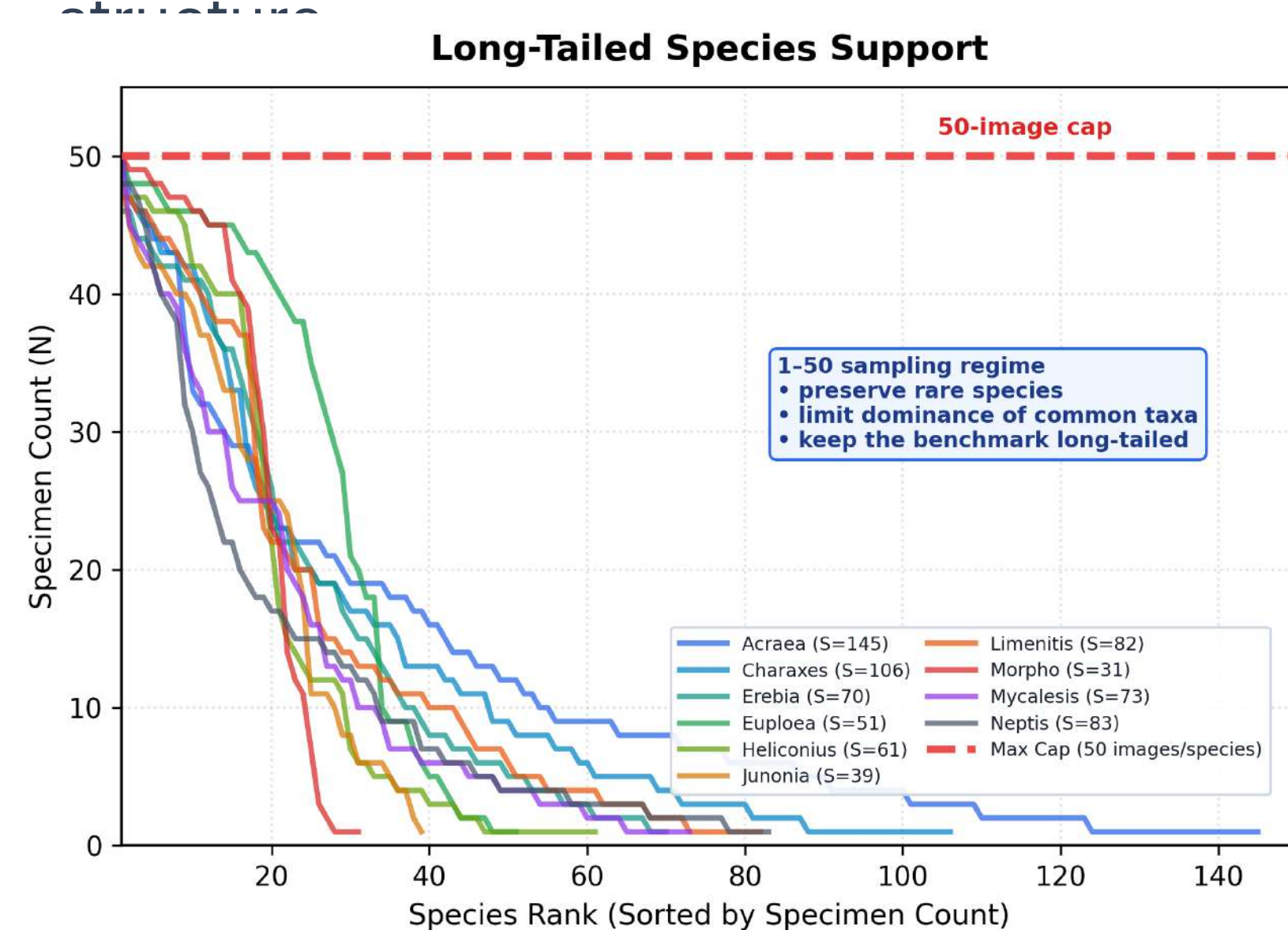
BENCHMARK IMAGES

741

RECOGNIZED SPECIES

### Key Design : 1-50 images/species

- Retain rare taxa • Cap dominant species • Preserve long-tailed



| Genus      | S   | N     | N/S   |
|------------|-----|-------|-------|
| Junonia    | 39  | 919   | 23.56 |
| Morpho     | 31  | 943   | 30.42 |
| Charaxes   | 106 | 1,403 | 13.24 |
| Heliconius | 61  | 1,020 | 16.72 |
| Limenitis  | 82  | 1,246 | 15.20 |
| Mycalesis  | 73  | 1,030 | 14.11 |
| Acraea     | 145 | 1,675 | 11.55 |
| Erebia     | 70  | 1,206 | 17.23 |
| Euploea    | 51  | 1,378 | 27.02 |
| Neptis     | 83  | 1,001 | 12.06 |

*Species-capped sampling retains rare species in the benchmark while preventing abundant taxa from dominating evaluation.*

