基于DAP-Seq方法鉴定玉米胚乳特异表达转录因子O2在全基因组水平上的结合位点

宁丽华¹, 王建晖², 时丕彪³, 伍楚善⁴, 董志诚², 赵涵^{1*}

1. 江苏省农业生物学重点实验室, 江苏省农业科学院种质资源与农业生物技术研究所, 南京 210014;

2. 广州大学生命科学学院广州高等教育中心, 广州 510006;

3. 江苏省盐城市新洋农业试验站, 盐城 224049;

4. 盐城师范学院海洋与生物工程学院, 盐城 224007

* 联系人, E-mail: zhaohan@jaas.ac.cn

2019-05-29 收稿, 2019-07-02 修回, 2019-07-04 接受, 2019-08-21 网络版发表

国家重点研究发展计划(2016YFD0101002)、国家自然科学基金(31701453)和江苏省重点研究开发项目(BE2017365)资助

摘要 Opaque2(O2)是一个调控玉米胚乳发育的重要转录因子. 为了揭示O2的转录调控网络, 利用DAP-Seq技术挖掘其在全基因组范围上结合位点共检测到11385个位点. 通过结合已报道RNA-Seq数据, 本研究DAP-Seq共检测到317个O2直接结合并调控的靶基因, 其中包括97个已报道ChIP-Seq检测靶基因以及220个DAP-Seq特异检测的靶基因. 而基序(motif)富集分析显示, DAP-Seq检测O2结合317靶标基因与220个特异结合靶标基因所富集的基序均与已报道O2结合motif相同. 另外, GO富集分析显示, 与ChIP-Seq一致的O2靶基因主要参与zein蛋白合成及氨基酸代谢途径, 而特异检测的O2靶基因, 除参与上述途径外, 还主要参与糖代谢途径以及多个胁迫响应相关途径. 本研究利用DAP-Seq技术进一步揭示了O2在胚乳发育过程发挥调控作用, 该结果是对前人研究结果的验证和补充.

关键词 Opaque2, 转录因子, DAP-Seq, 转录因子结合位点, 调控网络

Opaque2(O2)是玉米胚乳特异表达的转录因子(transcription factor, TF), O2蛋白含有碱性氨基酸基序/亮氨酸拉链, 属于bZIP家族^[1,2]. O2通过结合“ACGT”基序来激活玉米储藏蛋白如 α -醇溶蛋白的编码基因在胚乳中特异表达, 当O2突变后导致玉米籽粒的醇溶蛋白大幅降低同时提高籽粒中赖氨酸的含量^[3,4]. 另外, O2突变还造成胚乳呈现软质不透明的特性, 从而增加玉米的对病虫害及机械损伤的易感性, 说明O2参与玉米的抗性过程^[4]. 还有研究表明, O2突变后玉米籽粒的淀粉合成途径受到影响, 糖代谢水平下降以及氨基酸代谢途径发生改变等, 这使得突变体的玉米胚乳、幼苗以及雄穗发育过程均不同于野生型^[5,6]. 为了研究O2是

否直接参与调节这些代谢途径以及了解O2复杂的基因调控网络, 人们利用ChIP-Seq结合RNA-Seq方法检测一些O2直接结合并调控的靶标基因, 包括玉米各种zein亚家族基因, 与病原体防御相关的**b32**基因, 参与赖氨酸代谢的**LKR/SDH**基因, 参与籽粒碳氮代谢相关的**cyPPDK**基因以及与胚乳发育相关的**NKD2**基因等^[7,8].

虽然前人已检测到很多O2靶基因, 并且这些基因功能涉及范围广泛, 但这并不能解释O2突变后所产生所有表型改变. 为了进一步解析O2的功能, 并了解O2在全基因组水平上结合位点情况, 我们对O2转录因子进行DNA亲和纯化测序(DNA affinity purification

引用格式: 宁丽华, 王建晖, 时丕彪, 等. 基于DAP-Seq方法鉴定玉米胚乳特异表达转录因子O2在全基因组水平上的结合位点. 科学通报, 2019, 64: 2537–2548

Ning L H, Wang J H, Shi P B, et al. Mapping genome-wide binding sites of endosperm specific expression transcription factor O2 using DAP-Seq (in Chinese). Chin Sci Bull, 2019, 64: 2537–2548, doi: [10.1360/N972019-00334](https://doi.org/10.1360/N972019-00334)

sequencing, DAP-Seq)分析. DAP-Seq是目前新开发的一种体外检测TF-DNA互作的分析方法,该方法在不依赖于所研究TF抗体的前提下,使用体外表达的TF蛋白对裸露的全基因组DNA片段进行孵育,以确定其结合位置和基序^[9]. DAP-Seq检测方法具有成本低、获取TF结合信息快的特点. 另外该方法所使用的全基因组DNA携带了影响TF结合的特性(基序上下有序列和甲基化修饰等),这使得其在检测TF结合基序和结合位点方面比其他体外检测方法更准确. 目前DAP-Seq在人类、拟南芥以及玉米相关转录因子研究上得到应用^[9,10].

