

· 热点评述 ·

多组学整合网络：一把精准解码玉米功能基因组的钥匙

郭丽, 王雪涵, 田丰*

中国农业大学, 植物生理学与生物化学国家重点实验室, 国家玉米改良中心, 北京 100193

摘要 高通量组学技术的快速发展使生命科学进入大数据时代。科学家们从基因组、转录组、蛋白质组和代谢组等多组学数据中剥茧抽丝, 逐步揭示生物体内复杂而巧妙的调控网络。近日, 华中农业大学李林课题组联合杨芳课题组和严建兵课题组构建了玉米(*Zea mays*)首个多组学整合网络。该网络包括3万个玉米基因在三维基因组水平、转录水平、翻译水平和蛋白质互作水平的调控关系, 由280万个网络连接组成, 构成1 412个调控模块。利用该整合网络, 研究团队预测并证实了5个调控玉米分蘖、侧生器官发育和籽粒皱缩的新基因。进一步结合机器学习方法, 他们预测出2 651个影响玉米开花期的候选基因, 鉴定到8条可能参与玉米开花期的调控通路, 并利用基因编辑技术和EMS突变体证实了20个候选基因的生物学功能。此外, 通过对整合调控网络的进化分析, 他们发现玉米两套亚基因组在转录组、翻译组和蛋白互作组水平上存在渐进式的功能分化。这套集合多组学数据构建的整合网络图谱是玉米功能基因组学研究的重大进展, 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, 是解锁玉米功能基因组学的一把新钥匙。

关键词 玉米, 多组学, 整合网络图谱, 机器学习

郭丽, 王雪涵, 田丰 (2023). 多组学整合网络: 一把精准解码玉米功能基因组的钥匙. 植物学报 58, 1–5.

早在1万年前人类开始陆续驯化出多种作物。现代玉米(*Zea mays ssp. mays*)起源于中美洲墨西哥地区巴尔萨斯河流域, 由其野生祖先种大刍草(*Z. mays ssp. parviglumis*)驯化而来(Matsuoka et al., 2002)。自1492年哥伦布发现美洲新大陆并将玉米种子带到欧洲后, 玉米便开始向全世界传播, 现已成为我国乃至世界第一大粮食作物。2009年玉米B73自交系参考基因组序列的发布使玉米基因组学研究进入新时代。随着测序技术的更新迭代和测序成本的降低, 多个玉米自交系的基因组序列相继发表。2021年, 美国Hufford研究团队公布了玉米巢式关联定位群体(nested association mapping, NAM) 26个亲本的基因组序列(Hufford et al., 2021); 2022年, 华中农业大学严建兵团队公布了玉蜀黍属“超级泛基因组”图谱(Gui et al., 2022), 这些基因组序列的释放极大地促进了玉米重要性状的遗传解析、功能基因的挖掘及玉米进化机制的阐释。

转录组、代谢组和蛋白质组等多维组学技术的广

泛应用, 为玉米功能基因组学研究带来了前所未有的发展契机。Tu等(2020)利用转录组测序技术全面解析了玉米叶片中表达的转录因子结合位点, 构建了玉米叶片转录调控网络。Xu等(2021)对玉米未成熟雌穗进行单细胞转录组测序, 构建了玉米花序发育组织转录调控图谱。Wen等(2018)通过代谢组测序技术分析了513份玉米自交系的61条基础代谢途径, 鉴定出760个参与不同代谢通路的候选基因。Zhu等(2021)利用开发的翻译组测序技术, 成功构建了玉米多组织的翻译组图谱。Walley等(2016)通过转录组和蛋白质组联合分析, 建立了一个大规模的与玉米发育过程相关的基因表达谱; 并发现基于基因表达量构建的共表达网络与基于蛋白质表达量构建的调控网络存在较大差异, 仅有15%的调控网络保守。尽管单一组学的研究为我们提供了重要的数据库信息, 但仍有一定的局限性。如何利用多组学数据库精准预测基因功能及鉴定重要性状的分子调控通路仍是该领域研究人员面临的巨大挑战。

收稿日期: 2022-12-05; 接受日期: 2022-12-13

基金项目: 国家自然科学基金(No.32025027)

