北京林业大学学报
JOURNAL OF BEIJING FORESTRY UNIVERSITY

毛竹与拟南芥顶端分生组织的单细胞转录组比较分析

孙镭 赵韩生

Comparative analysis of single-cell transcriptomes in *Phyllostachys edulis* and *Arabidopsis thaliana* shoot apical meristem

Sun Lei, Zhao Hansheng

引用本文:
孙镭, 赵韩生. 毛竹与拟南芥顶端分生组织的单细胞转录组比较分析[J]. 北京林业大学学报, 2025, 47(7):1–10. doi: 10.12171/j.1000–1522.20250038
Sun Lei, Zhao Hansheng. Comparative analysis of single–cell transcriptomes in *Phyllostachys edulis* and *Arabidopsis thaliana* shoot apical meristem[J]. Journal of Beijing Forestry University, 2025, 47(7):1–10. doi: 10.12171/j.1000–1522.20250038

在线阅读 View online: https://doi.org/10.12171/j.1000–1522.20250038

您可能感兴趣的其他文章
Articles you may be interested in

基于全长转录组序列、核基因与叶绿体基因分析琼岛杨在杨属的亲缘关系
Genetic relationship of *Populus qiongdaoensis* in *Populus* based on full–length transcriptome sequences, nuclear genes and chloroplast genes
北京林业大学学报, 2021, 43(10): 28–37. https://doi.org/10.12171/j.1000–1522.20200286

毛竹不同构造中细胞化学成分及热稳定性研究
Chemical composition and thermal stability of cells in different structures of *Phyllostachys edulis*
北京林业大学学报, 2023, 45(8): 156–162. https://doi.org/10.12171/j.1000–1522.20230104

杨树*PdKNAT7*基因调控拟南芥次生细胞壁厚度的功能研究
Function of *PdKNAT7* gene in poplar regulating the thickness of secondary cell wall in *Arabidopsis thaliana*
北京林业大学学报, 2022, 44(11): 1–9. https://doi.org/10.12171/j.1000–1522.20210083

拟南芥株高性状全基因组上位互作网络构建
Network construction of genome–wide epistatic interaction for plant height traits in *Arabidopsis thaliana*
北京林业大学学报, 2023, 45(9): 21–32. https://doi.org/10.12171/j.1000–1522.20220146

白桦木质部原生质体初生细胞壁再生过程转录组分析
Transcriptome analysis of regeneration process of primary cell wall in xylem protoplast of *Betula platyphylla*
北京林业大学学报, 2022, 44(8): 12–22. https://doi.org/10.12171/j.1000–1522.20200376

胡杨PeMAX2调控拟南芥耐旱性
Populus euphratica PeMAX2 regulating drought tolerance in *Arabidopsis thaliana*
北京林业大学学报, 2024, 46(6): 106–117. https://doi.org/10.12171/j.1000–1522.20220494

DOI:10.12171/j.1000−1522.20250038

毛竹与拟南芥顶端分生组织的单细胞转录组比较分析

孙　镭　　赵韩生
（国际竹藤中心竹藤资源基因科学与基因产业化研究所，竹藤科学与技术国家林业和草原局重点实验室，北京 100102）

摘要:【目的】探索跨物种单细胞整合方法在植物中的应用，揭示毛竹与拟南芥顶端分生组织（SAM）细胞类型在单细胞转录组水平上的保守性与差异性。**【方法】**基于鉴定出的直系同源基因对，对毛竹和拟南芥的 SAM 单细胞转录组数据进行整合，并利用典型相关分析（CCA）进行批次矫正，以消除物种间的技术偏差。随后，使用一致流形逼近与投影（UMAP）方法对跨物种整合的细胞分布模式进行可视化分析，并通过斯皮尔曼相关性分析和跨物种保守标志基因鉴定，探究不同细胞类型的转录组相似性及其功能特征。**【结果】**（1）共鉴定出 23 个细胞类群，在 UMAP 可视化空间中，大部分细胞按物种来源聚集在不同区域，但部分细胞类群在两个物种间出现明显重叠，表明这些细胞在转录组水平上具有较高相似性。（2）进一步分析发现，重叠区域主要由表皮细胞和维管组织细胞组成，表明这类组织在植物演化过程中较为保守；相关性分析结果进一步证实了这些保守细胞类型之间基因表达的相似性。而其他细胞类群则表现出较大的物种特异性，且为多种细胞类型的混合。（3）共鉴定出 183 个跨物种保守的标志基因，这些基因在两个物种中的特定细胞类型中均特异表达，功能分析揭示了其在植物发育和进化中的关键作用。**【结论】**本研究揭示了毛竹与拟南芥保守细胞类型的核心功能特征，为植物进化机制提供了新的视角，并为非模式植物的功能基因发掘和研究奠定了数据基础。

关键词：毛竹；拟南芥；直系同源基因；跨物种整合；单细胞转录组；保守细胞类型；标志基因
中图分类号：Q811.4　　**文献标志码：**A　　**文章编号：**1000−1522(2025)07−0001−10
引文格式：孙镭，赵韩生. 毛竹与拟南芥顶端分生组织的单细胞转录组比较分析［J］. 北京林业大学学报，2025，47(7)：1−10. Sun Lei, Zhao Hansheng. Comparative analysis of single-cell transcriptomes in *Phyllostachys edulis* and *Arabidopsis thaliana* shoot apical meristem［J］. Journal of Beijing Forestry University, 2025, 47(7): 1−10.

Comparative analysis of single-cell transcriptomes in *Phyllostachys edulis* and *Arabidopsis thaliana* shoot apical meristem

Sun Lei　Zhao Hansheng
(Key Laboratory of National Forestry and Grassland Administration on Bamboo & Rattan Science and Technology, Institute of Gene Science and Industrialization for Bamboo and Rattan Resources, International Centre for Bamboo and Rattan, Beijing 100102, China)

Abstract:［**Objective**］This study aimed to explore the application of cross-species single-cell integration methods in plants to uncover the transcriptional conservation and divergence of shoot apical meristem (SAM) cell types between *Phyllostachys edulis* (moso bamboo) and *Arabidopsis thaliana*.［**Method**］Based on identified orthologous gene pairs, we integrated single-cell transcriptomic data from the SAM of moso bamboo and arabidopsis, using canonical correlation analysis (CCA) for batch correction to eliminate species-specific technical biases. Uniform manifold approximation and projection (UMAP) was applied to visualize the distribution patterns of integrated cells. Furthermore, Spearman correlation analysis and identification of cross-species marker genes were performed to assess transcriptional similarity and

收稿日期: 2025−02−10　**修回日期:** 2025−02−28
基金项目:“十四五”国家重点研发计划（2021YFD2201000）。
第一作者: 孙镭。主要研究方向：植物单细胞转录组学。Email：raysun61@163.com　地址：100102 北京市朝阳区东大街 8 号国际竹藤中心竹藤资源基因科学与基因产业化研究所。
责任作者: 赵韩生，研究员，博士生导师。主要研究方向：竹子种质资源与基因组学。Email：zhaohansheng@icbr.ac.cn　地址：同上。
本刊网址: http://j.bjfu.edu.cn；http://journal.bjfu.edu.cn

functional characteristics of different cell types. [**Result**] (1) A total of 23 cell clusters were identified. In the UMAP space, most cells clustered separately based on their species origin, but some cell clusters showed significant overlap, suggesting high transcriptomic similarity. (2) Further analysis revealed that the overlapping region mainly consisted of epidermal and vascular tissue cells, indicating the evolutionary conservation of these tissues. Correlation analysis further supported the high similarity in gene expression patterns among conserved cell types, while other cell clusters exhibited greater species specificity and consisted of mixed cell types. (3) 183 cross-species conserved marker genes were identified. These genes were specifically expressed in corresponding cell types of both species. Functional analysis revealed their key roles in plant development and evolution. [**Conclusion**] This study revealed the core functional characteristics of conserved cell types between moso bamboo and arabidopsis, providing new perspectives on plant evolutionary mechanisms and laying a data foundation for functional gene discovery and research in non-model plants.