本研究通过将DAP-Seq与前人报道RNA-Seq进行整合分析,发现有97个O2直接靶标基因与前人结果一致,同时还有220个DAP-Seq特异检测到O2直接靶标基因. GO分析发现O2直接调控基因主要参与编码籽粒储藏蛋白及参与氨基酸合成与代谢相关酶,同时O2还直接调控与碳骨架分配相关基因的转录. 特异检测的O2靶标基因主要参与糖代谢途径. 本研究进一步揭示了O2在胚乳发育过程发挥调控作用,该结果是对前人研究结果的验证和补充.

1 材料与方法

1.1 材料

授粉后15 d的B73玉米胚乳.

1.2 方法

根据Bartlett 等人^[11]报道,对DAP-Seq操作方法进行一定修改,具体方法如下:

(i) DNA片段化及建库. 利用CTAB法提取授粉后15 d玉米籽粒DNA,使用Bioruptor Plus仪器打断至片段大小为250 bp. 利用磁珠(BECKMAN COULTER AM-Pure XP A63881)筛选目标片段然后利用建库试剂盒(NEXTflex-Rapid DNA-Seq Kit)完成End repair以及加A过程.

(ii) 蛋白表达. 所使用构建表达载体的引物F: 5'-CAGAGCGATAACGCGATGGAGCACGTCATCT-CAATGGA-3'和R: 5'-AGCCCGAATTCGTTTTTAATACATGTCCATGTGTATGGCCCCA-3'将O2的开放阅读框全场重组到pFN9K Halotag[®]T7SP6 Flexi[®]载体. 然后,按照麦胚蛋白表达系统(Promega TNT SP6 wheat Germ Master Mix)使用说明,对O2重组质粒进行表达.

(iii) DNA与蛋白结合以及Western-blotting检测. 将表达蛋白溶液与Halotag-beads混合, 25°C, 1500 r/min震荡1 h, 然后加入DNA文库再震荡1 h. 震荡结束后清洗Halotag-beads, 向离心管中加入30 μ L EB溶解. 然后再98°C 1000 r/min震荡10 min. 加热结束后立刻将离心管放在磁力架上,待溶液澄清后2 min, 将上清转移至新离心管,并保存于-20°C,用于后续实验. 然后向含有Halotag-beads离心管中加入40 μ L ddH₂O, 再加入10 μ L 5 $\times$ SDS-PAGE 上样缓冲溶液,然后将离心管放在100°C加热10 min. 取20 μ L液体进行Western blotting,检查Halotag-beads上是否有O2蛋白表达.

(iv) 测序. 利用PCR扩增将Index加入到与O2蛋白结合的DNA片段上,然后利用磁珠进一步筛选除加上Index的DNA片段,再利用KAPA文库定量试剂盒检测DNA文库质量合格. 最后由上海瀚宇生物科技有限公司完成测序工作.

(v) 测序数据分析. 利用Bowtie2软件^[12]将DAP-Seq测序读长比对到玉米B73基因组(B73 RefGen_v3)^[13],然后利用MACS 2.0^[14]软件检测O2相对于对照GFP所富集的峰. O2结合峰利用IGV进行可视化分析^[15]. 利用MEME-ChIP软件分析O2靶标基因富集motif. 利用agriGO database v1.2在线分析GO分析^[16].

2 结果

2.1 基于DAP-Seq识别O2结合位点

为了检测O2在玉米全基因组上的直接靶基因,我们提取授粉15 d后B73玉米胚乳DNA进行DAP-Seq分析,试验设置两个重复,将全基因组DNA(gDNA)打断使DNA集中区域在250 bp(图1(a)). 然后根据试剂盒说明制备DAP-Seq gDNA文库. 利用麦胚表达系统将O2全长ORF序列进行表达,以GFP达作为阴性对照;将gDNA文库加入与O2蛋白孵育后,获得O2结合DNA片段,并用PCR引物扩增以引入索引;利用Illumina HiSeq-2500平台对回收的基因组片段进行测序,产生配对读长. Western-blotting实验验证Halo-tag结合的蛋白表达情况,发现O2转录因子和GFP均能正常表达(图1(b)).

O2结合DNA样品产生2822万个读长,其中有98.51%读长被定位到玉米参考基因组;而GFP结合DNA样品平均产生2245万个读长,有91.81%读长被定位到玉米参考基因组. 将测序数据进行前处理后,然后

图 1 (网络版彩色)DNA片段化、O2蛋白表达以及O2结合峰分布图. (a) DNA片段化检测; (b) 蛋白与DNA结合后west-blotting检测; (c) DAP-Seq检测O2结合峰分布, 基因区域从起始转录位点上游5 kb到终止转录位点下游5 kb; (d) DAP-Seq检测O2结合峰分布, 基因区域从起始转录位点上游1 kb到终止转录位点下游1 kb

Figure 1 (Color online) The detection of DNA fragmentation and O2 protein expression and the analysis of O2 binding peak distribution: (a) DNA fragment detection; (b) detection after O2 protein using west-blotting; (c) distribution of O2 peaks in the maize genome based on localization of peak summits. The gene region is from 5 kb upstream of the transcription start site to 5 kb downstream of the transcription ending site; (d) distribution of O2 peaks in the maize genome based on localization of peak summits. The gene region is from 1 kb upstream of the transcription start site to 1 kb downstream of the transcription ending site

利用MACS2软件去识别O2结合峰的序列. 本研究将两个重复均检测到的峰作为可重现峰用于后续分析, 设 $\log_{10}(q\text{-value})=5$ 为检测阈值, 共检测到11385个可重现的O2结合区域. DAP-Seq检测的峰分布在玉米所有染色体上, 其中有34.06% 峰位于基因起始转录位点上游5 kb之外的区域, 即基因间区(图1(c)). 另外有32.25%位于基因起始位点上游5 kb内, 13.02%位于基因转录终止位点下游5 kb内, 5'-非翻译区分布最少(1.66%). 为检测O2结合靶标基因, 将检测的重复峰落在基因的转录起始位点上下1 kb区域的注释基因, 定义为O2靶标基因,

根据这个标准我们共检测到5357个O2结合靶标基因.