## 7. Conclusion: What this work enables

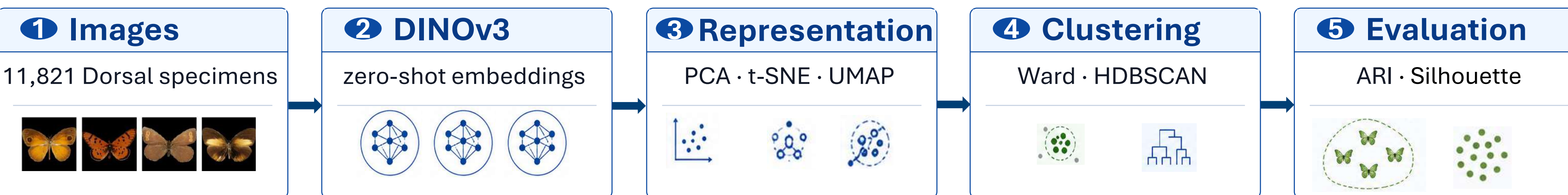
**Taxonomy-free organization** of large specimen archives by visual phenotype.

**Automated triage** of fragmented taxa, mixed clusters, atypical specimens, and label inconsistencies.

**A reproducible benchmark** for evaluating foundation-model representations in biological collections.

*Self-supervised vision can recover meaningful phenotype structure—but phenotype and taxonomy are not equivalent.*

## 4. Methodology: Taxonomy-Free Benchmarking Pipeline



### Zero Shot Visual Structuring

No taxonomic labels are used during feature extraction, representation reduction, or clustering.

### Post-hoc taxonomic audit

Species names are introduced only afterward to measure whether visual structure aligns with taxonomy.

