

宗琳皓,吴 彤,徐 德,等. 水稻 PEPCK 基因家族的生物信息学和表达分析[J]. 江苏农业科学,2025,53(7):22–28.

doi:10.15889/j.issn.1002-1302.2025.07.003

水稻 PEPCK 基因家族的生物信息学和表达分析

宗琳皓,吴 彤,徐 德,高志萍
(南京师范大学生命科学学院,江苏南京 210023)

摘要:磷酸烯醇式丙酮酸羧激酶(PEPCK)普遍存在于开花植物中,是将草酰乙酸(OAA)转变为磷酸烯醇式丙酮酸(PEP)的关键酶,参与植物体多条代谢途径。PEPCK 由 PEPCK 基因编码,本研究通过全基因组数据,鉴定编码水稻 PEPCK 蛋白的基因家族成员,并围绕水稻 PEPCK 基因家族的理化性质、亚细胞定位、保守基序、顺式作用元件、组织特异性表达、胁迫下表达以及进化关系等方面进行生物信息学分析。结果显示,水稻全基因组共有 4 个 PEPCK 基因,其中一个成员(*Oa03g0255500*)含有 2 条转录本,并且家族成员分布在水稻 3 号、4 号以及 10 号染色体上,具有相似的保守基序;顺式作用元件分析以及基因表达分析显示,该基因家族在水稻光合作用碳反应、种子萌发、防御与应激、抵御干旱胁迫等生理功能方面具有重要作用;利用不同碳代谢类型植物的 PEPCK 基因家族蛋白进行系统发育树构建,并对禾本科植物 PEPCK 基因家族进行共线性分析,结果显示,植物 PEPCK 基因在不同碳代谢类型植物中较为保守,且其进化在禾本科不同物种间经历相似途径。本研究为明确水稻 PEPCK 基因家族各成员的功能提供了参考依据。

关键词:水稻;磷酸烯醇式丙酮酸羧激酶;生物信息学;表达特征;共线性分析

中图分类号:S188;S511.01 **文献标志码:**A **文章编号:**1002-1302(2025)07-0022-07

植物磷酸烯醇式丙酮酸羧激酶(PEPCK)可逆催化草酰乙酸(OAA)脱羧,且该酶是催化糖异生途径的关键限速酶之一,同时也是连接糖异生、糖酵解及有机酸代谢等多种代谢途径的重要桥梁。前期研究表明,在贮油植物的种子萌发子叶还未具备光合作用能力前,PEPCK 可催化脂肪氧化产物 OAA 脱羧形成磷酸烯醇式丙酮酸(PEP),进而为植物生长提供有机碳源^[1]。在 C₄ 和景天酸代谢途径(CAM)植物中,PEPCK 还可以催化由苹果酸或天冬氨酸衍生的 OAA 脱羧,为卡尔文循环提供二氧化碳^[2-3]。在水稻中转入玉米 PEPCK 基因后,水稻脱落酸(ABA)信号通路增强、苹果酸代谢受到调节、气孔关闭得到促进,表明 PEPCK 在响应干旱胁迫中具有多重作用^[4]。除此之外,PEPCK 还参与了植物氮素与氨基酸的代谢、植物细胞 pH 值的调节等^[5-6]。

水稻是禾本科稻属的一年生植物,在全球粮食

生产体系中占据着举足轻重的地位。近年来,在水稻中与生长发育、逆境胁迫相关的基因被相继研究,PEPCK 早期被认为参与 C₄ 植物光合作用而只存在于 C₄ 植物中,近年来发现水稻中也存在 PEPCK^[7],但目前关于水稻 PEPCK 基因家族的研究还不够全面。本研究通过生物信息学方法,从水稻全基因组中鉴定出 PEPCK 基因家族成员,对其理化性质进行分析并进行亚细胞定位预测,通过注释文件得到 PEPCK 基因家族的染色体位置、基因结构、保守基序等信息,并进行启动子顺式作用元件的预测,同时针对水稻 PEPCK 基因家族的表达模式和对非生物胁迫的响应进行分析,期望揭示水稻 PEPCK 基因家族的特征及功能。此外,本研究选择不同碳代谢类型的代表植物,鉴定不同植物中 PEPCK 基因家族成员,并进行系统发育树的构建,同时通过不同物种间的共线性分析,得到水稻 PEPCK 基因家族的进化关系。以期后续深入开展水稻 PEPCK 基因家族在水稻分子育种和逆境胁迫中功能的研究提供理论参考。