2.2 DAP-Seq鉴定O2直接结合调控靶标基因

为了鉴定DAP-Seq检测的靶标基因是否有O2直接结合调控, 将DAP-Seq与前人报道的B73背景下O2突变体的RNA-Seq进行结合分析^[8], 发现DAP-Seq共检测到317个O2直接结合并结合靶标基因, 其中260个被O2激活, 57个被O2抑制(图2(a), 表S1). 与前人ChIP-Seq检测数据进行比较发现, 有97个DAP-Seq检测基因与前人结果一致, 与Zhan等人^[8]和Li等人^[7] 研究结果的重叠率分

别为50.54%和35.90%。其中这些重叠基因包括15个zein基因,以及早前已报道的*cyPPDK1/2*, *LKR/SDH*, *NKD2*和*bZIP17*等基因(表S1, 图3)。而其他220个基因则为DAP-Seq特异检测到O2直接结合调控的靶位点,这其中包括172个O2激活靶基因和48个O2抑制基因。上述结果说明, DAP-Seq能一定程度上解析O2结合位点,并对ChIP-Seq结果进行补充。

2.3 DAP-Seq鉴定O2结合位点分布及结合cis元件

为鉴定O2结合靶位点的cis元件,抽取317个O2直接结合并调控基因所对应峰的上下游200 bp基因组序列,并利用MEME-ChIP软件进行motif富集分析。结果显示富集最显著的motif序列是“YGCCACGTCAB”(E-value=1.3×e⁻³⁹),该基序中“CCACGTCA”与前人报道O2结合cis元件序列一致^[8]。为了进一步明确O2在基因组上结合的cis元件,同样抽取220个DAP-Seq特异检测到的O2结合峰的基因组序列,并进行MEME-ChIP分析,发现DAP-Seq特异结合的位点富集到“SYGC-CACGTCTC”的基序(E-value=7.6×e⁻⁸⁴)(图4(b))。该基序同样含有前人已报道的“CACGTCA”序列,并且从图5中可以看出该motif是在所分析峰的中心区域显著富集($P=1.4\times e^{-39}$)(图4(c))。另外,以100 bp为bin分析了DAP-Seq检测到的O2直接调控基因区域内峰分布的数目,结果发现O2主要集中结合转录起始位上下游-250~+200 bp基因组区域(图4(a))。这一结果与前人研究发

O2结合位点在转录起始位点上游300 bp区域内一致^[8,17]。由此可见, DAP-Seq检测鉴定到O2结合cis元件和结合位置与ChIP-Seq鉴定结果一致。

2.4 DAP-Seq鉴定O2结合靶基因GO分析

为了更深入地了解O2基因调控网络,对DAP-Seq检测的靶基因进行GO(Gene ontology)富集分析($q<0.05$)。结果显示, 260个O2直接激活基因主要参与四类相关过程(图5): 一是与有机酸代谢有关,包括有机酸代谢过程(GO:0006082); 有机酸分解代谢过程(GO:0016054); 代谢过程(GO:0043436); 羧酸代谢过程(GO:0019752); 单羧酸代谢过程(GO:0032787)以及羧酸分解代谢过程(GO:0046395)。这一类代谢途径涉及O2激活基因有67个,其中包括在籽粒的碳分配过程中发挥重要作用的*cyPPDK1/2*。二是与含氮复合物代谢相关的生物学过程,这一过程富集的GO条目最多,共富集到12个相关过程GO条目,包括有机氮化合物代谢过程(GO:1901564)、有机氮化合物分解代谢过程(GO:1901565)、胺代谢过程(GO:0009308)、 α -氨基酸代谢过程(GO:1901605)、 α -氨基酸分解代谢过程(GO:1901606)、 α -氨基酸生物合成过程(GO:1901607)、细胞氨基酸代谢过程(GO:0006520)、细胞氨基酸生物合成过程(GO:0008652)、细胞氨基酸分解代谢过程(GO:0009063)、丝氨酸家族氨基酸代谢过程(GO:0009069)、丝氨酸家族氨基酸生物合成(GO:0009070)以及谷氨酰

图2 (网络版彩色)O2直接结合调控基因的Venn图。(a) O2直接结合调控基因与前人DEGs的Venn图^[8];(b) DAP-Seq与ChIP-Seq检测O2直接结合调控基因结合基因的Venn图^[7,8]

Figure 2 (Color online) Venn diagram of O2 directly binding regulatory gene. (a) Venn diagram of numbers of O2-modulated genes and O2-bound genes detected by RNA-Seq^[8] and DAP-Seq. The direct binding of O2 regulatory gene to the Venn diagram with the DEGs in Zhan et al.'s study^[8]; (b) the venn diagram of numbers of O2 binding gene detected by DAP-Seq and ChIP-Seq^[7,8]