Key words: *Phyllostachys edulis*; *Arabidopsis thaliana*; orthologous genes; cross-species integration; single-cell transcriptomics; conserved cell types; marker genes

高等植物的地上结构都起源于茎顶端分生组织(shoot apical meristem，SAM)，SAM 通过持续的细胞分裂和分化产生新的组织和器官，从而维持植物的生长与发育[1]。所以，研究 SAM 对深入理解植物的发育进程和形态建成具有重要意义。近年来，单细胞转录组测序技术(包括 scRNA-seq 和 snRNA-seq)因突破了传统转录组测序技术(Bulk RNA-seq)仅能反映组织所有细胞平均基因表达水平的局限而快速发展，为探索这些起始细胞提供了强大的工具，使研究者能在单细胞分辨率下解析基因表达的异质性和动态变化[2]。目前，已有研究对玉米(*Zea mays*)[3]、拟南芥(*Arabidopsis thaliana*)[4]、毛竹(*Phyllostachys edulis*)[5]等模式植物或经济作物的 SAM 开展单细胞测序分析，成功鉴定出多种细胞类型及其特异基因。然而，鉴于植物细胞类型的复杂性和物种间的差异，一些重要的标志基因和分子机制在不同物种间的表达模式并不相同。例如，拟南芥中通过 CLV3-WUS 途径来保证 SAM 的正常维持和分化，但该途径在玉米中并不适用[3,6]。因此，对不同物种的 SAM 细胞进行比较分析，有助于揭示植物发育过程中基因表达特征的共性和差异，识别保守的标志基因并解析其发育调控机制。

相比于动物中大量基于单细胞的跨物种分析，植物领域的跨物种单细胞研究目前只在根尖细胞中有涉及[7−8]。Liu 等[7]通过整合拟南芥和水稻(*Oryza sativa*)的根尖细胞发现：根毛和皮层细胞在两个物种中具有相似性，并且基因 *Proline-rich protein-like* 1 (*PRPL*1)在两个物种的根毛中都特异表达；而其表皮、内皮层、中柱和根冠细胞则表现出显著差异，这些细胞在降维空间中并未重叠，且其标志基因的保守性也显著降低。Guillotin 等[8]则通过绘制玉米、高粱(*Sorghum bicolor*)和狗尾草(*Setaria italica*)的细胞图谱，鉴定出了跨物种保守的直系同源标记基因，并对不同物种的转录组分化速度进行了分析。这些研究表明，跨物种整合分析能够提供关于保守标志基因和特异功能的新认识，对于全面理解植物的进化和发育机制具有重要的实际意义。

毛竹是地球上生长速度最快的竹种之一，在快速生长期每天可以生长 1.14 m[9]。尽管毛竹和水稻、玉米等同属禾本科(Poaceae)，但是其茎秆高大且高度木质化，这些特性使其具有重要的经济价值[10]。毛竹的快速生长特性得益于其 SAM 中独特的细胞分裂与分化模式，这使得毛竹成为研究禾本科植物 SAM 的理想材料[5]。而拟南芥作为模式植物，拥有丰富的基因组和发育生物学研究基础。对毛竹和拟南芥的 SAM 进行单细胞转录组整合分析，不仅能够揭示二者在细胞类型和基因表达方面的异同，还有助于我们识别植物中高度保守的基因及其功能。

因此，本研究拟利用毛竹和拟南芥的 SAM 单细胞转录组测序数据，开展跨物种整合分析，系统比较二者 SAM 中的细胞组成及基因表达特征，旨在揭示毛竹与拟南芥 SAM 细胞类型在单细胞转录组水平上的保守性与差异性，挖掘重要标志基因，并深入解析其发育调控机制。研究结果将为理解植物 SAM 的发育演化过程提供新的视角，为植物跨物种单细胞研究提供参考框架，同时为未来的精准分子育种工作提供重要的候选基因资源。

1 数据收集与研究方法

1.1 数据收集

1.1.1 茎尖单细胞数据收集

拟南芥的茎尖单细胞转录组测序数据来自于

Beijing Institute of Genomics Data Center（http://bigd.big.ac.cn）中的 PRJCA003094 项目，包括 2 个不同发育阶段的茎尖组织和 1 个叶片样本[4]，数据格式为 10 × Genomics 单细胞测序技术的原始下机 FASTQ 文件。毛竹的单核转录组测序数据包括 3 个不同发育阶段的毛竹顶端分生组织[5]，文件下载于 Open Science Framework（https://osf.io/aj45e/?view_only=cdc05c8dae6d427b861aae2ffa4afd37），在 NCBI 上的项目号为 PRJNA1073297，数据格式为 10 × Genomics 单细胞测序技术生成的 barcodes.tsv.gz、features.tsv.gz 和 matrix.mtx.gz 基因计数文件。

1.1.2　基因组相关文件收集

本研究中用到的拟南芥基因组文件和拟南芥、水稻、玉米、高粱、杨树（*Populus trichocarpa*）的蛋白质序列文件均下载于 EnsemblPlants v60（https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-60/）。毛竹的蛋白质序列下载于 the GigaDB website（http://gigadb.org/dataset/view/id/100498/）[11]。

1.2　研究方法

1.2.1　数据预处理

参照 Zhang 等[4]研究，对拟南芥的下机 FASTQ 数据进行预处理。首先，使用“cellranger mkref”命令构建参考；随后，使用“cellranger count”命令结合“-r2-length = 98”参数生成单细胞基因计数文件（软件版本为 Cell Ranger 8.0.1）。R 包 Seurat 4.3.0[12]基于输出的基因计数文件创建 Seurat 对象（min.cells = 3，min.features = 200），并使用 DoubletFinder 2.0.4[13]检测去除双胞（测序中由于细胞重叠或捕获错误导致的假双细胞），关键参数与原始文献[4]保持一致。直接使用 Seurat 包读取基因计数文件，创建毛竹的 Seurat 对象（min.cells = 3，min.features = 200）。