* 通讯作者。E-mail: ft55@cau.edu.cn

从基因组到转录组、翻译组和蛋白质组,及从基因到表型的调控为层层加密环环相扣。应用多组学的整合分析能够极大地提高对基因功能和调控通路的预测能力。近日,华中农业大学李林课题组联合杨芳课题组和严建兵课题组发表了玉米首个多组学数据整合网络图谱(Han et al., 2022)。该图谱整合了三维基因组、转录组、翻译组以及蛋白质互作组4个层次的数据:(1)以Peng等(2019)利用ChIA-PET (chromatin interaction analysis using paired end tag sequencing)技术鉴定到约14 788个玉米近端和远端顺式调控元件的三维互作数据为基础,绘制了高分辨率的三维基因组图谱;(2)通过整合31份玉米不同组织和时期的转录组数据,构建了玉米基因共表达图谱;(3)对21份玉米不同组织和时期的翻译组数据进行整合,绘制了翻译组共翻译图谱;(4)利用高通量酵母双杂交技术(Yang et al., 2018),获得包含15 476个玉米蛋白的360 000多对互作数据,其中高可信度的蛋白互作关系达56 243对。最终,3万个玉米基因在多组学水平形成280万个调控连接,构成1 412个调控模块。该研究一个重要的创新点是通过改造酵母双杂交系统的载体质粒(Yang et al., 2018),将诱饵蛋白所在载体(pGBKT7)插入ATTB位点,将捕获蛋白所在载体(pGADT7)插入ATTP位点,分别在2个载体上插入来自噬菌体PhiC31的表达盒。利用PhiC31介导的ATTB位点和ATTP位点发生重组,使互作的2个蛋白所在质粒整合为1个大质粒,进一步通过扩增、酶切后加接头及二代测序技术获得高通量的蛋白互作网络,这是迄今为止最完善的玉米蛋白互作数据库。大规模分子实验数据与大数据挖掘分析的有效结合,使该数据库具备了信息全面和功能强大的优越性(Han et al., 2022)。

他们测试了该多组学整合网络图谱预测新基因的效率(Han et al., 2022)。分蘖减少和顶端优势增强是玉米驯化中关键的形态转变,这一驯化过程主要由*tb1* (*teosinte branched 1*)、*gt1* (*grassy tillers 1*)和*tru1* (*tassels replace upper ears 1*)等基因调控(Doebley et al., 1997; Gallavotti et al., 2004; Whipple et al., 2011; Dong et al., 2017)。该研究团队利用多维网络图谱鉴定到2个与*tb1*、*gt1*和*tru1*共表达的ALOG转录因子基因*ZmALOG1*和*ZmALOG2*;并进一步利用CRISPR基因编辑技术证明*ZmALOG1*和*ZmALOG2*

敲除突变体表现出多分蘖表型。*CUC* (*CUP-SHAPED COTYLEDON*)基因家族是拟南芥(*Arabidopsis thaliana*)中已克隆的影响器官发育边界的基因,但在玉米中还未见报道。该研究团队分析了玉米*CUC*基因(*ZmCUC3*、*ZmNAM1*和*ZmNAM2*)所在的调控网络,发现这些基因与调控玉米侧生器官形成和发育的*Tsh1* (Gallavotti et al., 2004)、*Ba1* (Gälweiler et al., 1998)和*ZmPIN1* (Xiao et al., 2022)存在显著互作,并进一步利用基因编辑技术证实*ZmNAM1*和*ZmNAM2*参与玉米侧生器官发育的调控;此外,他们对目前研究较为完善的玉米籽粒发育基因调控网络进行了分析。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个基因位于该整合图谱中。该研究团队成功预测并证实了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。