图 3 (网络版彩色)DAP-Seq检测O2靶基因峰图
Figure 3 (Color online) Peak of O2 target genes detected by DAP-Seq

胺家族氨基酸代谢过程(GO:0009064)。参与这一类代谢途径的基因有93个, 其中包括与胚乳发育相关的 *NKD2* (GRMZM5G884137)、*LKD/SDH* 基因 (GRMZM2G181362)、谷氨酰胺合成酶编码基因 (GRMZM2G024104)、脯氨酸氧化酶编码基因 (GRMZM2G053720和GRMZM2G117956)、丙氨酸-乙醛酸转氨酶编码基因 (GRMZM2G124353) 和 2-氧代异戊酸脱氢酶编码基因 (GRMZM2G037614) 等。第三类是参与糖代谢过程, 该类生物学过程所富集GO条目包括糖基化合物生物合成过程 (GO:1901659)、糖苷生物合成过程 (GO:0016138) 以及糖苷代谢过程 (GO:0016137)。参加这一类代谢途径的基因有磷酸海藻糖磷酸酶编码基因 (GRMZM2G140078和GRMZM2G118462)、 α -葡聚糖磷酸化酶编码基因 (GRMZM2G085577) 和肌醇半乳糖苷-蔗糖半乳糖基转移酶编码基

因 (GRMZM2G050177) 等。最后一类与抗逆相关过程, 包括两个GO条目分别是响应刺激 (GO:0050896) 和响应karrikin过程 (GO:0080167)。O2转录激活靶标基因中有157基因可能与玉米胁迫响应有关, 包括参与盐和氧化响应的 *LGL* (GRMZM2G312877)、与乙醇代谢相关的 *GBF1* (GRMZM2G011932), 以及参与玉米抗旱相关过程的 *NF-YB16* (GRMZM2G384528) 和 *NAC111* (GRMZM2G127379) 基因。另外, 该部分基因最显著富集的分子功能的GO条目是营养库活性 (nutrient reservoir activity) (GO:0045735)。此外, 还富集的氧化还原酶活性 (GO:0016491) 和转氨酶活性 (GO:0008483) 等分子功能GO条目。

对DAP-Seq检测到57个O2直接抑制靶基因进行GO分析, 共检测到7个显著富集GO条目均为生物学过程GO条目, 包括氧化还原过程 (GO:0055114)、化学物

图4 (网络版彩色) O₂结合位点基因组区域分布及220个DAP-Seq特异检测靶标基因所富集的motif. (a) O₂结合位点基因组区域分布; (b) 220个DAP-Seq特异检测靶标基因所富集的motif; (c) 显著富集基序在O₂靶标峰周围的位置分布

Figure 4 (Color online) The analysis of distribution of O₂ binding site and enriched motif of genes specifically detected from DAP-Seq O₂ binding cis-elements. (a) O₂ binding motif identified by MEME-ChIP in 1-kb flanking sequences around the genic peak summits and the density plot of this motif around the summits of peaks. (b) The significantly enriched motif sequence of O₂ binding sites specifically detected by DAP-Seq. (c) The motif's positional distributions around the summits of peaks associated with direct O₂ targets

响应(GO:0042221)、响应盐胁迫响应(GO:0009651)以及渗透胁迫响应(GO:0006970)等GO条目. 参与氧化还原过程的O₂抑制靶标基因有11个, 包括*Orphan154* (GRMZM2G300125)、编码APGase小亚基的*bt2* (GRMZM2G068506)、参与糖酵解途径的三磷酸甘油醛脱氢酶编码基因(*GAPDH*, GRMZM2G071630)以及过氧化物酶编码基因(GRMZM2G136158).

总的来说, O₂直接调控基因主要编码籽粒储藏蛋白zein、参与氨基酸合成和代谢一些相关酶, 以及参与调控碳骨架分配过程. 另外O₂还能直接调节糖代谢相关途径以及参与多个胁迫响应相关过程.

2.5 DAP-Seq特异检测O₂结合靶基因功能分析

为了进一步了解DAP-Seq检测到与ChIP-Seq重叠的O₂靶标基因与DAP-Seq特异检测O₂靶标基因功能的差异, 对这两部分基因分别进行GO分析(图6), 发现与ChIP-Seq重叠的88个O₂激活基因主要参与两类相关过程(图6(a)). 一是与有机酸代谢有关过程, 包括有机酸代谢过程(GO:0006082)、oxoacid代谢过程(GO:0043436)以及羧酸代谢过程(GO:0019752). 这一类生物学过程中涉及包括*cyPPDK1/2*的26个基因. 二是与含氮化合物代谢相关过程, 包括胺代谢过程(GO:0009308)以及多个与氨基酸相关的代谢过程: 细胞氨基酸分解代谢(GO:0009063)、 α -氨基酸代谢过程(GO:1901605)、 α -

氨基酸分解代谢过程(GO:1901606)、细胞氨基酸代谢过程(GO:0006520)和 α -氨基酸生物合成过程(GO:1901607). 这一类生物学过程参与的基因有*LKD/SDH*基因(GRMZM2G181362)、谷氨酰胺合成酶编码基因(GRMZM2G 024104)和脯氨酸氧化酶编码基因(GRMZM2G 053720和GRMZM2G117956)等. 另外最显著富集的分子功能GO条目是与zein基因相关的营养库活性条GO目(nutrient reservoir activity)(GO:0045735), 还有氧化还原酶活性(GO:0016491)和因子结合活性(GO:0048037)等. 在本研究中, 与ChIP-Seq重叠的9个O₂抑制基因没有检测到显著富集的GO条目. GO富集条目说明DAP-Seq检测到与ChIP-Seq重叠的O₂靶标基因主要参与调节籽粒储藏蛋白zein、参与氨基酸合成与代谢途径以及参与调控碳骨架向碳和氮化合物之间的分配过程.