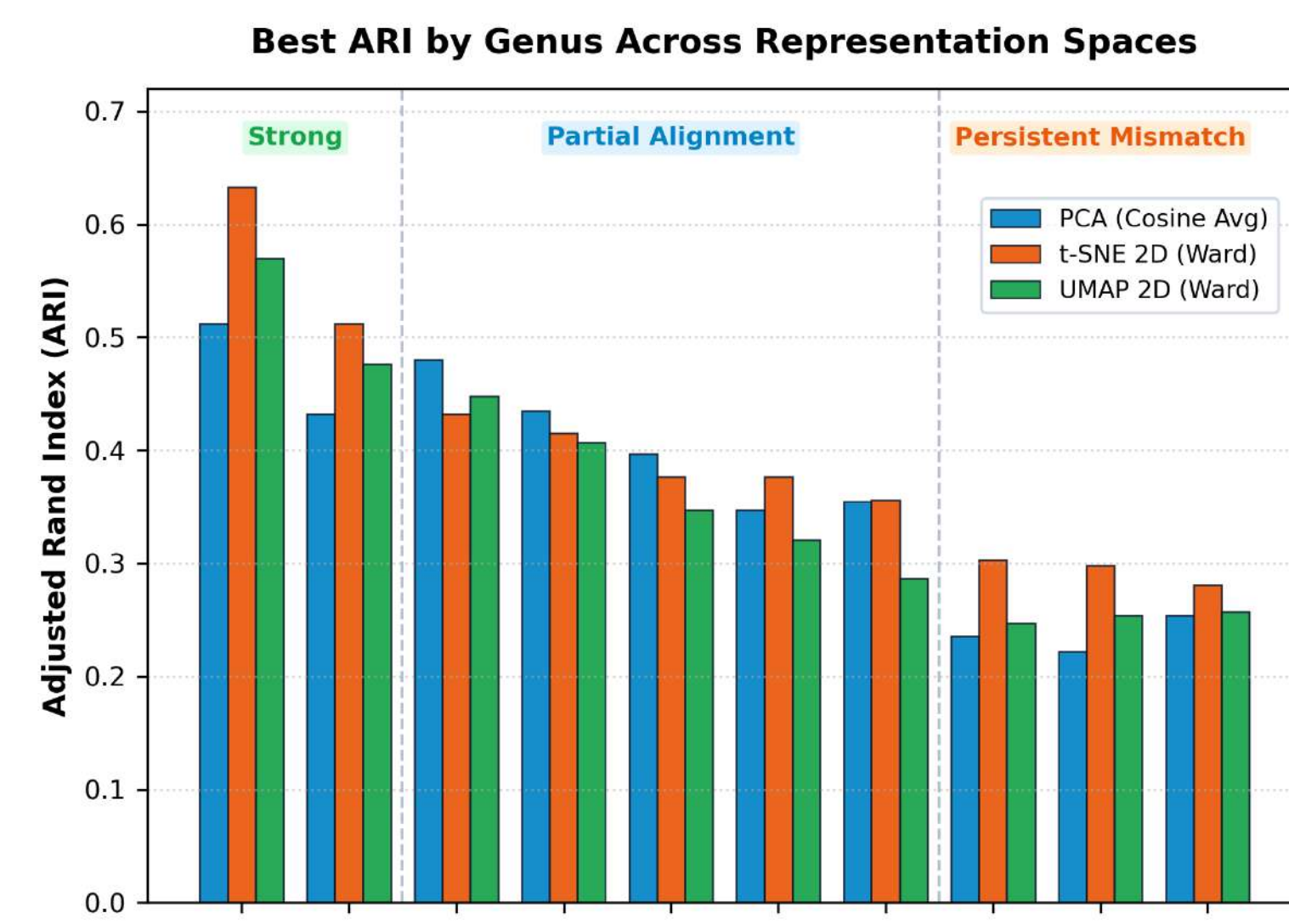
### Extrinsic vs. intrinsic metrics

ARI quantify taxonomic agreement; Silhouette measures visual coherence.

*The model is not trained to predict taxonomy; taxonomy is used only afterward as an audit signal.*