1.2.2　降维聚类和细胞注释

参照 Zhang 等[4]和 Guo 等[5]研究，对拟南芥和毛竹的数据进行降维聚类和细胞注释，由于软件版本和运行环境的不同，部分参数有所调整。拟南芥数据经过质控、标准化、高变基因筛选、细胞周期和原生质体基因回归、主成分分析（principal component analysis，PCA）、批次效应校正（Harmony 1.2.1[14]）、聚类（dims = 1∶50，res = 0.7）、UMAP 降维可视化（dims = 1∶50），最后利用已知的标志基因对细胞身份进行鉴定。类似地，对于毛竹，使用 Seurat 中的典型相关分析（canonical correlation analysis，CCA）方法对数据进行整合（nfeatures = 2 000，dims = 1∶30）后，再使用 Harmony 去除批次效应，并进行聚类（dims = 1∶20）。最后利用空间转录组与标志基因对细胞进行注释，细胞注释信息储存在 Seurat 对象的 meta.data 中用于后续分析。

1.2.3　共享基因对生成

为了实现跨物种的比较，需要统一两个物种的基因信息，以便后续的整合。参考 Song 等[15]的方法，将毛竹的基因替换为拟南芥中同源性最高的直系同源基因。具体步骤如下：首先，通过 OrthoFinder 2.5.4[16]对拟南芥、毛竹、玉米、水稻、高粱、杨树的蛋白序列进行同源物鉴定，以获得毛竹和拟南芥的直系同源物。接着，利用 DIAMOND 2.1.10[17]对毛竹和拟南芥的蛋白序列进行双向比对，筛选出最佳唯一比对结果。最后，将 DIAMOND 的比对结果与 OrthoFinder 的比对结果进行交叉验证，仅保留属于同一直系同源群（orthogroup）的基因对。

1.2.4　单细胞跨物种整合与聚类

基于上述鉴定出的基因对，将毛竹基因计数矩阵中的基因名称替换为相应的拟南芥直系同源基因名称。同时，对拟南芥的基因计数矩阵进行筛选，仅保留基因对中存在的基因。接着，对每个样本的基因计数矩阵进行标准化处理，并识别高变异基因（nfeatures = 2 000）。随后，利用 Seurat 的 CCA 方法，基于鉴定出的高变异基因，寻找跨物种的锚点（nfeatures = 2 000，dims = 1∶30），并进行数据整合（dims = 1∶30）。整合结果存储在 integrated assay 中，作为后续分析的基础。基于整合后的数据，进一步进行细胞周期基因和原生质体相关基因的得分归一化。然后，利用 PCA 进行线性降维（dims = 1∶30），后进行聚类分析（resolution = 0.9），并使用 UMAP 方法（dims = 1∶30）进行降维可视化。

1.2.5　细胞注释与相关性分析

首先，根据每个类群（cluster）中物种的占比来确定其来源：当某物种的占比超过 85% 时，定义该 cluster 来源于该物种；否则，将其定义为混合来源（记为“共有”）。接着，将两个物种原始 Seurat 对象中的 meta.data 信息添加到整合后的 Seurat 对象中，以获取每个 cluster 的细胞类型组成。根据每个 cluster 中细胞类型的比例进行注释：当一个细胞类型占主导时，直接注释为该类型；当多种细胞类型比例接近时，则标记为混合状态。随后，利用 Seurat 的 AggregateExpression 函数（slot = “integrated”），按照注释信息分组生成伪 Bulk 转录组表达矩阵。基于伪 Bulk 表达数据，使用 cor 函数计算不同细胞类型的 Spearman 相关性系数，并使用 ComplexHeatmap 包进行可视化，展示跨物种细胞类型间的基因表达相关性。

1.2.6　差异基因计算和功能分析

对于拟南芥的标志基因，使用 Seurat 包中的

FindAllMarkers 函数计算拟南芥原始数据中不同细胞类型的差异基因(min.pct = 0.1，logfc.threshold = 0.25)，按照 diff_pct 排序，选取排名靠前的基因进行可视化。对于跨物种保守基因的鉴定，使用 FindMarkers 函数对整合后的数据计算差异基因(logfc.threshold = 0.5，min.pct = 0.1)，例如对于 cluster8，首先计算 cluster8 中毛竹细胞(moso_cluster8)与其他类群细胞的差异基因，再计算 cluster8 中的拟南芥细胞(at_cluster8)与其他类群细胞的差异基因，最后取交集得到 cluster8 的跨物种保守的标志基因(保守基因)。使用 R 包 clusterProfiler[18] 和 enrichplot(https://doi.org/doi:10.18129/B9.bioc.enrichplot)的 treeplot 函数对生成的基因集进行 Gene Ontology 术语的统计与可视化。