对DAP-Seq特异检测到172个O₂激活基因进行GO富集分析(图6(b)), 发现在该部分基因富集的生物过程GO条目中有3个与ChIP-Seq检测基因富集GO条目相同, 分别是有机酸代谢途径(GO:0006082)、有机酸代谢途径(GO:0043436)以及 α -氨基酸代谢过程(GO:1901605). 其中参与 α -氨基酸代谢过程的基因有12个, 包括丙氨酸-乙醛酸氨基转移酶(GRMZM2G124353)、3-甲基-2-氧丁酸脱氢酶参与缬氨酸代谢过程(GRMZM2G037614)、甲基丁烯酰-CoA 羧化酶

图5 (网络版彩色)DAP-Seq检测260个O₂结合激活靶基因GO分析
Figure 5 (Color online) GO analysis of 260 target genes activated by O₂ binding

(GRMZM2G702490). 除上述3个生物过程外, DAP-Seq特异检测的O₂靶标基因还参与籽粒中糖相关代谢过程, 比如单糖代谢途径(GO:0005996)、己糖代谢途径(GO:0019318)、糖苷生物合成途径(GO:0016138)、糖苷生物代谢途径(GO:0016137)以及糖基化合物生物合成途径(GO:1901659, 0.039). 有19个参与这部分代谢相关的途径的O₂靶标基因, 包含有磷酸海藻糖磷酸酶编码基因(GRMZM2G140078和GRMZM2G118462)、 α -葡聚糖磷酸化酶编码基因(GRMZM2G085577)和肌醇半乳糖苷-蔗糖半乳糖基转移酶编码基因(GRMZM2G050177)等. 另外, DAP-Seq特异检测到的靶标基因有109个O₂激活基因参与胁迫响应相关生物学过程, 其中包括*NF-YB16*和*ZmNAC111*. 另外有22个基因还与次生代谢物合成(GO:0044550)和代谢GO:0019748)有关.

DAP-Seq特异检测到48个O₂结合并抑制的基因主要参与多个胁迫响应相关过程, 如化学物响应(GO:0042221)、响应盐胁迫响应(GO:0009651)以及渗透胁迫响应(GO:0006970). 另外有10个基因(包括*bt2*)参与还参与氧化还原过程(GO:0055114).

从上述结果可以看出, DAP-Seq特异检测到O₂直接靶标基因除了参与氨基酸合成与代谢途径以及调控碳骨架分配过程外, 有更多基因直接参与糖代谢相关途径以及多个胁迫响应相关过程.

3 讨论

O₂突变后会对玉米胚乳发育造成多种影响, 有些影响的机制已经明确, 但还有很多机制并不清楚^[8]. 为了进一步解析O₂的功能并全面了解O₂在基因组水平

图6 (网络版彩色)DAP-Seq检测O₂结合激活基因的GO分析。(a) 与ChIP-Seq重叠O₂结合激活靶基因富集生物学过程GO条目;(b) DAP-Seq特异检测到O₂结合激活靶基因富集生物学过程GO条目

Figure 6 (Color online) The GO analysis for O₂ activated targets detected by DAP-Seq. (a) Associated biological process GO of genes overlapped with ChIP-Seq; (b) associated biological process GO of genes specifically detected by DAP-Seq

上的结合位点情况,我们对O₂转录因子进行DAP-Seq,并将DAP-Seq结果与已发表RNA-Seq结果进行整合分析,发现DAP-Seq检测到317个O₂直接结合并调控的靶标基因。DAP-Seq检测直接靶标与在W22背景中ChIP-Seq检测结果^[7]重叠率为35.90%,与在B73背景中ChIP-Seq检测结果^[8]重叠率为50.54%,该结果一方面说明DAP-Seq能一定的程度解析O₂在全基因组范围上的结合位点,另一方面也说明DAP-Seq与ChIP-Seq之间存在一些差异,究其原因一方面可能是检测全基因组结合位点所用材料的遗传背景存在差异^[18],另一方面可能由DAP-Seq与ChIP-Seq两种方法的差异造成。DAP-Seq是体外检测转录因子结合位点的方法,而ChIP-Seq是体内检测转录因子结合位点的方法。结合前人分析^[7,8],本研究发现有一部分O₂靶标基因只能由ChIP-Seq特异检测而不能被DAP-Seq检测,说明该部分基因被结合时

不仅需要O₂转录因子还需要体内与O₂相应的辅助因子或者互作蛋白^[8,19]。GBF1能够激活*Adh1*的转录^[20], Li等人^[7]利用ChIP-Seq方法检测到*GBF1*和乙醇脱氢酶编码基因(*Adh1*)均为O₂的靶标基因,但是本研究只检测到*GBF1*基因而没有检测到*Adh1*,这说明O₂可能需要与GBF1协同作用来调控*Adh1*的表达。