除预测单个基因的生物学功能外,他们还测试了多组学整合网络在揭示重要农艺性状分子调控通路上的预测能力(Han et al., 2022)。开花期是影响玉米产量和决定玉米适应性的关键性状,众多研究者利用突变体、QTL克隆和比较基因组学方法鉴定了一系列调控玉米开花期的重要基因(Liang et al., 2021),但玉米开花期的分子调控网络尚不完善。李林课题组与杨芳课题组和严建兵课题组合作,基于多组学整合网络,利用机器学习的方法,预测出2 651个玉米开花期基因,并进一步鉴定到8条可能的玉米开花期调控通路,分别为光信号途径、生物钟途径、光周期途径、自主途径、赤霉素途径、花序组织特定表达途径、成花转变途径及其它途径。利用基因编辑技术和甲基磺酸乙酯(ethyl methanesulfonate, EMS)诱变突变体,对20个位于不同调控通路的基因进行生物学功能验证,其中18个基因的功能之前在玉米中未有报道。需要指出的是,他们鉴定出一条新的开花调控通路,该通路基因在拟南芥中的同源基因参与春化基因*FLC*的表观调控,说明拟南芥春化基因在玉米中可能发生了功能分化,具体分子机制有待进一步研究。

多组学整合网络对玉米基因组进化特征也提供了重要线索。Schnable等(2011)通过比较高粱(*Sorghum bicolor*)和玉米基因组,发现玉米由于基因组加倍存在2个亚基因组,分别为*maize1*和*maize2*,其中*maize1*为更具优势的亚基因组。该研究团队从多

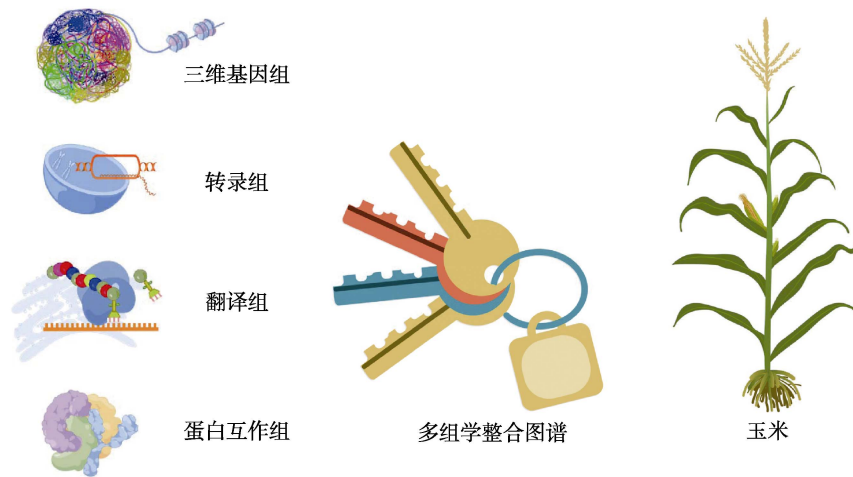


图1 多组学整合网络图谱是精准解锁玉米功能基因组学的新钥匙(本图由Figdraw绘制, 玉米植株图改自Chen等(2021))

Figure 1 Multi-omics integrative network map is a new key to accurately decode the maize functional genomics (this figure is drawn by figdraw and the plant picture is adapted from Chen et al., 2021)

组学水平比较了2个亚基因组的网络图谱, 发现2个亚基因组在转录组、翻译组和蛋白组的调控网络上存在渐进式的功能分化, 在共表达水平上2个玉米亚基因组之间不存在显著差异; 在共翻译水平上虽有差异, 但仍不显著; 在蛋白质互作水平上2个玉米亚基因组则存在显著的功能分化。基因组复制是植物基因组进化的重要特征之一。通过分析不同类型的复制基因在多组学水平的分化程度, 他们发现从共表达、共翻译到蛋白互作组, 玉米复制基因间调控网络的分化程度越来越高(Han et al., 2022)。

随着作物基础研究的持续深入和生物育种技术的不断进步, 作物育种进入4.0时代, 即BT+IT驱动的智能设计育种。智能设计育种的本质是大数据、机器学习等人工智能技术与多组学技术、基因编辑、合成生物学等生物技术的深度融合, 以实现作物新品种的定向、智能、高效设计和培育(王向峰和才卓, 2019; 汪海等, 2022)。其中, 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。该研究整合玉米三维基因组、转录组、翻译组和蛋白质互作数据, 构建了首个玉米多维组学整合网络图谱, 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。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(图1)。