2 结果与分析

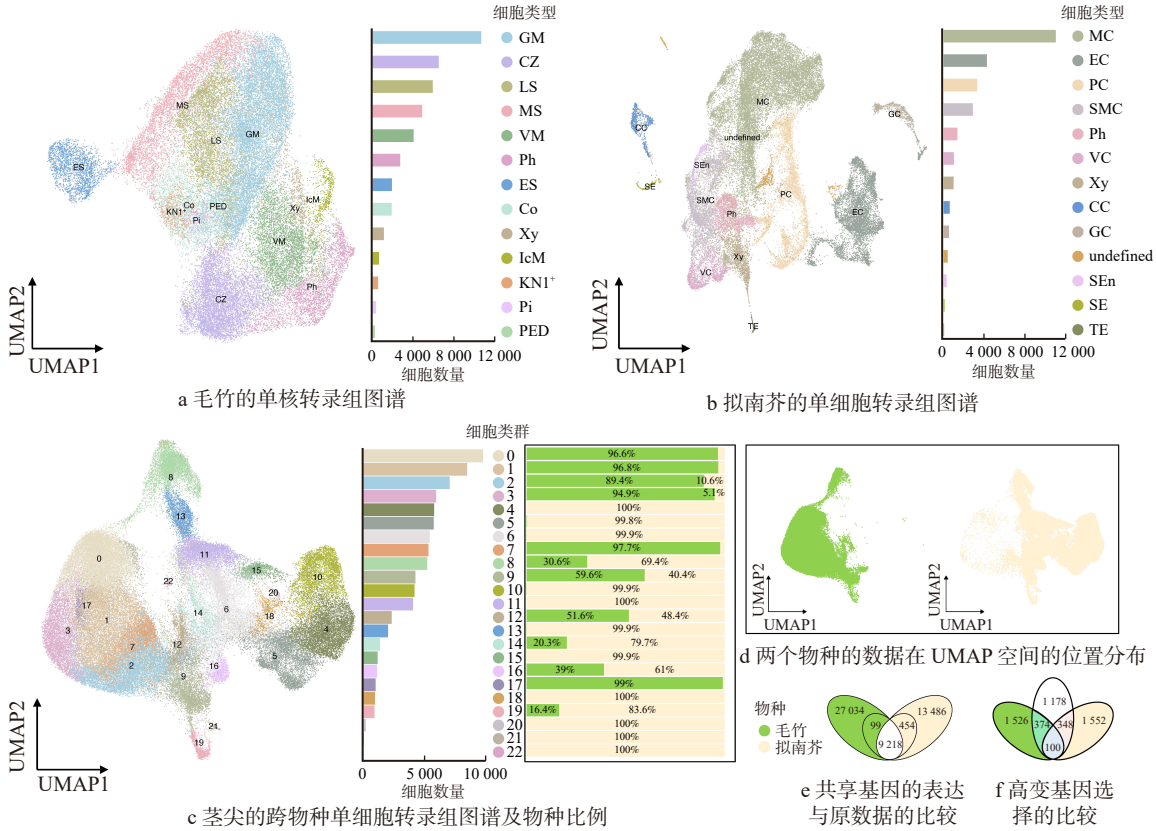
2.1 茎尖的跨物种单细胞转录组图谱构建

对毛竹的单核转录组(42 064 个细胞核)和拟南芥的单细胞转录组(40 804 个细胞)数据进行降维聚类和注释，分别鉴定出 13 种毛竹和 13 种拟南芥细胞类群(cluster)，包括了 13 种毛竹和 11 种(叶肉细胞 MC 和未知 undefined 类型除外)拟南芥的茎尖细胞类型(图 1a，b)。根据序列相似性对直系同源基因进行筛选，共获得 9 849 个一对一的直系同源基因对。基于这些基因对的表达矩阵对两个物种的数据进行整合，鉴定出 23 个 clusters(0 ~ 22，图 1c)，其中有 9 218 个基因在两个物种中均有表达(图 1e)。与原始数据相比，整合数据中分别包括了 474 个毛竹和 448 个拟南芥的高变基因(图 1f)。

UMAP 可视化显示，整合后的细胞群主要依据物种来源分布于两个区域，其中 cluster 8、9、12、14、16、19 位于 UMAP 图的中间位置，并且两个物种的这些细胞高度重叠(图 1c，d)。此外，这些 clusters 中两个物种细胞所占的比例相近，而其他 clusters 中的细胞则基本只来自某一个物种(图 1c，表 1)，表明这些位于中间区域的 clusters 可能代表了物种间功能保守的细胞类型。

2.2 细胞类型注释及相关性比较

对整合后的细胞进行注释发现：cluster 0、1、2、3、7、17 主要来源于毛竹，其中 cluster 0、1、2、3、



a 毛竹的单核转录组图谱　b 拟南芥的单细胞转录组图谱

c 茎尖的跨物种单细胞转录组图谱及物种比例　d 两个物种的数据在 UMAP 空间的位置分布

e 共享基因的表达与原数据的比较　f 高变基因选择的比较

a 图中：GM. 基本分生组织；CZ. 冷区细胞(分裂缓慢且转录活性较低的细胞群)；LS. 晚期笋鞭；MS. 中期笋鞭；VM. 维管分生组织；Ph. 韧皮部；ES. 早期笋鞭；Co. 皮层；Xy. 木质部；IcM. 居间分生组织；KN1+. 与标记基因 *KN1* 相关的分生组织细胞；Pi. 髓；PED. 原形成层富集区。
b 图中：MC. 叶肉细胞；EC. 表皮细胞；PC. 增殖细胞；SMC. 茎分生细胞；Ph. 韧皮部；VC. 维管细胞；Xy. 木质部；CC. 伴胞；GC. 保卫细胞；undefined. 未知；SEn. 茎内皮层；SE. 筛管；TE. 导管。

图 1　茎尖的跨物种单细胞转录组图谱构建

Fig. 1　Construction of cross-species single-cell transcriptomic atlas of shoot apex

表 1　毛竹和拟南芥茎尖单细胞图谱的细胞类群注释

Tab. 1　Cell type annotation of single-cell atlas of shoot apex in moso bamboo and arabidopsis

细胞类群	拟南芥细胞数量(占比)	毛竹细胞数量(占比)	主要物种来源	主要毛竹细胞类型	主要拟南芥细胞类型
0	330(3.39%)	9 413(96.61%)	毛竹	冷区/基本分生组织/晚期笋鞭	
1	275(3.25%)	8 184(96.75%)	毛竹	基本分生组织/晚期笋鞭/中期笋鞭	
2	747(10.6%)	6 297(89.4%)	毛竹	基本分生组织/维管分生组织	
3	304(5.11%)	5 642(94.89%)	毛竹	基本分生组织/中期笋鞭/晚期笋鞭	
4	5 786(99.97%)	2(0.03%)	拟南芥		叶肉细胞
5	5 741(99.76%)	14(0.24%)	拟南芥		增殖细胞/叶肉细胞
6	5 426(99.89%)	6(0.11%)	拟南芥		茎分生细胞/叶肉细胞
7	124(2.33%)	5 204(97.67%)	毛竹	基本分生组织/中期笋鞭/皮层/*KN1*相关的分生组织	
8	**3 632(69.45%)**	**1 598(30.55%)**	**共有**	**早期笋鞭**	**表皮细胞**
9	**1 724(40.39%)**	**2 544(59.61%)**	**共有**	**韧皮部**	**韧皮部/增殖细胞**
10	4 195(99.86%)	6(0.14%)	拟南芥		叶肉细胞
11	4 079(100%)	0(0%)	拟南芥		叶肉细胞/表皮细胞/增殖细胞
12	**1 137(48.42%)**	**1 211(51.58%)**	**共有**	**维管分生组织**	**木质部**
13	2050(99.9%)	2(0.1%)	拟南芥		表皮细胞
14	**1 117(79.67%)**	**285(20.33%)**	**共有**	**居间分生组织**	**维管细胞/茎分生细胞**
15	1 211(99.92%)	1(0.08%)	拟南芥		叶肉细胞
16	**707(60.95%)**	**453(39.05%)**	**共有**	**韧皮部**	**韧皮部**
17	10(0.95%)	1 044(99.05%)	毛竹	晚期笋鞭	
18	995(100%)	0(0%)	拟南芥		叶肉细胞 /茎分生细胞
19	**803(83.56%)**	**158(16.44%)**	**共有**	**韧皮部**	**伴胞**
20	247(100%)	0(0%)	拟南芥		保卫细胞
21	106(100%)	0(0%)	拟南芥		筛管
22	58(100%)	0(0%)	拟南芥		未知

注：空白表示没有对应的细胞类型；加粗的为两物种共有的保守细胞类群。

7 为多种细胞的混合类型，而 cluster 17 主要由晚期笋鞭细胞组成；cluster 4、5、6、10、11、13、15、18、20、21 和 22 主要来源于拟南芥，其中 cluster 4、10、15 被注释为叶肉细胞，13、20 和 21 分别注释为表皮细胞、保卫细胞和筛管，其他 clusters 为多种细胞的混合类型(表 1，图 2a)。在 UMAP 图中高度重叠的 cluster 8、9、12、14、16 和 19，在毛竹中分别主要被注释为早期笋鞭、韧皮部、维管分生组织、居间分生组织、韧皮部、韧皮部，在拟南芥中则分别主要被注释为表皮细胞、韧皮部或增殖细胞、木质部、维管细胞或茎分生细胞、韧皮部、伴胞(图 2b，c)。