*Zein*基因是O₂主要调控的靶标基因^[7,8],本研究发现有20个*zein*基因被O₂结合,这几乎涵盖所有*zein*亚家族,其中15个(包括10个19 kD和2个22 kD α -*zein*, 16 kD β -*zein*, 27和15 kD γ -*zein*, 以及18 kD δ -*zein*)与ChIP-Seq检测结果一致^[7,8]。不同的是本研究检测到O₂只结合了2个22 kD *zein*编码基因,明显少于ChIP-Seq检测数目,说明22 kD *zein*亚家族基因的激活比其他*zein*亚家族基因更依赖于植株体内O₂同源二倍体结构以及与其互作的因子,这或许能一定程度解释O₂与PBF相互作用后

比单独O2对22 kD *zein*基因具有更强的转录激活活性^[19]。

在非*zein*基因O2靶基因中有多个与ChIP-Seq重叠并且功能明确的基因,例如*LKR/SDH*基因编码赖氨酸-酮戊二酸还原酶/糖苷脱氢酶参与赖氨酸降解^[21],两个编码脯氨酸氧化酶基因(*GRMZM2G053720*和*GRMZM2G117956*)均为O2直接靶标基因^[7]。另外还检测到*cyPPDK1*和*cyPPDK2*,编码胚乳特异性胞质同工丙酮酸正磷酸二激酶异构体,该酶参与碳的分配过程^[22],但是很遗憾本研究并没有检测与淀粉合成相关的*SSIII*基因。*NKD2*基因编码转录因子是许多胚乳中糊粉层、胚乳细胞分化等胚乳发育过程以及储藏物质合成的核心调控因子^[23]。另外,本研究检测到*bZIP17*能够被O2直接结合并激活的靶标基因,该基因编码转录因子与O2相互作用后能增强对下游靶基因的转录活性^[8]。

DAP-Seq还特异检测到很多重要功能的非*zein*靶标基因(表S1,图3),其中*GRMZM2G124353*编码一种丙氨酸-乙醛酸氨基转移酶,参与甘氨酸和丙氨酸之间的转换代谢过程,在遗传定位分析中发现该基因与玉米氨基酸含量显著关联^[24]。*GRMZM2G037614*编码的3-甲基-2-氧丁酸脱氢酶参与缬氨酸代谢过程,通过GWAS分析鉴定到该基因参与氨基酸代谢网络^[25]。*GRMZM2G702490*编码甲基丁烯酰-CoA 羧化酶,该酶参与亮氨酸降解过程。前人研究发现O2不仅能够直接调控籽粒蛋白的合成而且还能直接结合并激活淀粉合成相关基因^[26],本研究检测到一个参与淀粉合成过程重要基因*bt2*,该基因编码玉米种子的腺苷酸葡萄糖焦磷酸酶小亚基(AGPase),AGPase能够将催化Glc-1-P和ATP产生ADP-葡萄糖和无机焦磷酸盐,是淀粉合成过程中的限速酶^[27],这一结果进一步说明O2能够直接参与调控淀粉合成相关途径。前人研究发现O2激活*SSIII*基因表达,而本研究发现*bt2*转录被O2抑制,说明O2可能以一种复杂的调控模式对淀粉合成途径进行调控,这一推测需要后期进一步研究。另外,*GRMZM5G856653*基因参与多个代谢途径,该基因编码的己糖激酶既调控植物体内储存糖和游离糖的利用率,还调控糖酵解和氧化戊糖磷酸途径的代谢速率^[28]。*GRMZM2G118462*是海藻糖-磷酸合酶编码基因,该酶可能与玉米产量的形成有关^[29]。另外,本研究还检测到多个与胁迫

响应相关基因,如*ZmNAC111*编码NAC家族转录因子,该基因5'-非翻译区发生突变后会导致玉米幼苗的抗旱性减弱,说明该基因在玉米抗旱过程发挥重要作用^[30]。总体来说,DAP-Seq特异检测到靶标基因说明O2除了直接调节籽粒储藏蛋白*zein*、参与氨基酸合成与代谢途径以及参与调控碳骨架的分配过程,还能直接调节糖代谢相关途径并参与胁迫响应过程。

早期研究显示O2是一个转录激活因子^[7,21],近期有研究发现O2也是一个转录抑制因子^[8],而本研究中显示O2具有转录抑制功能。将DAP-Seq与RNA-Seq结合分析发现有57个基因被O2直接结合并抑制,其中包含9个ChIP-Seq检测到靶标基因和48个DAP-Seq特异检测的直接抑制靶点(表S1)。本研究发现O2抑制*bt2*的转录,而从前人研究整个胚乳发育过程来看当O2基因开始表达时*bt2*基因表达量呈现下降趋势^[31],这说明O2可能一定程度上对这些基因表达产生抑制作用。另外还有研究发现将*o2*基因导入背景材料后,该导入系的AGPase活性高于其背景材料的活性,说明O2具有抑制AGPase活性功能。但是本研究DAP-Seq仅是结合15 DAP的RNA-Seq数据进行分析,并不能确定O2是否在胚乳发育后期与这些基因结合;另外前人研究表明O2突变后会导致蛋白质组的再平衡,即*zein*蛋白减少而非*zein*蛋白水平上升^[32],所以在*o2*突变体中*bt2*等基因表达水平上调是否由O2突变导致转录抑制功能丧失而成还有待进一步分析。总体来说,本研究进一步证实O2具有转录抑制功能。