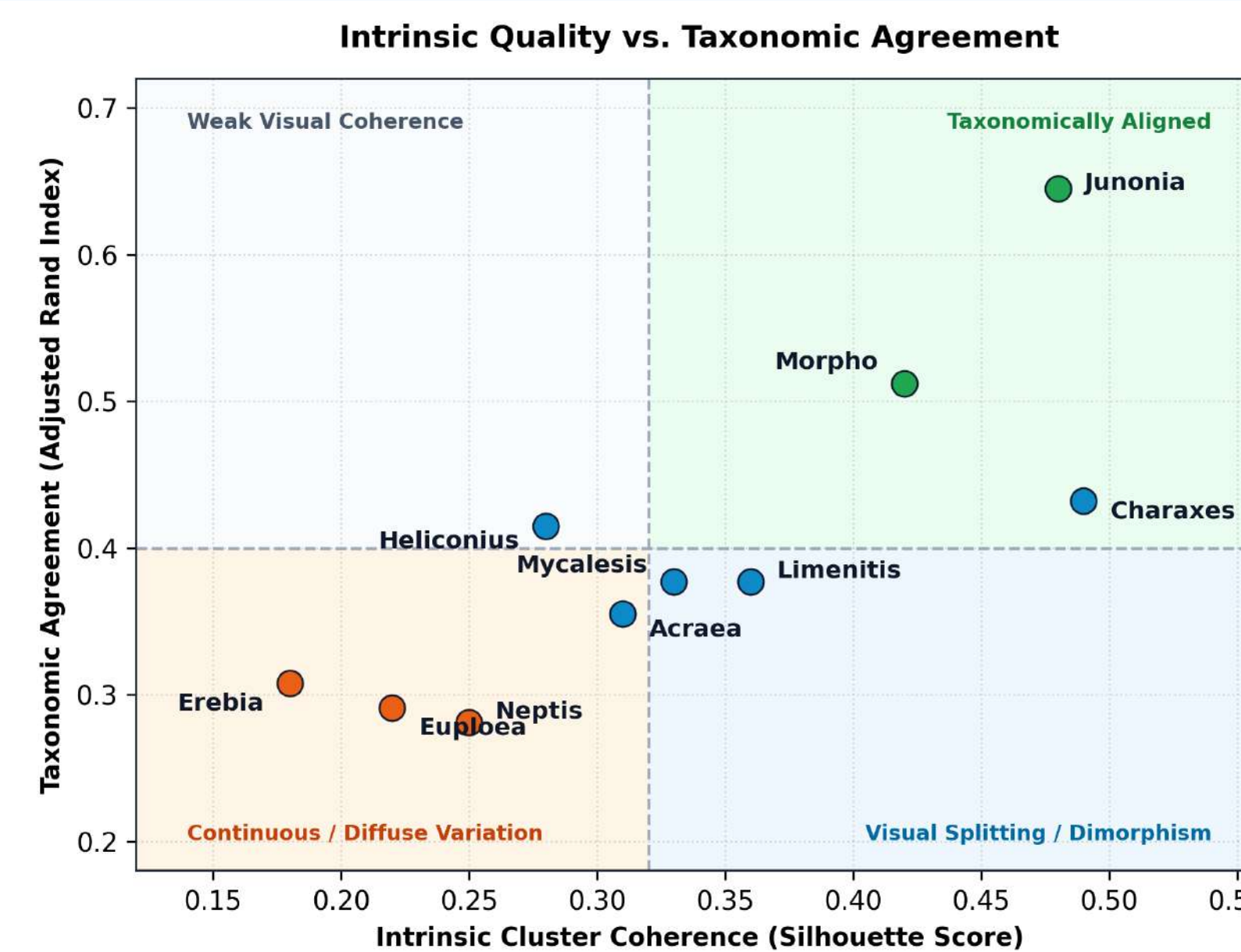
## 5. Quantitative Results

### 5.1 Phenotype–Taxonomy Alignment