对毛竹和拟南芥 clusters 的基因表达情况进行相关性分析，结果(图 2d)显示：在混合细胞类群中，毛竹的 cluster 7、2、1、0 与拟南芥的混合类群 cluster 5、10、14、18 整体相关性最强。而共有的细胞类群中，毛竹的 cluster 8、14、16、19 分别与拟南芥对应的 cluster 8、14、16、19 相关性最强。此外，图中所有相关性系数的 p 值均等于 0，表明这些相关性在统计上具有高度显著性。以上结果进一步表明在 UMAP 空间重叠的 clusters 在转录组水平上具有较高的相似性，并且主要来源于表皮和维管组织。

2.3　保守细胞类型的基因表达特征

对共有的 6 种保守细胞类群(cluster 8、9、12、14、16、19)进行分析(图 2a)。在之前的注释中，cluster 8 被分别注释为拟南芥的表皮细胞和毛竹的早期笋鞭(表 1)。从解剖学的角度来看，笋鞭源自最外层细胞的分化[5]，可以认为早期笋鞭是未完全成熟的表皮细胞，因此将 cluster 8 统一定义为表皮细胞(EC)。毛竹和拟南芥的 cluster 9 都被注释为韧皮部，且都包含了一些增殖活跃的细胞(维管分生组织、增殖细胞)，因此将 cluster 9 定义为韧皮部形成层(PPh)。同理，cluster 12、16 和 19 分别定义为木质部(Xy)、韧皮部(Ph)和伴胞(CC)。对于 cluster 14，在毛竹中被注释为居间分生组织，而在拟南芥中被

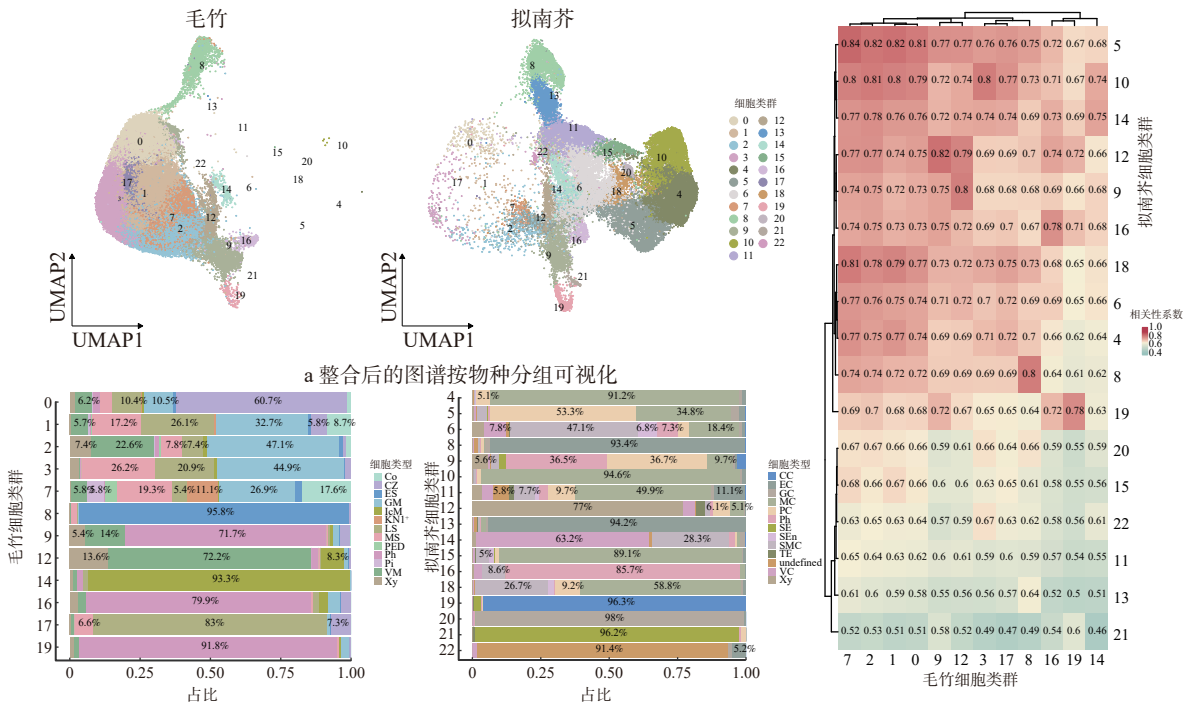


图 2 两个物种细胞类群的细胞类型成分与相关性比较

Fig. 2 Comparison of cell type composition and correlations between cell clusters of two species

注释为维管细胞/茎分生细胞。居间分生组织在单子叶植物中通过快速分裂活动促进节间的伸长[19]，和维管细胞、茎分生细胞一样，都具有增殖和分生能力，因此将 cluster 14 统一定义为分生组织(Me)。

以上的细胞类型都能被拟南芥中的特异标志基因所验证(图 3)。例如，表皮细胞(EC)高表达拟南芥中的表皮标志基因 *AT2G47240*(*CER8*)，韧皮部形成层(PPh)高表达拟南芥中的增殖细胞标志基因 *AT3G15680*，韧皮部细胞(Ph)则高表达拟南芥中的韧皮部标志基因 *AT1G52140* 等。值得注意的是，拟南芥中的部分特异标志基因在毛竹中并不适用。如表皮标志基因 *AT5G19290*(*MAGL16*)只在拟南芥的表皮细胞中高表达，伴胞标志基因 *AT1G10120*(*CIB4*)也仅在拟南芥中高表达，其他 cluster 中也都存在这种情况。虽然从整体水平上看，这 6 类细胞类型相对保守，但是依旧存在着物种间的差异。