4 结论

综上所述,本研究利用DAP-Seq检测到11385个位点O2结合DNA位点,结合前人RNA-Seq数据共发现317个O2直接结合并调控靶基因,这些靶基因功能与ChIP-Seq检测基因功能一致。而motif富集分析显示DAP-Seq检测O2结合317靶标基因与220个特异结合靶标基因所富集的基序均与已报道O2结合motif相同。另外,GO富集分析显示,与ChIP-Seq一致的O2靶基因主要参与*zein*蛋白合成及氨基酸代谢途径,而特异检测的O2靶标基因,除参与上述途径外,还主要参与糖代谢途径以及多个胁迫响应相关途径。可见,本研究利用DAP-Seq技术进一步揭示了O2在胚乳发育过程发挥调控作用,该结果是对前人研究结果的验证和补充。

参考文献

- 1 Schmidt R J, Burr F A, Aukerman M J, et al. Maize regulatory gene opaque-2 encodes a protein with a "leucine-zipper" motif that binds to zein DNA. *Proc Natl Acad Sci USA*, 1990, 87: 46–50
- 2 Mertz E T, Bates L S, Nelson O E. Mutant gene that changes protein composition and increases lysine content of maize endosperm. *Science*, 1964, 145: 279–280
- 3 Schmidt R J, Ketudat M, Aukerman M J, et al. Opaque-2 is a transcriptional activator that recognizes a specific target site in 22-kD zein genes. *Plant Cell*, 1992, 4: 689–700
- 4 Bass H W, Webster C, Obrian G R, et al. A maize ribosome-inactivating protein is controlled by the transcriptional activator Opaque-2. *Plant Cell*, 1992, 4: 225–234
- 5 Mehta S L, Srivastava K N, Mali P C, et al. Changes in the nucleic acid and protein fractions in Opaque-2 maize kernels during development. *Phytochemistry*, 1972, 11: 937–942
- 6 Giroux M J, Boyer C, Feix G, et al. Coordinated transcriptional regulation of storage product genes in the maize endosperm. *Plant Physiol*, 1994, 106: 713–722
- 7 Li C, Qiao Z, Qi W, et al. Genome-wide characterization of *cis*-acting DNA targets reveals the transcriptional regulatory framework of *Opaque2* in maize. *Plant Cell*, 2015, 27: 532–545
- 8 Zhan J, Li G, Ryu C H, et al. Opaque-2 regulates a complex gene network associated with cell differentiation and storage functions of maize endosperm. *Plant Cell*, 2018, 30: 2425–2446
- 9 O'Malley R C, Huang S S, Song L, et al. Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell*, 2016, 165: 1280–1292
- 10 Galli M, Khakhar A, Lu Z, et al. The DNA binding landscape of the maize auxin response factor family. *Nat Commun*, 2018, 9: 4526
- 11 Bartlett A, O'Malley R C, Huang S S C, et al. Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nat Protoc*, 2017, 12: 1659–1672
- 12 Langmead B, Salzberg S L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 2012, 9: 357–359
- 13 Schnable P S, Ware D, Fulton R S, et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science*, 2009, 326: 1112–1115
- 14 Zhang Y, Liu T, Meyer C A, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, 2008, 9: R137
- 15 Thorvaldsdóttir H, Robinson J T, Mesirov J P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings BioInf*, 2013, 14: 178–192
- 16 Du Z, Zhou X, Ling Y, et al. agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res*, 2010, 38: W64–W70
- 17 Wu Y, Messing J. Rapid divergence of prolamin gene promoters of maize after gene amplification and dispersal. *Genetics*, 2012, 192: 507–519
- 18 Bernard L, Ciceri P, Viotti A. Molecular analysis of wild-type and mutant alleles at the Opaque-2 regulatory locus of maize reveals different mutations and types of O2 products. *Plant Mol Biol*, 1994, 24: 949–959
- 19 Vicente-Carbajosa J, Moose S P, Parsons R L, et al. A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proc Natl Acad Sci USA*, 1997, 94: 7685–7690
- 20 de Vetten N C, Ferl R J. Characterization of a maize G-box binding factor that is induced by hypoxia. *Plant J*, 1995, 7: 589–601
- 21 Kemper E L, Neto G C, Papes F, et al. The role of Opaque2 in the control of lysine-degrading activities in developing maize endosperm. *Plant Cell*, 1999, 11: 1981–1993
- 22 Sheen J. Molecular mechanisms underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. *Plant Cell*, 1991, 3: 225–245
- 23 Gontarek B C, Neelakandan A K, Wu H, et al. NKD transcription factors are central regulators of maize endosperm development. *Plant Cell*, 2016, 28: 2916–2936
- 24 Wen W, Jin M, Li K, et al. An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize. *Plant J*, 2018, 93: 1116–1128
- 25 Deng M, Li D, Luo J, et al. The genetic architecture of amino acids dissection by association and linkage analysis in maize. *Plant Biotechnol J*, 2017, 15: 1250–1263
- 26 Zhang Z, Zheng X, Yang J, et al. Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. *Proc Natl Acad Sci USA*, 2016, 113: 10842–10847
- 27 Huang B, Hennen-Bierwagen T A, Myers A M. Functions of multiple genes encoding ADP-glucose pyrophosphorylase subunits in maize endosperm, embryo, and leaf. *Plant Physiol*, 2014, 164: 596–611
- 28 Zhou M L, Zhang Q, Sun Z M, et al. Trehalose metabolism-related genes in maize. *J Plant Growth Regul*, 2014, 33: 256–271
- 29 Paes G P, Viana J M S, Silva F F E, et al. Linkage disequilibrium, SNP frequency change due to selection, and association mapping in popcorn