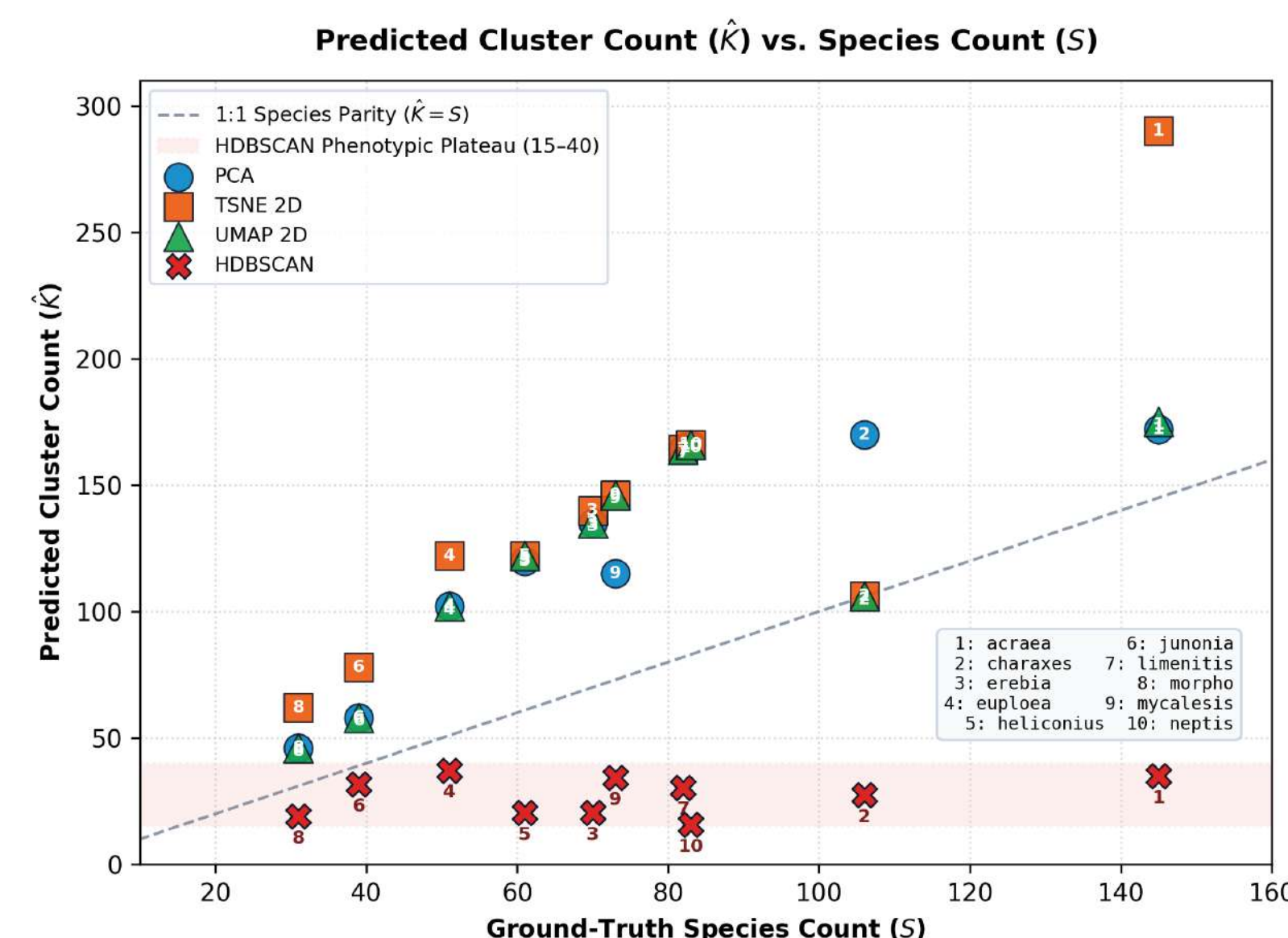
*Phenotype–taxonomy alignment is driven by genus-specific biology.*