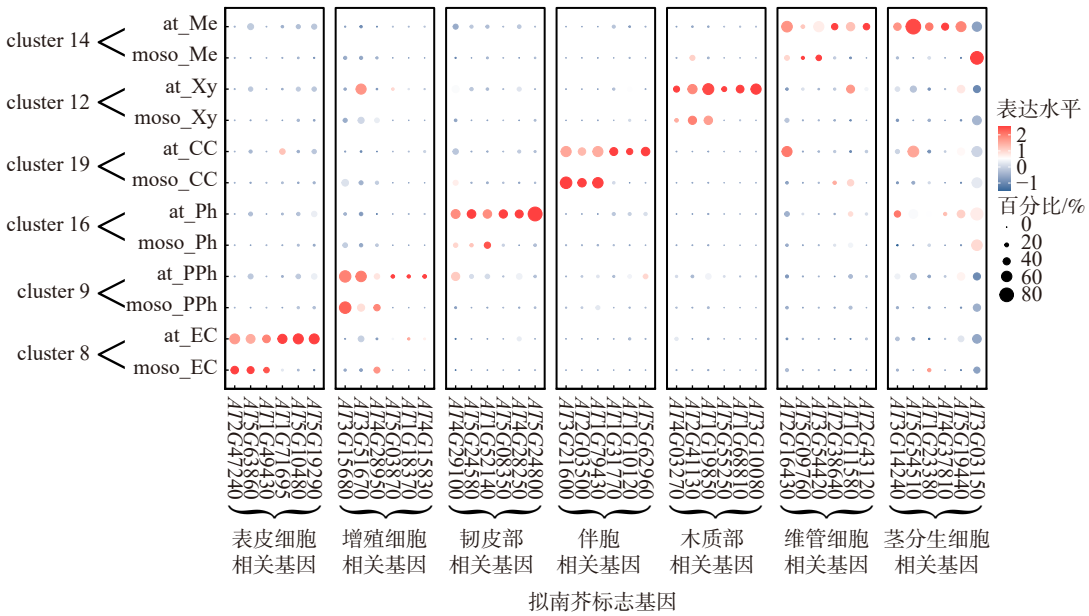


图 3 拟南芥细胞标志基因在保守细胞类群中的表达

Fig. 3 Expression of arabidopsis cell marker genes in conserved cell clusters

2.4　保守基因鉴定及功能分析

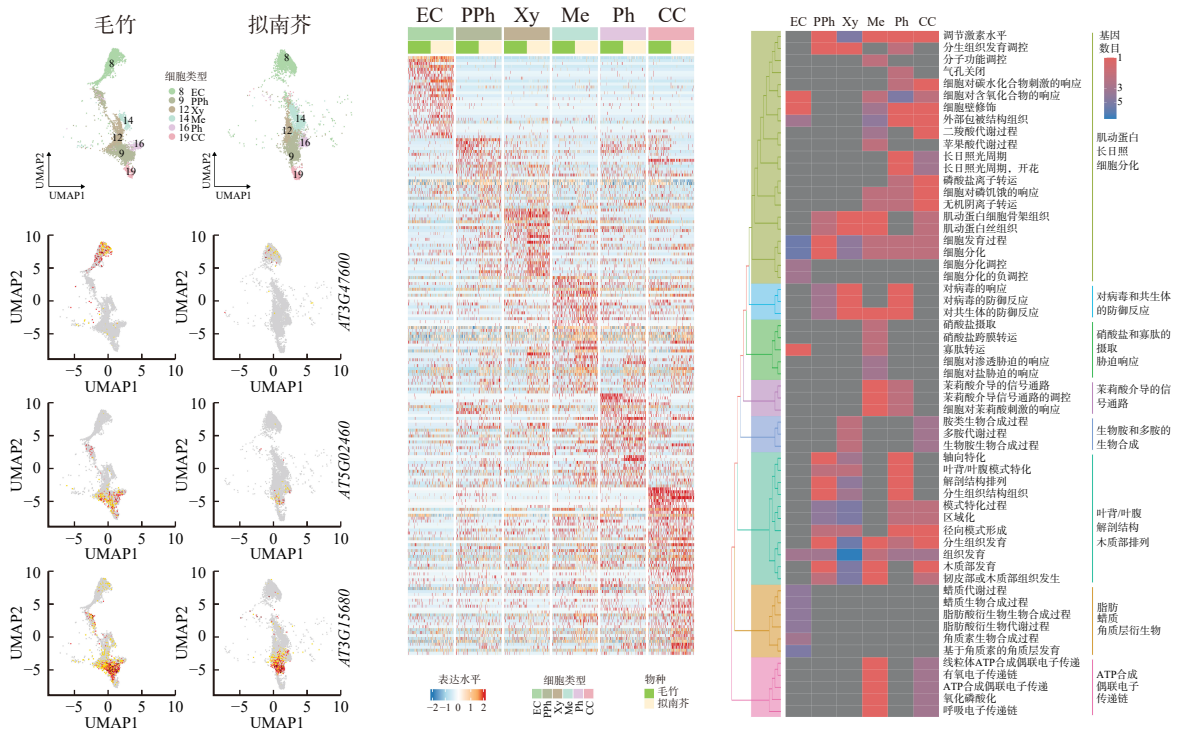
进一步对两个物种保守细胞类群的差异基因进行筛选，共鉴定出 6 类保守细胞类群的 183 个跨物种保守标志基因（表 2）。从数量上看，大部分基因在两个物种间是不保守的。例如，在表皮细胞类群（cluster 8）中，拟南芥和毛竹分别鉴定出 771 和 685 个差异基因，但是只有 28 个基因是共享的，这一部分重叠的基因反映了物种间进化保守性的重要特征。例如，*AT3G47600*（*MYB94*）参与表皮蜡质生物合成[20]，在两个物种的表皮细胞类群（cluster 8）均特异表达（图 4a）；而 *AT5G02460*（*PEAR2*）是一种调控韧皮部发育的关键转录因子[21]，在韧皮部形成层（cluster 9）和韧皮部（cluster 16）中表达强烈（图 4a）。此外，*AT3G15680* 也同为两个物种韧皮部形成层（cluster 9）的标志基因，已被证实在拟南芥和杨树的韧皮部形成层中过表达[22]。

图 4b 展示了所有保守基因在不同细胞类群中的表达情况，这些基因在同一个 cluster 中的两个物种中表现一致。对这些保守基因的生物学功能进行分析（图 4c），结果显示：表皮细胞类群（cluster 8）的

表 2　保守细胞类型注释与保守基因鉴定

Tab. 2　Conserved cell type annotation and conserved gene identification

细胞类群	细胞类型	毛竹差异基因数量	拟南芥差异基因数量	保守基因数量	每种细胞类群的前5个保守基因	
					毛竹	拟南芥
cluster 8	表皮细胞 (EC)	685	771	28	*PH02Gene35874* *PH02Gene40895* *PH02Gene27506* *PH02Gene43735* *PH02Gene04338*	*AT2G38540* (*LTP1*) *AT2G26250* (*KCS10*) *AT3G47600* (*MYB94*) *AT1G17200* (*CASPL2A1*) *AT2G47240* (*LACS1*)
cluster 9	韧皮部形成层 (PPh)	800	863	24	*PH02Gene47962* *PH02Gene42087* *PH02Gene04163* *PH02Gene42526* *PH02Gene18974*	*AT3G15680* *AT3G61670* *AT1G73850* *AT5G02460* (*PEAR2*) *AT5G43810* (*AGO10*)
cluster 12	木质部 (Xy)	743	779	23	*PH02Gene38420* *PH02Gene29179* *PH02Gene00395* *PH02Gene11634* *PH02Gene08468*	*AT2G34710* (*PHB*) *AT2G41130* (*bHLH106*) *AT1G19850* (*MP*) *AT5G60690* (*REV*) *AT1G45145* (*TRX5*)
cluster 14	分生组织 (Me)	729	628	40	*PH02Gene00458* *PH02Gene06790* *PH02Gene32562* *PH02Gene14153* *PH02Gene08468*	*AT5G20900* (*JAZ12*) *AT2G24260* (*LRL1*) *AT4G37870* (*PCK1*) *AT2G16700* (*ADF5*) *AT1G45145* (*TRX5*)
cluster 16	韧皮部 (Ph)	750	705	32	*PH02Gene35120* *PH02Gene25754* *PH02Gene09203* *PH02Gene42526* *PH02Gene28947*	*AT4G08290* (*UMAMIT20*) *AT5G64700* (*UMAMIT21*) *AT1G52140* *AT5G02460* (*PEAR2*) *AT4G29080* (*PAP2*)
cluster 19	伴胞 (CC)	841	567	57	*PH02Gene46403* *PH02Gene45375* *PH02Gene39987* *PH02Gene39870* *PH02Gene09030*	*AT3G04030* (*MYR2*) *AT3G21600* *AT1G79430* (*APL*) *AT4G19840* (*PP2-A1*) *AT2G03500* (*EFM*)



a 部分已知功能基因的表达特征　b 所有保守基因在不同细胞类型中的表达水平　c 保守基因的 GO 生物学过程分类统计

图 4　保守基因的表达模式和功能分析

Fig. 4　Expression patterns and functional analysis of conserved genes

保守基因主要与角质和脂肪酸的生物合成相关；韧皮部形成层（cluster 9）、木质部（cluster 12）、韧皮部（cluster 16）、伴胞（cluster19）的保守基因中都包含了与韧皮部与木质部发育及分生结构排列模式相关的功能；而分生组织（cluster 14）单独显示出硝酸盐跨膜运输与渗透胁迫反应的功能等。这些结果展示了两个物种中保守细胞类群共有的重要生物学过程。