- chromosome regions containing QTLs for quality traits. [Genet Mol Biol](#), 2016, 39: 97–110
- 30 Mao H, Wang H, Liu S, et al. A transposable element in a NAC gene is associated with drought tolerance in maize seedlings. [Nat Commun](#), 2015, 6: 8326
- 31 Chen J, Zeng B, Zhang M, et al. Dynamic transcriptome landscape of maize embryo and endosperm development. [Plant Physiol](#), 2014, 166: 252–264
- 32 Wu Y, Messing J. Proteome balancing of the maize seed for higher nutritional value. [Front Plant Sci](#), 2014, 5: 252

补充材料

表S1 DAP-Seq检测的317个O2直接结合并调控靶基因

本文以上补充材料见网络版csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “基于DAP-Seq方法鉴定玉米胚乳特异表达转录因子O2在全基因组水平上的结合位点”

Mapping genome-wide binding sites of endosperm specific expression transcription factor O2 using DAP-Seq

Lihua Ning¹, Jianhui Wang², Pibiao Shi¹, Chushan Wu³, Zhicheng Dong² & Han Zhao^{1*}

¹ Institute of Crop Germplasm and Biotechnology, Jiangsu Provincial Key Laboratory of Agrobiolgy, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China;

² Guangzhou Higher Education Mega Center, School of Life Sciences, Guangzhou University, Guangzhou 510006, China;

³ Xinyang Agricultural Experiment Station of Yancheng City, Yancheng 224049, China;

⁴ Ocean and Biological Engineering College, Yancheng Teachers University, Yancheng 224007, China

* Corresponding author, E-mail: zhaohan@jaas.ac.cn

Opaque2 (O2) is an important transcription factor regulating maize endosperm development, and mutation of the wild type O2 gene reduces the amount of zein and increases the lysine contents, thereby, leading to development of opaque and starchy endoplasm commonly referred to as *o2* mutant. The opaque and starchy nature of the *o2* mutant endosperm renders the kernels resistance to pests, diseases and mechanical damage caused during transportation besides having high nutritional quality. O2 regulates the genes involved in various biological functions or processes, including the synthesis and metabolism of proteins, carbohydrates and lipids. Previous studies have identified O2 target genes regulating different biological functions. However, the detailed information about the genome-wide analysis of O2 targets genes is insufficient and needs further investigation. Taking advantage of DNA affinity purification sequencing (DAP-Seq) over Chromatin Immunoprecipitation sequencing (ChIP-Seq) being a high throughput and easy approach to capturing whole genome transcription factor and DNA interaction profile, the gene regulatory network of O2 was investigated by coupling with mRNA abundance analyses. Using DAP-Seq, we detected a total of 11385 O2 binding sites throughout the genome. Genome-wide comparison of O2 TF showed that DAP-Seq peaks captured significant fractions of ChIP-Seq peaks (35.9%–50.54%), indicating the potential of DAP-seq to preferentially captures *in vitro* transcription factor binding sites. By integrating previously RNA-Seq analysis of wild type (B73) vs *o2* mutant endosperm, we identified 317 directly bound target genes of O2 in this study, correlated with 97 target genes detected by ChIP-Seq and 220 additional target genes specifically detected by DAP-Seq. Among the direct target genes detected by DAP-Seq, 260 target genes were directly bound and activated by O2, including 20 zein gene, *cyPPDK1/2*, *NF-YB16* and *ZmNAC111*; and another 57 target genes were directly bound and repressed by O2, such as *bt2*, *GAPDH*, *Orphan 154* etc. To further gain a comprehensive insight into the functions of the O2-regulated gene network, we performed Gene Ontology (GO) terms significantly enriched among O2 direct target genes using agriGO tool. The results indicated that O2 mainly regulates genes involved in endosperm storage product synthesis and accumulation, directly activates genes encoding storage proteins and enzymes involved in amino acid biosynthesis or metabolism in maize endosperm. Moreover, O2 also regulated genes involved in carbon partitioning between carbohydrates and proteins in maize endosperm. However, 220 O2 binding target gene specifically detected by DAP-Seq were also mainly involved in the glucose metabolism pathway and response to various stresses. To identify cis-regulatory sequences mediating O2 binding, we performed a de novo motif discovery on the O2 binding sites detected by DAP-Seq, using the MEME-ChIP program. We found that the most significant motif containing “CCACGTCA” sequence, the reported O2 cis motif, was enriched in the 400 bp region centered at the summits of peaks for 317 direct O2 target genes. In addition, the common cis-motif also enriched within 400 bp region centered at the summits of peaks for 220 specific binding target genes. The motif was shown to be centrally enriched in the analyzed genomic regions. Taken together, our results provide insight into the complexity of the O2 regulatory network, which validate and supplement the previous research findings.

Opaque2, transcription factor, DAP-Seq, TF binding site, regulatory network

doi: [10.1360/N972019-00334](https://doi.org/10.1360/N972019-00334)