参考文献

- 汪海, 赖锦盛, 王海洋, 李新海 (2022). 作物智能设计育种——自然变异的智能组合和人工变异的智能创制. 中国农业科技导报 **24**(6), 1–8.
- 王向峰, 才卓 (2019). 中国种业科技创新的智能时代——“玉米育种4.0”. 玉米科学 **27**, 1–9.
- Chen Q, Li W, Tan L, Tian F (2021). Harnessing knowledge from maize and rice domestication for new crop breeding. *Mol Plant* **14**, 9–26.
- Doebley J, Stec A, Hubbard L (1997). The evolution of apical dominance in maize. *Nature* **386**, 485–488.
- Dong ZB, Li W, Unger-Wallace E, Yang JL, Vollbrecht E, Chuck G (2017). Ideal crop plant architecture is mediated by tassels replace upper ears1, a BTB/POZ ankyrin repeat gene directly targeted by TEOSINTE BRANCHED1. *Proc Natl Acad Sci USA* **114**, E8656–E8664.
- Gallavotti A, Zhao Q, Koyuzuka J, Meeley RB, Ritter MK, Doebley JF, Pè ME, Schmidt RJ (2004). The role of barren stalk1 in the architecture of maize. *Nature* **432**, 630–635.
- Gälweiler L, Guan CH, Müller A, Wisman E, Mendgen K, Yephremov A, Palme K (1998). Regulation of polar auxin

- transport by *AtPIN1* in *Arabidopsis* vascular tissue. *Science* **282**, 2226–2230.
- Gui ST, Wei WJ, Jiang CL, Luo JY, Chen L, Wu SS, Li WQ, Wang YB, Li SY, Yang N, Li Q, Fernie AR, Yan JB** (2022). A pan-*Zea* genome map for enhancing maize improvement. *Genome Biol* **23**, 178.
- Han LQ, Zhong WS, Qian J, Jin ML, Tian P, Zhu WC, Zhang HW, Sun YH, Feng JW, Liu XG, Chen G, Farid B, Li RN, Xiong ZM, Tian ZH, Li J, Luo Z, Du DX, Chen SJ, Jin QX, Li JX, Li Z, Liang Y, Jin XM, Peng Y, Zheng C, Ye XN, Yin YJ, Chen H, Li WF, Chen LL, Li Q, Yan JB, Yang F, Li L** (2022). A multi-omics integrative network map of maize. *Nat Genet* Doi: <https://www.nature.com/articles/s41588-022-01262-1>
- Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou SJ, Liu JN, Ricci WA, Guo TT, Olson A, Qiu YJ, Della Coletta R, Tittes S, Hudson AI, Marand AP, Wei S, Lu ZY, Wang B, Tello-Ruiz MK, Piri RD, Wang N, Kim DW, Zeng YB, O'Connor CH, Li XR, Gilbert AM, Baggs E, Krasileva KV, Portwood JL II, Cannon EKS, Andorf CM, Manchanda N, Snodgrass SJ, Hufnagel DE, Jiang QH, Pedersen S, Syring ML, Kudrna DA, Llaca V, Fengler K, Schmitz RJ, Ross-Ibarra J, Yu JM, Gent JI, Hirsch CN, Ware D, Dawe RK** (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662.
- Liang YM, Liu HJ, Yan JB, Tian F** (2021). Natural variation in crops: realized understanding, continuing promise. *Annu Rev Plant Biol* **72**, 357–385.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J** (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci USA* **99**, 6080–6084.
- Peng Y, Xiong D, Zhao L, Ouyang WZ, Wang SQ, Sun J, Zhang Q, Guan PP, Xie L, Li WQ, Li GL, Yan JB, Li XW** (2019). Chromatin interaction maps reveal genetic regulation for quantitative traits in maize. *Nat Commun* **10**, 2632.
- Schnable JC, Springer NM, Freeling M** (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci USA* **108**, 4069–4074.
- Tu XY, Mejía-Guerra MK, Valdes Franco JA, Tzeng D, Chu PY, Shen W, Wei YY, Dai XR, Li PH, Buckler ES, Zhong SL** (2020). Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat Commun* **11**, 5089.
- Walley JW, Sartor RC, Shen ZX, Schmitz RJ, Wu KJ, Urich MA, Nery JR, Smith LG, Schnable JC, Ecker JR, Briggs SP** (2016). Integration of omic networks in a developmental atlas of maize. *Science* **353**, 814–818.
- Wen WW, Jin M, Li K, Liu HJ, Xiao YJ, Zhao MC, Alseikh S, Li WQ, de Abreu e Lima F, Brotman Y, Willmitzer L, Fernie AR, Yan JB** (2018). An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant J* **93**, 1116–1128.
- Whipple CJ, Kebrom TH, Weber AL, Yang F, Hall D, Meeley R, Schmidt R, Doebley J, Brutnell TP, Jackson DP** (2011). *grassy tillers1* promotes apical dominance in maize and responds to shade signals in the grasses. *Proc Natl Acad Sci USA* **108**, E506–E512.
- Xiao YG, Guo JY, Dong ZB, Richardson A, Patterson E, Mangrum S, Bybee S, Bertolini E, Bartlett M, Chuck G, Eveland AL, Scanlon MJ, Whipple C** (2022). Boundary domain genes were recruited to suppress bract growth and promote branching in maize. *Sci Adv* **8**, m6835.
- Xu XS, Crow M, Rice BR, Li F, Harris B, Liu L, Demesa-Arevalo E, Lu ZF, Wang LY, Fox N, Wang XF, Drenkow J, Luo AD, Char SN, Yang B, Sylvester AW, Gingeras TR, Schmitz RJ, Ware D, Lipka AE, Gillis J, Jackson D** (2021). Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Dev Cell* **56**, 557–568.
- Yang F, Lei YY, Zhou ML, Yao QL, Han YC, Wu X, Zhong WS, Zhu CH, Xu WZ, Tao R, Chen X, Lin D, Rahman K, Tyagi R, Habib Z, Xiao SB, Wang D, Yu Y, Chen HC, Fu ZF, Cao G** (2018). Development and application of a recombination-based library versus library high-throughput yeast two-hybrid (RLL-Y2H) screening system. *Nucleic Acids Res* **46**, e17.
- Zhu WC, Xu J, Chen SJ, Chen J, Liang Y, Zhang CJ, Li Q, Lai JS, Li L** (2021). Large-scale translome profiling annotates the functional genome and reveals the key role of genic 3' untranslated regions in translomic variation in plants. *Plant Commun* **2**, 100181.

