棉花多组学数据库发布，助力棉花遗传育种

作者：writer 来源：爱科学

本文原地址：https://www.iikx.com/news/progress/20353.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

棉花多组学数据库发布，助力棉花遗传育种。近日，华中农业大学信息学院、中国农科院棉花研究所以及新疆农垦科学院棉花所等多单位在《核酸研究》（Nucleic Acids Research）上联合发表研究成果。该研究通过搜集和整合棉花基因组、转录组、变异组、表观遗传、表型组和代谢组等6个组学的数据，构建出目前最为系统和全面的棉花多组学数据库，为棉花遗传育种研究提供了重要的数据资源和分析平台。

当前，我国棉花新品种选育主要以传统遗传育种方法为主。该方法虽然取得了大量重要成果，但其挖掘候选重要育种价值基因的精度和效率都较低。近年来已逐渐发展成熟的多组学技术可以为研究者提供更广泛、更多维度的信息来加速育种进程，为解决传统遗传育种方法中存在的问题、加速实现精准育种提供了新的途径。

为发挥多组学技术在育种中的价值，水稻、玉米、高粱等多个主要作物的多组学数据库平台已相继被建立，但目前仍缺少一个综合型的棉花多组学数据库平台。

为解决这一问题，研究者通过整合25个棉花基因组、76个组织样本的转录组、5个物种的表观遗传学、4180个样本的群体遗传变异数据、20个表型和768个代谢物含量等公共多组学数据，构建了目前最为系统和全面的棉花多组学数据库——CottonMD。

该数据库包含大量来自不同组学的信息，用户可以通过输入基因ID或物理位置信息来检索基因相关的多组学信息。以ATAF1基因为例，用户可以利用基因组模块获取4个同源基因的结构和功能信息，并通过转录组模块查询同源基因在不同组织、时期以及胁迫环境下的表达特征。这些功能为用户快速准确地理解基因的功能提供了快速方便的工具。

在该数据库中，研究者利用全基因组关联分析（GWAS）、表达数量性状位点定位（eQTL）、孟德尔随机化（SMR）和共定位分析等多组学关联分析方法，对不同组学的棉花数据进行关联，并将分析结果和工具整合到数据库中以方便用户查询、分析和利用。

然后，研究者以纤维伸长率位点FE1为例，系统介绍了利用CottonMD解析位点调控基因表达和表型的机制的方法。用户首先查询SMR和COLOC模块中该位点的SMR和共定位分析结果，确定出该位点中的候选基因Ghi_D04G09151；利用Variaiton模块进行单倍型分析，确定不同单倍型对应的基因表达和表型的效应；利用JBrowser浏览器查询该基因区域的遗传变异进和表观信号，推测遗传变异影响基因表达和表型的机制。

与现有的其他数据库相比，CottonMD是首个利用多种关联分析方法挖掘变异—基因表达—表型

之间关联信息并提供相应数据可视化查询的数据库。

此外，该数据库提供了最丰富的棉花多组学数据以及多种在线多组学分析以及种质资源管理工具。在CottonMD中，所有模块均支持25个已发表棉花基因组的基因ID进行搜索、浏览和数据，以服务全球范围内的相关研究工作。（来源：中国科学报王方）

相关论文信息：https://doi.org/10.1093/nar/gkac863

作者：Zhiquan Yang等 来源：《核酸研究》

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发