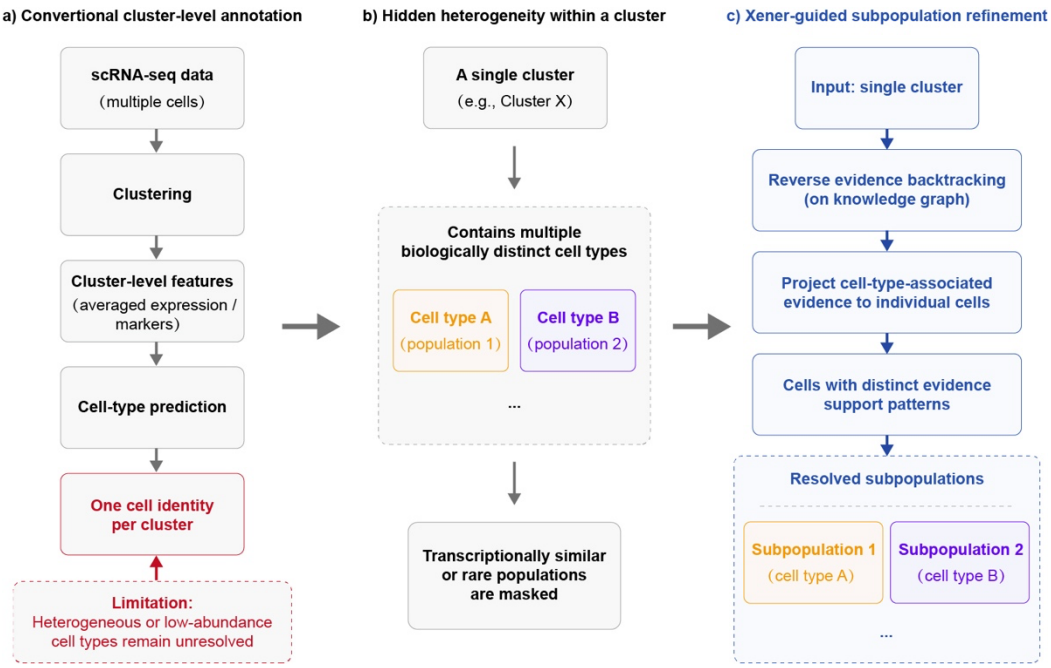
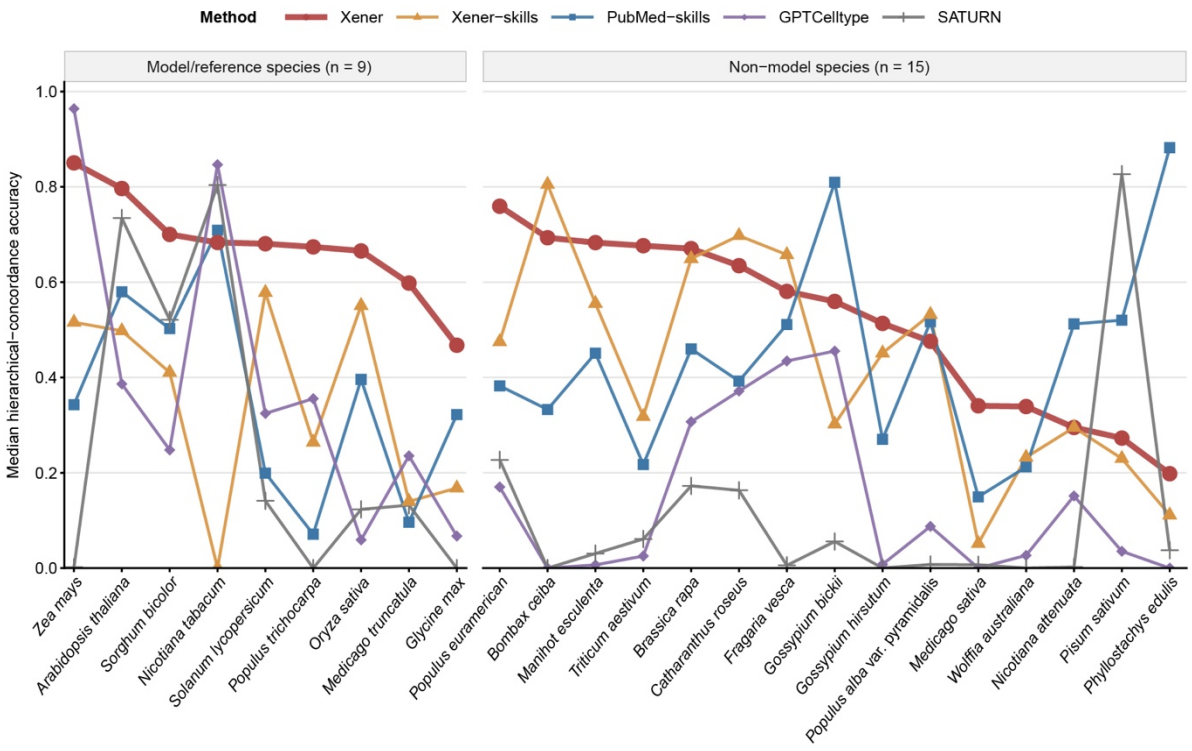
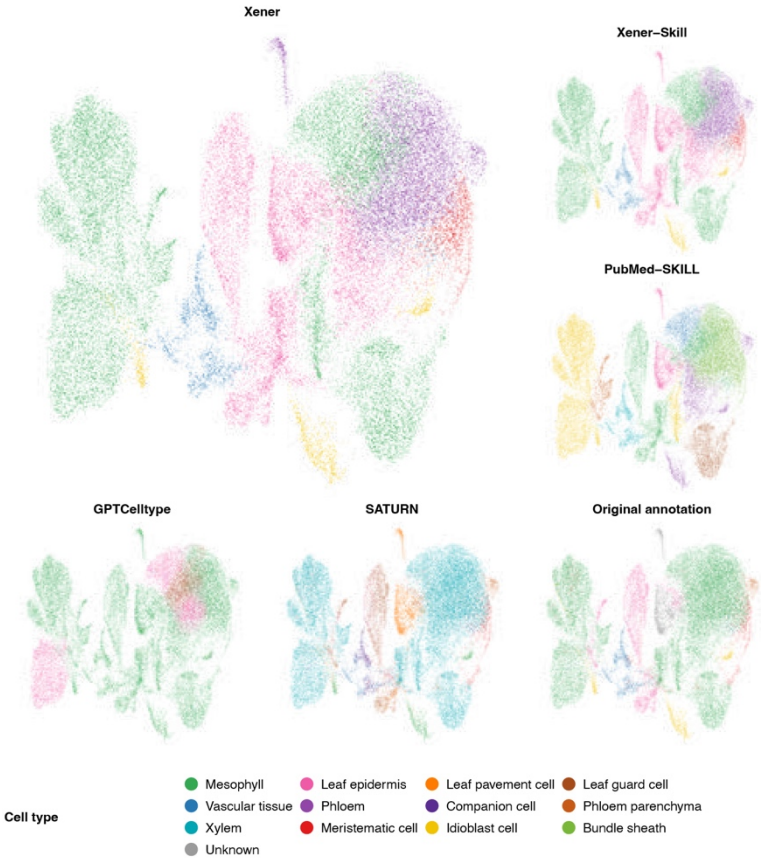


