



2024年第26期总378期

## 农业生物技术专题

### 本期导读

#### ▶ 学术文献

1. 四川农大利用生理/转录组分析揭示芦笋嫩茎生长转录响应机制
2. 中国农业大学汪海团队实现转录调控序列的人工智能设计
3. 西北农林科技大学开发了创新的深度学习模型预测表型
4. 农科院蔬菜花卉所利用转录组/代谢组解析番茄风味物质机制
5. 山东农大解析醋栗番茄基因组

中国农业科学院农业信息研究所

联系人：周诚昊;顾亮亮

联系电话：010-82109850

邮箱：[agri@ckcest.cn](mailto:agri@ckcest.cn)

2024年6月24日

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>

## 学术文献

### 1. 四川农大利用生理/转录组分析揭示芦笋嫩茎生长转录响应机制

**简介:** 2024年6月19日,四川农业大学郑阳霞教授等在BMC Plant Biology发表了一篇题为Physiological and transcriptome analysis of changes in endogenous hormone and sugar content during the formation of tender asparagus stems的研究论文。该研究对芦笋的生长模式进行了记录,并通过生理和转录组分析选择了四个连续伸长阶段的茎段进行研究,以深入了解芦笋茎在伸长和生长过程中的动态变化及其背后的机制。本研究记录了芦笋的生长模式,发现当芦笋长度达到25厘米时,其生长速率显著加快。在芦笋嫩茎的伸长区,蔗糖、果糖、葡萄糖及其他糖类的浓度显著下降。相反,生长素和赤霉素(GAs)水平升高,同时参与蔗糖降解的酶活性也增加。研究发现,生长素和赤霉素与蔗糖降解相关酶之间存在显著的正相关关系。随着茎的伸长,ABA(脱落酸)含量逐渐增加。组织切片显示,细胞伸长是茎伸长的内在表现。此外,通过转录组分析筛选出的差异基因富集在淀粉和蔗糖代谢、植物激素合成代谢和信号转导等路径中。在茎伸长过程中,ARF、GA20ox、NCED、PIF4等基因的表达水平上调,而DAO、GA2ox等基因的表达水平下调。基因表达水平与激素含量的变化一致,并影响细胞长度的伸长。综上,该研究揭示了芦笋嫩茎在伸长和生长过程中的基因表达水平、内源激素和糖类变化,为未来深入探讨芦笋茎生长发育的分子机制提供了宝贵的见解,并为芦笋的栽培和生产实践提供了理论基础。

**来源:** BMC Plant Biology

**发布日期:**2024-06-19

**全文链接:**

[http://agri.nais.net.cn/file1/M00/10/42/Csgk0GZ1EGeAEQQwAG3x\\_Ne8TII961.pdf](http://agri.nais.net.cn/file1/M00/10/42/Csgk0GZ1EGeAEQQwAG3x_Ne8TII961.pdf)

### 2. 中国农业大学汪海团队实现转录调控序列的人工智能设计

**简介:** 2024年6月18日,中国农业大学农学院汪海团队联合美国康奈尔大学、丹麦奥胡斯大学、北京大学现代农业研究院、坦桑尼亚农业科学院等单位,在PNAS在线发表了题为“Modeling 0.6 million genes for the rational design of functional cis-regulatory variants and de novo design of cis-regulatory sequences”的研究论文。该研究利用17个植物物种(拟南芥、毛果杨、大豆、甜菜、蒺藜苜蓿、黄瓜、葡萄、番茄、土豆、谷子、狗尾草、玉米、高粱、二穗短柄草、水稻、小立碗藓、莱茵衣藻)的60万个基因以及6256套转录组数据,开发了一个名为PhytoExpr的深度学习模型。该模型以近端转录调控区DNA序列(5kb启动子和5kb终止子)为输入,预测基因的中位数表达量以及该序列来自哪个物种。为了无偏倚地评估模型的预测准确度,以基因家族为单位划分训练集和测试集,确保模型只在它从未见过的基因家族上进行测试。研究者构建了两种模型结构:CNN+stacking和transformer,并训练随机森林模型对上述两种深度学习模型结构的超参数进行了优化,然后比较它们在四种不同任务中的预测精度。结果显示,采用transformer和多任务学习(multi-task learning)架构时,模型在mRNA丰度预测(图1d)和序列物种来源预测上都有着更高的准确度。为了评估PhytoExpr模型对新物种的泛化能力,研究者使用17个物种中的16个物种训练模型,在模型没有见过的新物种的新基因家族上测试其准确度。结果显示,除莱茵衣藻外,模型可以对从未见过的物种中的从未见过的基因家族作出可靠的预测。这表明PhytoExpr可以有效地拓展到与模型训练集包含的物种有相近进化关系的新物种中。而莱茵衣藻与其他植物分化了约10亿年,其转录调控的语法可能已经大