### 5.2 Visual Coherence Does Not Equal Taxonomic Agreement



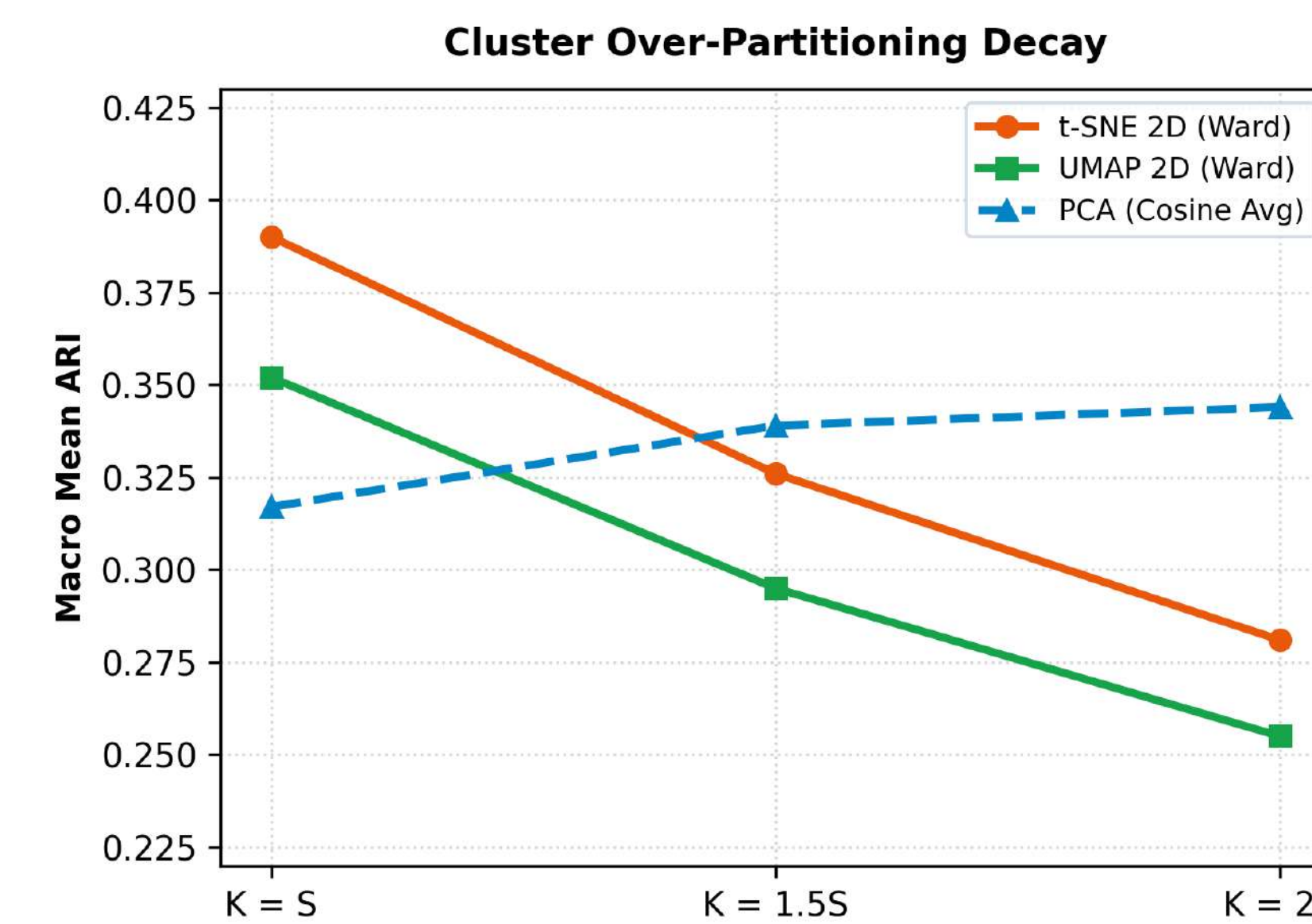
*Phenotypic organization and taxonomy are related—but non-equivalent.*