3　讨　　论

随着单细胞测序技术的发展，跨物种整合分析为解析细胞类型保守性与功能演化提供了新的视角。毛竹和拟南芥分别作为禾本科与十字花科的代表性植物，有着显著不同的生长发育方式。对其茎尖分生组织（SAM）的单细胞转录组进行跨物种整合分析，有助于揭示植物干细胞维持与分化过程中潜在的保守机制，为理解植物形态多样性提供分子基础。然而，在植物领域，尤其是非模式物种中，此类研究仍处于起步阶段。

对于跨物种整合，首先面临的挑战是数据复现的复杂性。由于软件版本和计算环境的差异，数据的重现变得异常繁琐[23]。本研究投入了大量时间对数据的复现和注释进行优化，以确保整合分析的可靠性。其次，共享基因筛选、批次效应矫正等步骤直接影响了整合结果的准确性和生物学解释能力[24]。由于植物基因组多倍化和基因重复事件频繁[25]，直系同源基因常以多对多形式存在，这使得选择准确的共享基因对成为关键问题。目前，共享基因的选择主要根据同源物的同源性和表达量进行筛选[15]，这种方法在哺乳动物中通常能够得到 8 000 左右的保守基因对[26]。尽管最近的研究[27]针对植物提出了基于共表达网络筛选最相关同源基因的方法，但是只提供了少数植物的基因对库（例如玉米和拟南芥有 4 899 对[28]）。本研究也尝试过基于该方法生成适用于毛竹的共享基因集，最终获得的基因对数量有限（宽松的阈值下得到 3 506 个共享基因对），且整合效果不佳，表现为细胞类群识别模糊、种类偏少。对于整合算法，另外一种常用到的 Harmony 算法[14]也被尝试。然而，Harmony 算法在保留物种特异性信号的同时未能很好识别细胞类型间的共享特征，导致整合结果呈现出强烈的物种区分。因此，我们最终基于序列相似性筛选出的直系同源基因对，并利用 Seurat 中的 CCA 算法对数据进行整合。结果表明，当物种间差异较大时，使用更全面的共享基因空间和具有适当矫正力度的算法是获得期望整合结果的关键。

此外，高变基因是单细胞降维聚类时使用的主要特征[12]。在本研究中，用于整合的 2 000 个高变基因中仅有 822 个与原始数据一致（图 1f）。这种高变

基因选择上的不一致可能导致特定基因特征的丢失，从而影响细胞类群的区分能力（如毛竹的 cluster 0、1、2、3、4 未能被准确区分）。同时，单细胞分析中的细胞类群划分主要是基于基因表达的特征，而传统解剖学方法则通过观察细胞位置和形态分类，二者无法完全对应[29]（如毛竹中的 $KN1^+$ 细胞）。并且由于物种本身的差异，两个物种的 SAM 细胞类型分类结果也并不一致。例如，毛竹中的早期笋鞭在拟南芥中并不存在，但其与拟南芥的表皮细胞同位于 SAM 的最外层，且具有相似的基因表达特征；而拟南芥中的茎内皮层在毛竹中没有对应的细胞类型。所以，对整合后的细胞类群进行注释需具有丰富的生物学经验，尤其是在处理混合状态的细胞群时，若缺少原始身份信息，细胞类型的判断难度会进一步加大。

进一步分析保守细胞类型的基因表达特征发现，对于 cluster 14（Me），来自拟南芥的标志基因在毛竹中大多表现为非特异性（图 3），这也反映了两个物种在分生组织发育模式上的显著差异。此外，通过跨物种转录组图谱，将毛竹韧皮部细胞细分成了更多的细胞类型，并且成功利用拟南芥的基因鉴定出保守的伴胞（CC），这为细胞注释提供了可以参考的方法。进一步的保守基因鉴定展示了两个物种间进化保守的生物学功能。除了一些已有研究基础的基因外，我们还鉴定出了一些未被研究的跨物种保守基因。例如，*AT*1*G*52140 是韧皮部的保守基因，但尚无相关研究报道。值得注意的是，保守细胞类型的差异基因（即被舍弃的不保守基因）在两个物种中展示出了一些共同的特征。具体而言，拟南芥中的不保守基因普遍富集到“线粒体 ATP 合成”“细胞呼吸”等高能量代谢通路，暗示其通过爆发式氧化供能支持细胞活动；而毛竹中则一致偏向“大分子合成负调控”“RNA 剪接”等基因表达精密调控通路，揭示其通过全局代谢抑制与转录后修饰维持稳态。这些结果反映了两个物种在进化历程中为适应各自的生长发育模式、生理需求以及生态环境，在基因功能层面发生了分化。拟南芥侧重于能量的高效产生与利用，以满足其相对快速的生长繁殖需求；毛竹则通过精细的基因调控机制，实现对大分子合成等过程的精准控制，保障自身在复杂环境下的持续生长与发育。

总之，本研究首次实现了毛竹和拟南芥 SAM 的单细胞转录组整合分析，识别出茎尖中的多个保守细胞类型，并鉴定出潜在的共享核心基因，为深入理解植物进化机制提供了新视角，也为构建更广泛的植物单细胞图谱提供了可参考的方法框架。但是，当前植物 SAM 单细胞转录组研究仍处于起步阶段。到目前为止，仅有玉米、拟南芥和毛竹的数据被公布。由于玉米数据集的细胞捕获量有限，为了使整合结果更具可比性，本研究选择了毛竹和拟南芥的茎尖分生组织进行整合，这在一定程度上限制了研究的广度。其次，非模式植物缺乏可直接使用的同源基因数据库，导致共享基因的选择和整合策略受限。因此，未来的研究可以扩展到更多物种，构建更全面的植物基因同源关系数据库，并开发出更适合植物的跨物种整合方法，进一步提高研究结果的准确性和普适性。

4　结　　论

本研究为了探索不同物种茎尖分生组织（SAM）细胞功能的保守性与差异性，基于直系同源基因对，系统比较了毛竹和拟南芥的 SAM 单细胞转录组数据，研究主要得到以下结论。

（1）共鉴定出 23 个细胞类群，在 UMAP 可视化空间中，大部分细胞按物种来源聚集在不同区域，但部分细胞类群在两个物种间出现明显重叠，表明这些细胞在转录组水平上具有较高相似性。