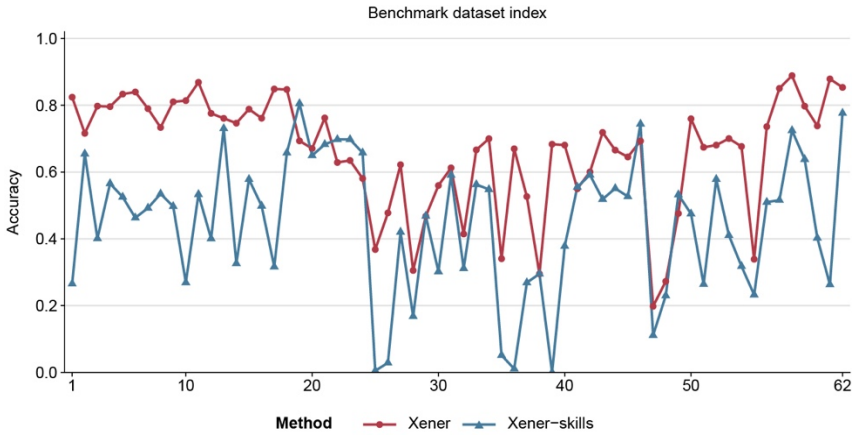
Supplementary Figure S1. Conceptual workflow showing how Xener resolves hidden intracluster heterogeneity. (a) Conventional cluster-level annotation aggregates expression and marker information across multiple cells, assigns one cell identity to each cluster, and can therefore leave heterogeneous or low-abundance cell types unresolved. (b) A single cluster may contain multiple biologically distinct cell types; aggregation can mask transcriptionally similar or rare populations. (c) Xener-Refine starts from an annotated cluster, backtracks candidate cell-type evidence through the knowledge graph, projects cell-type-associated evidence onto individual cells, and identifies cell-specific support patterns to resolve evidence-supported subpopulations. Refinement is restricted to candidate identities supported by the cluster-level Xener-Anno result, thereby avoiding unrelated labels and eliminating the need for reclustering. scRNA-seq, single-cell RNA sequencing.