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>

相径庭，因此没有见过莱茵衣藻的深度学习模型无法对该物种触类旁通。作者进一步使用PhytoExpr评估了玉米每个基因的转录调控区每个碱基对基因表达的影响，并将碱基重要性得分与碱基保守性、转录因子结合位点进行了比较。结果显示，高保守碱基、与转录因子结合的碱基都显著富集在PhytoExpr预测的高重要性核苷酸中。这表明虽然PhytoExpr没有使用功能注释数据（例如Conserved Non-coding Sequence、ChIP-Seq等）进行训练，PhytoExpr仍能够识别出调控区序列中的功能元件。接下来作者使用PhytoExpr解决前文中提到的三个应用问题。首先，利用PhytoExpr模型系统评估了1730万个来自玉米HAPMAP3的SNP对基因表达的影响，发现模型预测的大效应SNP中显著富集罕见变异（MAF < 0.01），说明这些变异更可能受到负选择，同时SNP效应评分与SNP到TSS或TTS的距离有很强的相关性。因此研究者猜测PhytoExpr预测的大效应SNP可能与自然群体中基因表达水平的变异存在很强的相关性。利用以前报道的基于282个玉米自交系的7个不同组织转录组eQTL数据，依据cis-eQTL曼哈顿图中的关联强度对SNP进行排序，发现cis-eQTL中排序靠前的SNP在模型预测的大效应SNP中显著富集，证明PhytoExpr在转录调控区功能变异发掘中的有效性。第二，作者开发了两种算法利用PhytoExpr改进自然启动子。第一种策略称为基于重要性分数的半暴力算法，利用该方法对四个玉米调控区序列进行设计，并得到了预期的表达量。第二种策略称为虚拟进化，即在遗传算法中使用PhytoExpr作为选择算子。以植物启动子作为初始序列，通过对特定区域的多轮虚拟进化，该策略成功提高了目标启动子的转录强度。第三，作者利用虚拟进化技术实现了转录调控区DNA序列的从头设计。随机序列在计算机中经过多代虚拟进化，产生了适用于拟南芥的增强子和适用于玉米的增强子，这些增强子虽然和自然序列没有相似性，但是可以分别在拟南芥和玉米中启动报告基因的表达。因此，将PhytoExpr模型嵌入遗传算法，可以实现DNA序列在计算机中的进化，高效地探索DNA序列的无限变异空间，为植物合成生物学提供无限的元件。

来源：PNAS

发布日期:2024-06-18

全文链接:

[http://agri.nais.net.cn/file1/M00/10/42/Csgk0GZ1DnqAeDzgACM\\_XLqDD1s191.pdf](http://agri.nais.net.cn/file1/M00/10/42/Csgk0GZ1DnqAeDzgACM_XLqDD1s191.pdf)

### 3. 西北农林科技大学开发了创新的深度学习模型预测表型

**简介:** 2024年6月13日，西北农林科技大学植物生物信息学与系统生物学研究团队在Plant Communications上发表了题为“Dual-Extraction Modeling: A multimodal deep learning architecture for phenotypic prediction and functional gene mining of complex traits”的研究论文。该研究开发了一种创新的深度学习模型Dual-Extraction Modeling (DEM)。DEM模型能够从多种异质的组学数据集中提取代表性特征，进而预测与复杂性状相关的表型，并识别影响这些性状的功能基因。DEM模型架构采用了双重提取的策略，分别在高维和低维特征空间中进行建模。通过多头自注意力网络，DEM能够从个体组学特征矩阵和联合矩阵中提取全局注意力特征向量，进而优化模型权重并输出最终预测结果。这种方法不仅提升了性状表型预测的准确性，还通过事后解释策略提高了模型对性状相关功能基因的识别能力。在随后一系列全面的基准测试实验中，DEM模型在植物的数量性状和质量性状预测方面均展现出了优越的准确性、鲁棒性、泛化能力和灵活性。此外，DEM模型在人类疾病预测中同样表现优异。特别值得一提的是，DEM在预测影响多性状的多效性基因方面表现出了显著的可解释能力。在此基础上，研究团队开发了一款便于使用的软