### 5.3 Dense Phenotypic Modes Are Coarser Than Species Richness



*Dominant phenotypic structure is often coarser than species-level taxonomy.*

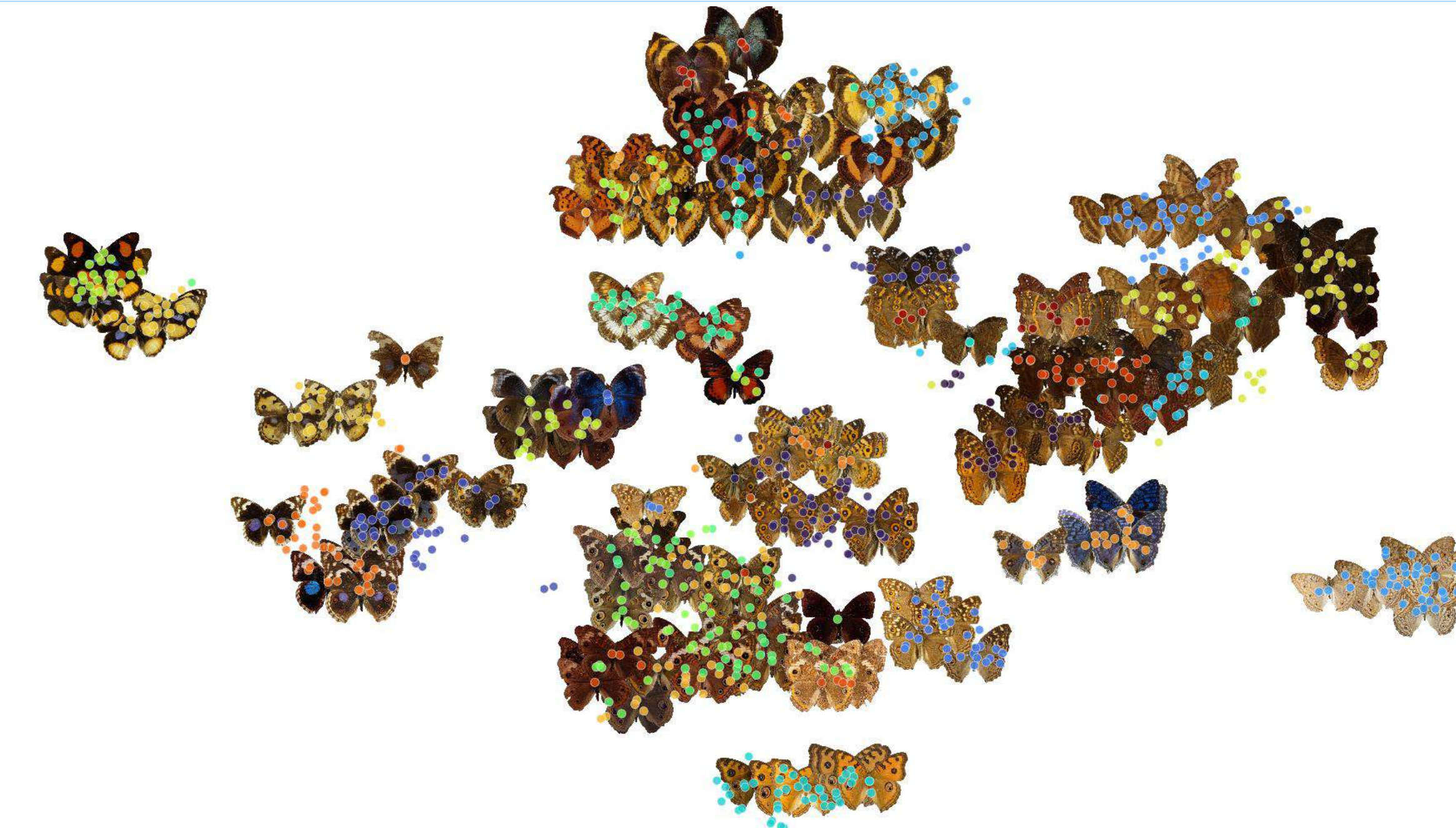
### 5.4 Nonlinear Manifolds Reveal Fine Scale Phenotypic Structure