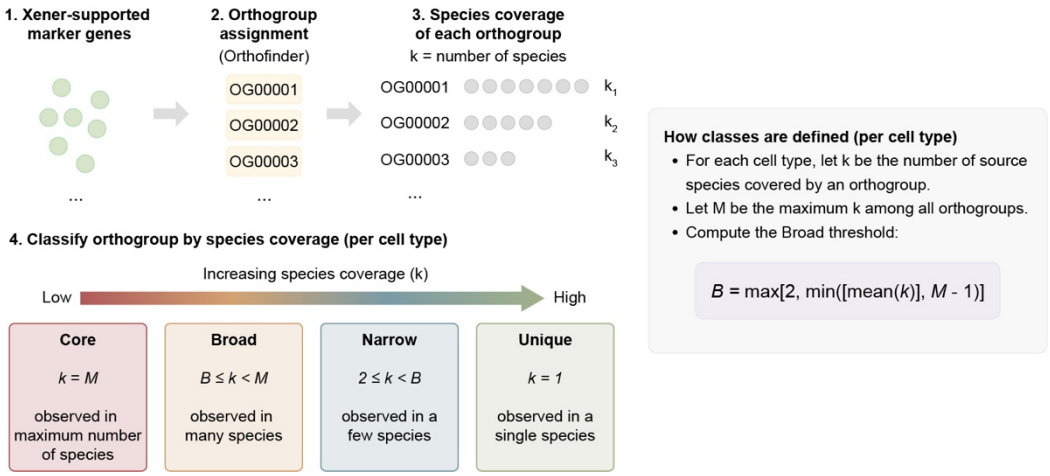
Supplementary Figure S2. Species-level median hierarchical concordance across model/reference and non-model plant species. Median Plant Ontology-aware hierarchical-concordance accuracy is shown for Xener, Xener-Skill, PubMed-SKILL, GPTCelltype, and SATURN in nine model/reference species (left) and 15 non-model species (right). For species represented by multiple benchmark accessions, each symbol denotes the median dataset-level accuracy for that species; species represented by one accession contribute the corresponding single value. Xener, Xener-Skill, PubMed-SKILL, and GPTCelltype were evaluated as reference-independent workflows, whereas SATURN used reference-dependent cross-species label transfer. Concordance was scored against the published labels after standardization to Plant Ontology terms and included exact matches, parent-child relationships, or labels sharing the same immediate ontology parent. Lines connect species-level values within each method only to guide the eye and do not represent a continuous or phylogenetic trajectory.



Supplementary Figure S3. Comparison of cell-type annotations for the *Catharanthus roseus* leaf dataset SRP335448. Uniform manifold approximation and projection (UMAP) embeddings display the same cells and coordinates colored by published annotation, Xener annotation, Xener-Skill annotation, PubMed-SKILL annotation, GPTCelltype annotation, and SATURN annotation. Each point represents one cell; colors indicate the cell-type label assigned by the corresponding annotation source. Published labels were withheld during prediction and used only for benchmark evaluation. Differences in the number and granularity of displayed labels reflect method-specific assignments after harmonization to Plant Ontology terminology.