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>

件，使得研究人员能够轻松运用DEM的各项功能进行多组学数据的分析与解读。该项研究成果不仅为复杂性状的研究提供了一个强有力的工具，还通过高效的特征提取、表型预测和事后解释方法，推动了对性状机制和疾病机理的理解。DEM的高性能和易用性将有望加速智能育种和精准医疗研究的进程。，有望在性状改良和疾病预测中发挥重要作用。

来源: Plant Communications

发布日期:2024-06-13

全文链接:

<http://agri.nais.net.cn/file1/M00/03/6F/Csgk0WZ1D1WAHpaMAIitroMXeJ0533.pdf>

#### 4. 农科院蔬菜花卉所利用转录组/代谢组解析番茄风味物质机制

**简介:** 2024年6月10日，Horticultural Plant Journal在线发表了中国农业科学院蔬菜花卉研究所崔霞团队题为“Integration of transcriptome and metabolome reveals regulatory mechanisms of volatile flavor formation during tomato fruit ripening”的研究论文。本研究以风味物质含量存在差异显著的野生醋栗番茄和栽培番茄为研究对象，采用固相微萃取-气相色谱-质谱联用技术，共鉴定出35种挥发性风味物质。通过对该35种挥发性风味物质的含量进行分析，野生番茄的总含量远高于栽培番茄，并且果实的挥发性风味物质含量在果实破色时期和红熟时期显著增加。通过转录组和代谢组联合分析，分别构建了影响重要风味物质2-异丁基噻唑的关键结构基因SITNH1以及己醛和顺式-2-庚烯醛的关键结构基因SILOX的转录调控网络。本研究鉴定到了bHLH和AP2/ERF家族的转录因子通过结合在SITNH1的启动子序列上进而激活其转录，影响2-异丁基噻唑的合成。此外，本研究还鉴定了HD-Zip家族中的一个转录因子对脂肪酸通路中的SILOX起调控作用，进而影响己醛和顺式-2-庚烯醛的合成。本研究对番茄果实在成熟过程中的转录组和挥发性风味物质的代谢组进行联合分析，挖掘调控番茄挥发性风味物质的转录因子并构建了调控网络，为研究番茄果实挥发性风味的形成提供了新的思路，对为进一步培育优良风味的番茄提供了理论基础。

来源: Horticultural Plant Journal

发布日期:2024-06-10

全文链接:

<http://agri.nais.net.cn/file1/M00/03/6F/Csgk0WZ1F0qACRv8AG2IqEc1JjA669.pdf>

#### 5. 山东农大解析醋栗番茄基因组

**简介:** 2024年6月4日，山东农业大学生命科学学院李传友团队在上发表题为“Chromosome-level genome assembly of Solanum pimpinellifolium”的研究论文，该研究在染色体水平上解析了醋栗番茄LA1589的基因组，为今后研究番茄驯化和基因组育种提供了宝贵的资源。番茄是全球重要的蔬菜作物之一，也是果实发育和抗性研究的经典模式作物。在番茄驯化过程中，持续的选择降低了遗传多样性，导致番茄“丢失”了许多优异的农艺性状。因此，在现代番茄育种中，野生番茄经常被用作重要的种质供体。醋栗番茄（Solanum pimpinellifolium）作为栽培番茄的野生祖先，具有抗病、抗逆等优异性状，使其在培育抗病、耐逆番茄新种质过程中具有巨大潜力。研究人员利用PacBio三代测序、Bionano光学图谱、Hi-C技术和二代测序策略，构建了醋栗番茄LA1589染色体水平的高质量基因组图谱。组装的基因组大小为833 Mb，contig N50长度为31 Mb，大约

更多资讯 尽在农业专业知识服务系统:<http://agri.nais.net.cn/>

98.8%的contig序列锚定到12条染色体，重复序列约占整个基因组的74.47%。BUSCO评估结果达到98.3%，同时，LAI得分达到14.49分，说明该基因组的组装质量较高。基因组注释共发掘41,449个蛋白编码基因，其中89.17%具有功能注释信息。

**来源:** Scientific Data

**发布日期:**2024-06-04

**全文链接:**

[http://agri.nais.net.cn/file1/M00/10/42/Csgk0GZ1F\\_GARI6UABnkJ-cHr0s462.pdf](http://agri.nais.net.cn/file1/M00/10/42/Csgk0GZ1F_GARI6UABnkJ-cHr0s462.pdf)