组学大数据“秒懂”功能基因的朋友圈

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/21413.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！

组学大数据“秒懂”功能基因的朋友圈。



玉米实验田。受访者供图

经典的遗传学手段克隆并解析了一批重要的功能基因。然而，在功能基因组提出二十几年之后的今天，水稻和玉米中克隆的功能基因仍不足其所有基因的10%，已克隆基因的新功能还不断被发现。

如何快速克隆功能基因、解析重要性状变异的分子机制，并全局解码重要作物的遗传变异的奥秘，仍然面临着巨大的挑战。

北京时间2022年12月30日，《自然—遗传》在线发表了华中农业大学作物遗传改良全国重点实验室、湖北洪山实验室教授李林课题组联合杨芳课题组、严建兵课题组的研究论文。该研究构建了玉米第一代多组学整合网络图谱，涉及到基因组、转录组、翻译组和蛋白互作组多个遗传层级的200万个网络关系，并利用机器学习方法成功预测了一批重要功能基因,鉴定出调控玉米开花期等

重要性状的分子调控通路。

中国农业大学教授田丰等当天在《植物学报》发表热点述评称，玉米多维组学整合网络的构建是玉米功能基因组学研究的重大进展,不仅为玉米重要性状新基因克隆、分子调控通路解析和玉米基因组进化分析提供了新工具,也为玉米基因组设计育种提供了重要基因资源和分子模块,为玉米智能育种奠定了重要基础。

5年克隆一个基因：功能基因解析进展缓慢

2008年，中国科学院院士、华中农业大学教授张启发在国际期刊《分子植物》(Molecular Plant)上发表了水稻功能基因组倡议(Rice 2020)，计划在2020年解析水稻所有基因的功能。

彼时，多种农作物的功能基因组解析工作全面开花，越来越多的科研人员投入到这一庞大和艰巨的研究当中。



李林在田间实验。受访者供图

李林读博期间的主要工作就是克隆和解析一个玉米籽粒油份主效功能基因。他做了非常完善的分子与遗传实验，也做了分子育种应用的评估。

然而，2010年，当他将花费了5年时间精细定位并克隆的基因投给了一本学术期刊，原本信心满满的他却收到评审人的意见：这个基因已经被别人克隆研究过了。

当时就非常沮丧，认为花5年图位克隆一个基因太低效了。于是，李林开始思考能不能快速全局地解析基因功能。

与此同时，农作物功能基因组的全面解析并没有像预想的那样快速推进。

2013年，李林在美国做博士后期间，产生了通过生物大数据方法来系统解析每一个基因功能的想法。

那时我就在做eQTL定位与共表达网络，并以此来解析玉米的全局基因的调控网络。2013年底，李林在美国《公共科学图书馆—遗传》(PLoS Genetics)上发表了玉米最早的通过群体RNA-seq进行eQTL研究调控关系的论文。

当他希望进一步深入研究，拟从多维组学的角度构建各个尺度以及介尺度水平的基因与基因调控网络，从而全面解析生物遗传奥秘时，却在与合作导师讨论时被否定了。可能觉得工作量太大，认为这是不可能完全的任务。李林回忆说。

不过，这个想法却在李林心中扎下了根。2016年，李林回国，成为华中农业大学的教师。严建兵让他组织大家讨论未来的重大课题。此时，杨芳团队已经开发了高通量酵母双杂交系统，并开始解析作物蛋白组学的网络结构。

我就把我的想法提出来与大家讨论，最终与杨芳老师一拍即合。得益于国内快速发展的科研实力、华中农业大学的平台支持，条件成熟了，三个团队共同推动了这个宏大项目的开展。李林说，他们正式拉开了在玉米基因组、转录组、翻译组，以及蛋白组学等水平构建玉米多维网络图谱的序幕。

彼时，只有不到10%的水稻基因的功能得以被解析，Rice 2020仍然任重道远。

锻造生物网络大数据之剑

生物种业是农业的基础与核心，而生物育种是生物种业的关键技术。严建兵告诉《中国科学报》，生物育种经历了1.0、2.0、3.0时代，正在大踏步向4.0时代的BT+IT驱动的智能育种进军。

无论生物育种处于哪个阶段，都离不开控制生物遗传变异的功能基因克隆与分子机制解析。严建兵说，经典的遗传学与分子生物学手段对重要性状的某个单一重要位点进行定位、克隆、分子互作实验，从而明确重要目标基因的上游调控基因、分子伴侣、以及下游的靶位点，进而构建功能基因的调控网络，最终解析一个基因控制重要性状变异的分子机制。

然而，以水稻和玉米为代表的农作物功能基因解析进展缓慢。快速克隆功能基因并解析重要性状变异的分子机制是迈步智能育种4.0时代的重要制约因素。

生物学研究业已进入大数据时代。基于生物大数据，从全局水平构建所有基因的上下游及分子伴侣网络已经成为可能，为我们全局解决尽可能多的基因功能，进而全面破解生物遗传变异的奥秘提供了前所未有的机会。严建兵说。