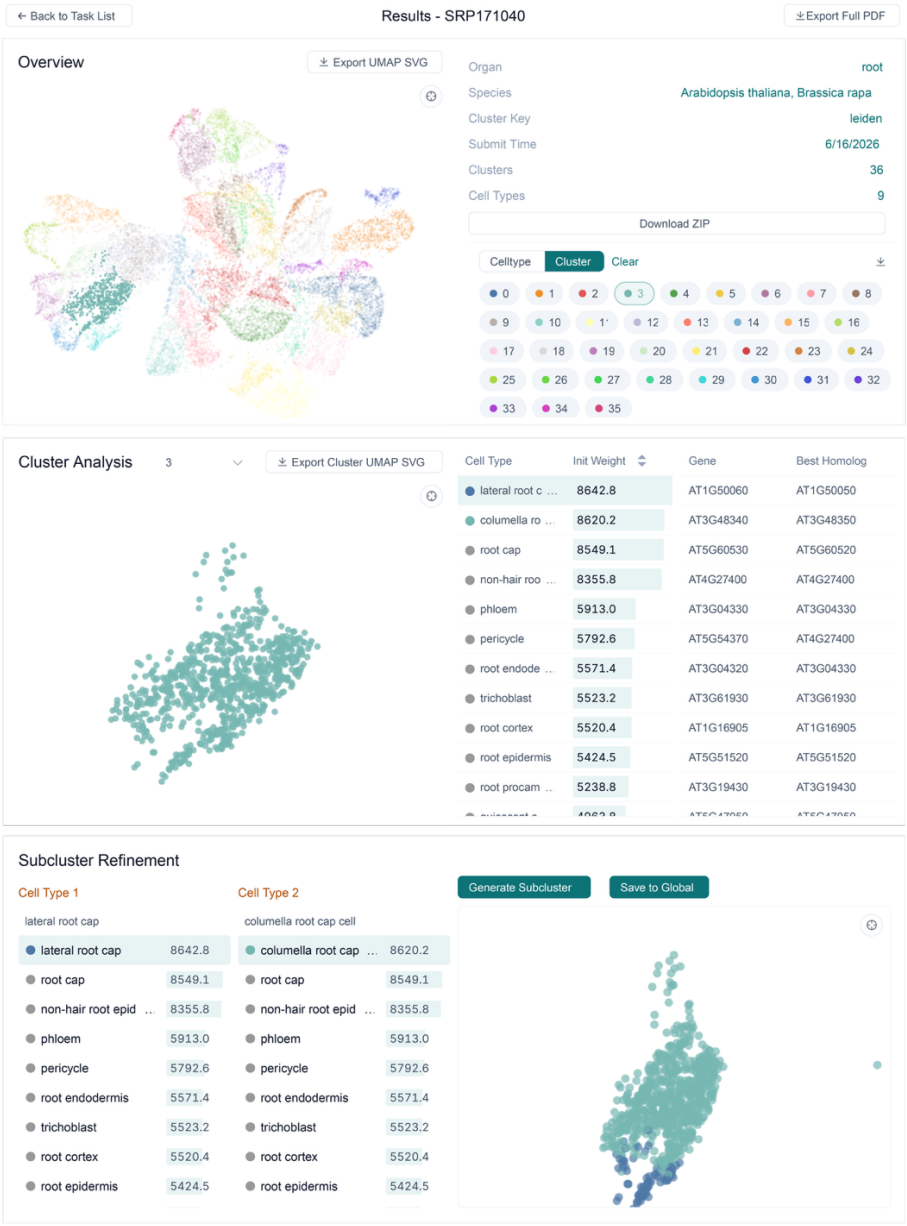


Supplementary Figure S4. Dataset-level hierarchical-concordance accuracy of Xener and Xener-Skill across 62 benchmark datasets. Each point represents one accession-level single-cell dataset evaluated against its published cell-type labels using the same Plant Ontology-aware hierarchical-concordance criterion. Datasets are displayed by benchmark index from 1 to 62. Xener results are shown as red circles and Xener-Skill results as blue triangles; connecting lines are included only to guide the eye. Across the benchmark cohort, the median hierarchical concordance was 0.693 for Xener and 0.498 for the agent-based Xener-Skill interface.



Supplementary Figure S5. Orthogroup-based classification of cross-species marker conservation. Xener-supported marker genes were assigned to orthogroups with OrthoFinder, and conservation was evaluated separately for each standardized cell type. For an orthogroup, species coverage k was defined as the number of benchmark

species containing at least one orthogroup member associated with the focal cell type; multiple members from the same species were counted once. For each cell type, M denotes the maximum k among its orthogroups, and the lower threshold for the Broad category was calculated as B = max[2, min(mean(k), M - 1)]; when M was less than or equal to 2, B was set to M. Orthogroups were classified as Unique when k = 1, Narrow when 2 <= k < B, Broad when B <= k < M, and Core when k = M and M > 1. Marker records inherited the conservation category of their corresponding orthogroup. The horizontal arrow indicates increasing species coverage from low to high.



Supplementary Figure S6. Interactive visualization and subcluster refinement in the Xener web interface. The overview panel summarizes the SRP171040 root dataset and displays the global UMAP colored by 36 Leiden clusters, together with the task metadata and nine assigned cell types. Leiden cluster 3 is selected for detailed analysis, showing the cluster-specific UMAP and candidate cell types ranked by their initial annotation weights, supporting genes, and best homologs. Lateral root cap and columella root cap cell are selected as candidate identities for subcluster refinement. The resulting UMAP preview displays their cell-level separation before the refined assignments are saved to the global annotation.