1 材料与方法

1.1 试验时间及地点

试验于 2023 年 10 月至 2024 年 3 月,在南京师范大学生命科学学院江苏省生物多样性与生物技

收稿日期:2024-03-14

基金项目:国家自然科学基金(编号:32101639);江苏省优势学科项目(PAPD)。

作者简介:宗琳皓(2003—),男,江苏南通人,主要从事生物信息学研究。E-mail:licht_hao@163.com。

通信作者:高志萍,博士,副教授,主要从事植物学研究。E-mail:08295@njnu.edu.cn。

术重点实验室进行。

1.2 水稻 PEPCK 基因家族的鉴定

从 Ensembl 数据库 (<http://plants.ensembl.org/>) 获得水稻、高粱、小米、玉米、小麦、豌豆、拟南芥、番茄以及菠萝的基因组数据以及注释文件。通过 Pfam 数据库^[8] (<http://pfam.xfam.org/>) 检索 PEPCK 保守结构域的隐马尔科夫模型(数据库编号 PF01293),通过 HMMER 软件查找水稻全基因组所编码蛋白中含有 PEPCK 结构域的序列,以作为水稻 PEPCK 基因家族的候选成员。为获得更为准确的结果,利用拟南芥中已被充分研究的 PEPCK 蛋白序列信息进行 BLAST 序列比对。将 2 种方法得到的水稻 PEPCK 基因家族候选成员合并,去除重复值后利用 NCBI CDD (<https://www.ncbi.nlm.nih.gov/cdd>) 进一步验证搜索到的蛋白序列结构域,最终筛选出水稻 PEPCK 基因家族成员。其他植物中 PEPCK 基因家族成员参考与水稻相同的方法进行鉴定。

1.3 水稻 PEPCK 基因家族的理化性质分析与亚细胞定位预测

利用 ExPASy^[9] (<https://web.expasy.org/>) 对水稻 PEPCK 基因家族的氨基酸数目、等电点、分子量、平均疏水性、不稳定指数等参数进行分析。通过 WoLF - PSORT^[10] (<https://wolfpsort.hgc.jp/>) 对水稻 PEPCK 基因家族成员进行亚细胞定位预测。

1.4 水稻 PEPCK 基因家族的蛋白保守基序与基因结构分析

使用 MEME^[11] (<https://meme-suite.org/>) 对水稻 PEPCK 基因家族成员进行保守基序分析,并利用 TBtools 软件^[12] 进行水稻 PEPCK 基因家族 motif 的可视化。根据水稻基因组注释信息数据,使用 TBtools 软件构建水稻 PEPCK 基因家族的基因结构图。

1.5 水稻 PEPCK 基因家族的染色体定位

依据水稻基因数据以及注释文件,通过 TBtools 软件进行水稻 PEPCK 基因家族染色体定位的分析与可视化。

1.6 水稻 PEPCK 基因家族的启动子顺势作用元件预测分析

利用 TBtools 软件提取水稻 PEPCK 基因家族成员上游 1 500 bp 序列,通过 PlantCARE 在线网站 (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 进行顺式作用元件预测,在 R 语言中完成数据处理、分析及可视化。

1.7 水稻 PEPCK 基因家族的表达分析

从 RGAP 数据库 (<https://rice.uga.edu/expression.shtml>) 下载水稻 PEPCK 基因家族成员在不同组织以及不同生长发育阶段的转录组数据,在 R 语言中对基因表达数据进行分析及可视化,结果以聚类热图的形式呈现。利用水稻表达数据库 RED^[13] (<https://ngdc.cnecb.ac.cn/red/index>) 获得水稻 PEPCK 基因家族在不同胁迫下的表达谱数据,水稻 PEPCK 基因家族各成员在不同处理下的基因表达量以折线图的形式呈现。

1.8 水稻 PEPCK 基因家族成员的进化关系分析

利用鉴定获得的水稻、高粱、小米、玉米、小麦、豌豆、拟南芥、番茄以及菠萝的 PEPCK 基因家族成员,于 MEGA 11 软件^[14] 中进行蛋白序列的多序列比对,并构建不同碳代谢类型植物 PEPCK 基因家族的系统进化树。

1.9 水稻 PEPCK 基因家族成员的共线性分析

利用 MCScanX 软件^[15] 进行共线性分析,分别分析水稻与小麦及水稻与玉米物种间的共线性关系,并利用 TBtools 软件进行物种间共线性分析结果的可视化。