*Finer partitions expose within-species phenotypic substructure.*

## 6. Qualitative Results: Alignment Regimes

### 6.1 Strong Alignment : Junonia



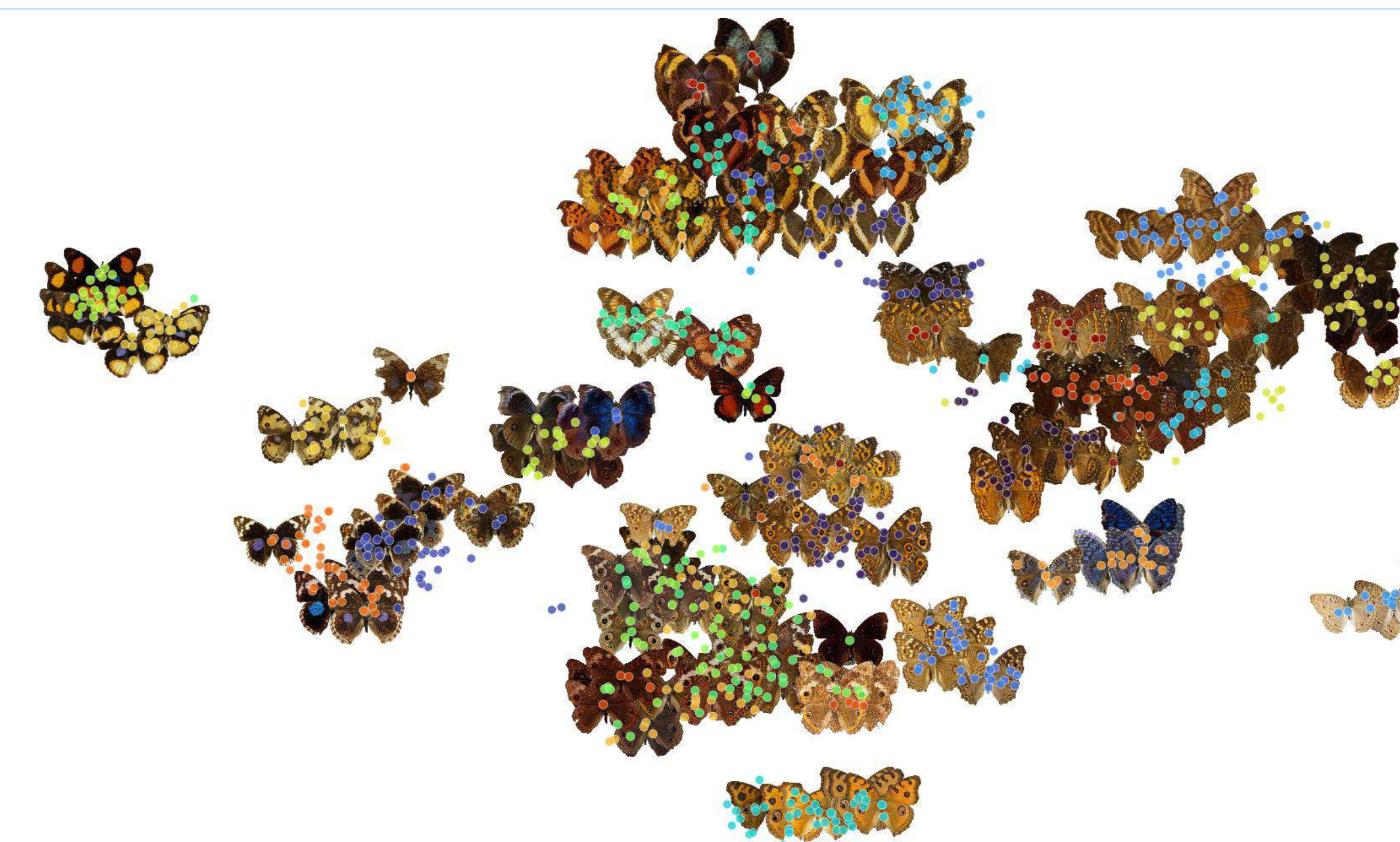
*Junonia: Discrete phenotypic neighborhoods broadly track species.*

### 6.2 Partial Alignment : Charaxes



*Charaxes: Coherent phenotypic modes fragment nominal species.*

### 6.3 Persistent Mismatch : Erebia



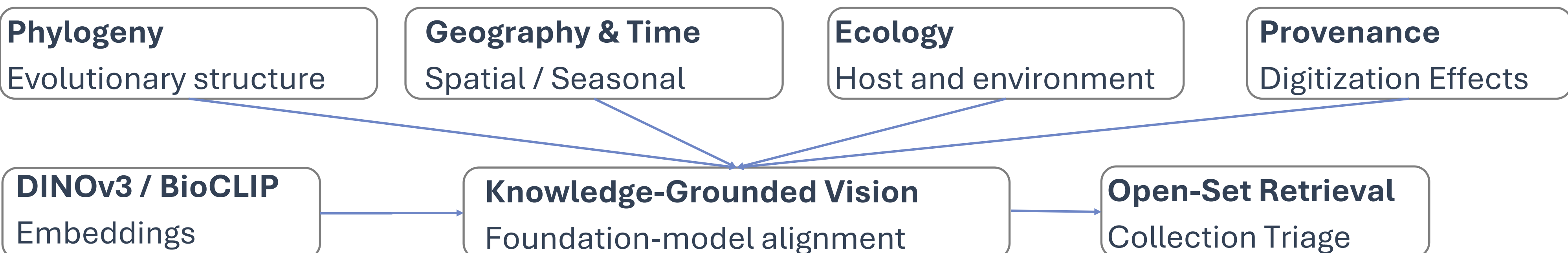
*Erebia: Continuous phenotypic variation weakens species separation.*

## 9. Collaboration Opportunities

- Geometric ML & graph representation learning
- Open-set retrieval & anomaly detection
- Vision foundation models for biodiversity
- Cross-collection / multimodal biological AI

*Seeking collaborators across computer vision, geometric ML, biodiversity informatics, and natural-history collections.*

## 8. Future Direction: From Visual Phenotype to Knowledge-Grounded Vision



*Align visual embeddings with relational biological structure for Open-Set Retrieval and Collection Triage.*