Multi-omics Integrative Network Map, a Key to Accurately Decoding the Maize Functional Genomics

Li Guo, Xuehan Wang, Feng Tian*

State Key Laboratory of Plant Physiology and Biochemistry, National Maize Improvement Center of China, China Agricultural University, Beijing 100193, China

Abstract Life science is entering into the era of big data due to the rapid development of high-throughput omics technology. Multi-omics data such as genome, transcriptome, proteome, metabolome have greatly facilitated dissecting the complex and sophisticated regulatory networks of organisms. Recently, a collaborative team led by Lin Li, Fang Yang and Jianbing Yan from Huazhong Agricultural University constructed the first multi-omics integrative network map of maize. This map comprises over 30 000 genes and 2.8 million network edges at the levels of genome, transcriptome, translome, and proteome, finally forming 1 412 regulatory modules. Using the integrative network map, the research team successfully predicted and confirmed five new functional genes regulating the development of tiller, lateral organ, and kernel in maize. Based on the integrative map and machine learning, the research team identified 2 651 maize flowering time genes that are enriched in eight candidate subnetworks. The biological functions of 20 flowering candidate genes were further validated using CRISPR/Cas9 gene editing technology and EMS mutants. Furthermore, evolutionary analysis of the integrative network map showed that the two subgenomes of maize had undergone a progressive functional differentiation from the levels of co-expression, co-translation to interactome. The construction of the multi-omics integrative network map represents an important breakthrough in maize functional genomics, which provides a new tool for cloning new genes, identifying novel molecular regulatory pathways, and revealing maize genome evolutionary features. This multi-omics integrative network map is a new key to decode maize functional genomics.

Key words maize, multi-omics, integrative network map, machine learning

Guo L, Wang XH, Tian F (2023). Multi-omics integrative network map, a key to accurately decoding the maize functional genomics. *Chin Bull Bot* **58**, 1–5.

* Author for correspondence. E-mail: ft55@cau.edu.cn

(责任编辑: 孙冬花)