（2）进一步分析发现，重叠区域主要由表皮细胞和维管组织细胞组成，表明这类组织在植物演化过程中较为保守；相关性分析结果进一步证实了这些保守细胞类型之间基因表达的相似性。

（3）鉴定到 183 个跨物种保守的标志基因，这些基因在两个物种中的特定细胞类型中均特异表达，揭示了其在植物发育和进化中的关键作用。其中包含尚未被深入研究的基因（如 *AT*1*G*52140、*AT*3*G*15680 等），这为后续的生物学功能研究提供了候选靶点。

以上研究结果揭示了两个物种保守细胞类型的核心功能，为理解植物的进化机制提供了新的视角，也为植物跨物种单细胞研究提供了参考框架，同时为非模式植物的功能基因挖掘和精准分子育种研究提供了重要的候选基因资源。

参　考　文　献

[1] 薛红卫, 焦雨铃, 徐通达, 等. 植物细胞与发育生物学研究进展 [J]. 植物生理学报, 2023, 9(12): 2145−2187.
Xue H W, Jiao Y L, Xu T D, et al. Recent advances in plant cell and developmental biology[J]. Plant Physiology Journal, 2023, 9(12): 2145−2187.

[2] 安妍妍, 张明惠, 谭何新. 单细胞测序在植物研究中的应用 [J]. 科学通报, 2024, 69(12): 1586−1597.
An Y Y, Zhang M H, Tan H X. The application of single-cell sequencing in plant research[J]. International Journal of Molecular Sciences, 2024, 69(12): 1586−1597.

[3] Satterlee J W, Strable J, Scanlon M J. Plant stem-cell organization

and differentiation at single-cell resolution[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(52): 33689−33699.

[4] Zhang T Q, Chen Y, Wang J W. A single-cell analysis of the Arabidopsis vegetative shoot apex[J]. Developmental Cell, 2021, 56(7): 1056−1074.

[5] Guo J, Luo D, Chen Y, et al. Spatiotemporal transcriptome atlas reveals gene regulatory patterns during the organogenesis of the rapid growing bamboo shoots[J]. New Phytologist, 2024, 244(3): 1057−1073.

[6] Dinneny J R, Benfey P N. Plant stem cell niches: standing the test of time[J]. Cell, 2008, 132(4): 553−557.

[7] Liu Q, Liang Z, Feng D, et al. Transcriptional landscape of rice roots at the single-cell resolution[J]. Molecular Plant, 2021, 14(3): 384−394.

[8] Guillotin B, Rahni R, Passalacqua M, et al. A pan-grass transcriptome reveals patterns of cellular divergence in crops[J]. Nature, 2023(617): 785−791.

[9] Chen M, Guo L, Ramakrishnan M, et al. Rapid growth of moso bamboo (*Phyllostachys edulis*): cellular roadmaps, transcriptome dynamics, and environmental factors[J]. Plant Cell, 2022, 34(10): 3577−3610.

[10] 汤雯欣, 顾一鸣, 刘昊玮等. 中国毛竹材资源现状研究 [J]. 四川建材, 2021, 47(12): 11−12.
Tang W, Gu Y, Liu Hi, et al. Research on current situation of bamboo resources in China[J]. Sichuan Building Materials, 2021, 47(12): 11−12.

[11] Zhao H, Gao Z, Wang L, et al. Chromosome-level reference genome and alternative splicing atlas of moso bamboo (*Phyllostachys edulis*)[J]. GigaScience, 2018, 7(10): 1−12.

[12] Hao Y, Hao S, Erica A, et al. Integrated analysis of multimodal single-cell data[J]. Cell, 2021, 184(13): 3573−3587.

[13] Mcginnis C S, Murrow L M, Gartner Z J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors[J]. Cell Systems, 2019, 8(4): 329−337.

[14] Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony[J]. Nature Methods, 2019, 16(12): 1289−1296.

[15] Song Y, Miao Z, Brazma A, et al. Benchmarking strategies for cross-species integration of single-cell RNA sequencing data[J]. Nature Communications, 2023, 14(1): 6495.

[16] Emms D M, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics[J]. Genome Biology, 2019, 20(1): 238.

[17] Buchfink B, Reuter K, Drost H G. Sensitive protein alignments at tree-of-life scale using DIAMOND[J]. Nature Methods, 2021, 18(4): 366−368.

[18] Yu G, Wang L G, Han Y, et al. clusterProfiler: an R rackage for comparing biological themes among gene clusters[J]. OMICS: A Journal of Integrative Biology, 2012, 16(5): 284−287.

[19] Nagai K, Mori Y, Ishikawa S, et al. Antagonistic regulation of the gibberellic acid response during stem growth in rice[J]. Nature, 2020, 584(7819): 109−114.

[20] Lee S B, Suh M C. Cuticular wax biosynthesis is up-regulated by the MYB94 transcription factor in *Arabidopsis*[J]. Plant and Cell Physiology, 2015, 56(1): 48−60.

[21] Qian P, Song W, Zaizen-Iida M, et al. A Dof-CLE circuit controls phloem organization[J]. Nature Plants, 2022, 8(7): 817−827.

[22] Cho H, Cho H S, Nam H, et al. Translational control of phloem development by RNA G-quadruplex–JULGI determines plant sink strength[J]. Nature Plants, 2018, 4(6): 376−390.

[23] Puntambekar S, Hesselberth J R, Riemondy K A, et al. Cell-level metadata are indispensable for documenting single-cell sequencing datasets[J]. PLoS Biology, 2021, 19(5): e3001077.

[24] Shafer M E R. Cross-species analysis of single-cell transcriptomic data[J]. Frontiers in Cell and Developmental Biology, 2019, 7: 175.

[25] 张以星, 齐浥颖, 王宇浩, 等. 多倍体植物基因组学研究进展 [J]. 基因组学与应用生物学, 2024, 43(8): 1293−1302.
Zhang Y X, Qi Y Y, Wang Y H, et al. Advances in polyploid plant genomics research[J]. Genomics and Applied Biology, 2024, 43(8): 1293−1302.

[26] Zernecke A, Erhard F, Weinberger T, et al. Integrated single-cell analysis-based classification of vascular mononuclear phagocytes in mouse and human atherosclerosis[J]. Cardiovascular Research, 2023, 119(8): 1676−1689.

[27] Passalacqua M J, Gillis J. Coexpression enhances cross-species integration of single-cell RNA sequencing across diverse plant species[J]. Nature Plants, 2024, 10(7): 1075−1080.

[28] Xu X, Passalacqua M, Rice B, et al. Large-scale single-cell profiling of stem cells uncovers redundant regulators of shoot development and yield trait variation[J]. BioRxiv, 2024, 6: 583414.

[29] Northcutt A J, Kick D R, Otopalik A G, et al. Molecular profiling of single neurons of known identity in two ganglia from the crab *Cancer borealis*[J]. Proceedings of the National Academy of Sciences, 2019, 116(52): 26980−26990.

（责任编辑 赵田芸
责任编委 卢孟柱）