无论什么基因的研究，最终都要建立这个基因的分子网络模型。那么，为什么不一次性地把所有基因的上下游与分子伴侣关系都解析出来，从而就可以全局地了解尽可能多的基因的功能呢?李林解释说，生命体有几万个基因，要确定这些基因的功能，其实就是要确定这些基因与基因之间的调控关系。

生命体内部的基因，与人类社会中人非常相似。要确定一个人在人类社会中的功能或者作用，可以通过他的家庭亲属关系、朋友圈关系，以及工作圈关系等。同样，要了解一个基因的功能，也就是要了解它在不同遗传层次下与其它基因的关系。甚至，基于物以类聚，人以群分的逻辑，就可以推测出任何基因的功能。

基于此，他们在基因行使功能、传递遗传信息的不同层级(基因组、转录组、翻译组和蛋白互作组等)内部分别进行基因与基因关系的鉴定。



团队成员给实验玉米授粉。受访者供图

该研究对参考自交系B73全生育期不同组织/时期的样品进行多维组学大数据测定，获得了31个不同组织或发育时期的mRNA-Seq数据、21个不同组织或发育时期的circRNA-Seq、sRNA-Seq数据和21个组织的Ribo-Seq数据。

杨芳介绍，他们使用高通量酵母系统构建了玉米蛋白互作网络，获得了36万多个蛋白—蛋白互作对，高置信度的互作有56243个。整合已有的基因组水平ChIA-PET网络与该研究产生的转录组水平共表达网络、翻译组水平共翻译网络和蛋白互作网络，构建了玉米第一代多组学整合网络图谱，涉及到200万个互作关系。

这是首次如此全面地在一个物种中构建了基因组、转录组、翻译组以及蛋白组的网络大数据图谱，就好像锻造出了生物网络大数据之剑，为全面系统解析玉米遗传变异机制提供了基础。李林说。

革新经典遗传学研究的范式

基于成功构建的玉米多维网络大数据图谱，该研究在全基因组水平探究了重复基因在网络中的功能分化，揭示出玉米两个远古亚基因组从转录组到蛋白互作组表现出渐进式的功能分化。

他们还重构了玉米已经克隆的株型功能基因以及籽粒发育相关功能基因的分子网络。田丰等指出，截至目前,玉米中共有63个调控籽粒发育的基因被克隆,其中62个基因位于该团队此次发布的整合图谱中。他们成功预测并证实了1个未知功能的PPR蛋白能够影响玉米籽粒的皱缩。上述研究结果充分证明了该整合网络图谱具有强大的预测基因功能的能力。

进而，他们重点关注了玉米重要农艺性状——开花期。为了保障大数据预测的准确性，他们与华中农业大学理学院教授陈洪团队紧密合作，不断开发前沿的人工智能算法，快速进行性状解析，为系统解析基因功能以及性状变异的遗传机制提供了新的手段，在一定程度上革新了经典遗传学研究的范式。

论文评审人认为，该研究描述了一项庞大的实验工作，用以鉴定玉米基因组编码的大分子之间的复杂分子关系。特别是论文中机器学习对于网络大数据的挖掘应用，是一种创新的功能基因分子网络图谱的解读方式。

通过对第一代玉米整合多组学网络大数据进行挖掘，他们预测了2651个候选的开花期基因，并根据它们是否控制相同的性状而划分为8个子网络途径。

为了验证预测结果的准确性，2020年开始，他们在海南、湖北、山东和河北四地对预测结果进行分子实验与大田实验验证，确定了20个预测基因与开花期性状相关，并对其分子机制进行了初步阐释。

除前人已经报道的途径外，他们还鉴定了在玉米中全新的分子网络途径，加深了对玉米开花期的理解，为玉米开花期的智能设计育种提供了理论基础与基因资源。

田丰等人指出，基因组、表型组、转录组、蛋白组和代谢组等多维组学大数据是利用机器学习等人工智能技术精准挖掘关键基因和分子模块进行基因组智能设计育种的基础。

论文评审人认为，该项研究用功能验证的方式支撑了基于网络大数据的基因功能的人工智能预测;构建了玉米不同生育期不同组织的综合分子网络图谱，对于整个玉米研究来说是一个重要的资源。

李林告诉《中国科学报》，这篇论文已成功解析了玉米多维组学内部的网络关系，下一步还要搞

清楚多维组学之间的调控关系。此外，这一研究范式还可应用到水稻、小麦等其他作物的功能基因组研究上。(来源：中国科学报 李晨)

相关论文信息：http://doi.org/10.1038/s41588-022-01262-1

作者：李林等 来源：《自然—遗传》

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