2 结果与分析

2.1 水稻 PEPCK 基因家族的鉴定与理化性质分析

本研究在水稻全基因组中共鉴定出 4 个水稻 PEPCK 家族基因,分别为 *Os03g0255500*、*Os04g0592500*、*Os10g0204300* 以及 *Os10g0204400*,其中 *Os03g0255500* 有 2 个转录本,分别命名为 *Os03g0255500.1* 与 *Os03g0255500.2*。对水稻 PEPCK 基因家族成员进行蛋白理化性质分析与亚细胞定位预测,结果(表 1)表明,该家族基因编码的氨基酸数目在 48 ~ 663 个之间,编码氨基酸数目最少的为 *Os04g0204300* 基因(48 个),而氨基酸数目最多的为 *Os03g0255500.1* (663 个);分子量大小介于 5 356.22 ~ 73 204.02 之间;等电点位于 6.28 ~ 9.60 之间;不稳定指数位于 8.21 ~ 38.80 之间;平均疏水性处在 -0.379 ~ 0.004 之间;从亚细胞定位可以看出,水稻 PEPCK 蛋白主要定位在叶绿体与细胞质中,预示着其可能参与水稻光合作用及糖异生等生理过程。

2.2 水稻 PEPCK 基因家族的蛋白保守基序与基因结构分析

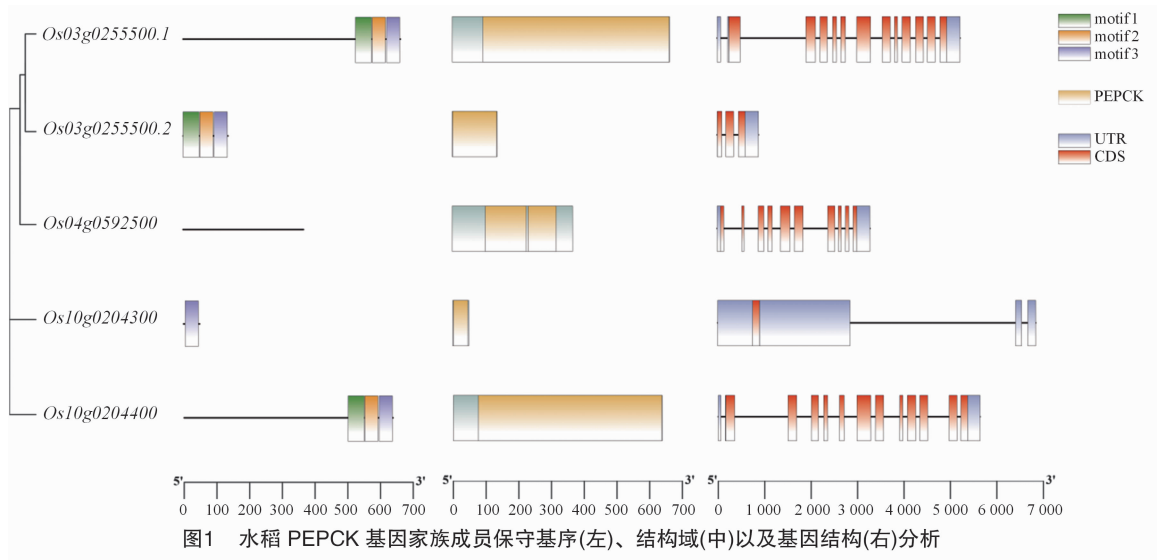
对水稻 PEPCK 基因家族进行保守基序分析,发

表 1 水稻 PEPCK 基因家族蛋白的理化性质分析与亚细胞定位

基因名称	氨基酸数目	分子质量 (u)	等电点	不稳定指数	平均疏水性	亚细胞定位
<i>Os03g0255500.1</i>	663	73 204.02	6.79	38.80	-0.379	细胞质
<i>Os03g0255500.2</i>	136	15 071.43	9.22	18.59	-0.240	叶绿体
<i>Os04g0592500</i>	367	39 279.02	9.12	36.71	0.004	叶绿体
<i>Os10g0204300</i>	48	5 356.22	9.60	8.21	-0.154	叶绿体
<i>Os10g0204400</i>	637	70 899.40	6.28	34.97	-0.358	细胞质

现 3 个保守基序 (motif1 ~ motif3), 其存在对于 PEPCK 结构域的功能及空间结构的稳定性起到关键作用。在这些保守基序中, motif3 相对最为保守, 除 *Os04g0592500* 外, 其他 PEPCK 基因家族蛋白都含有 motif3; 除 *Os04g0592500* 和 *Os10g0204300* 外都

含有 motif1、motif2 以及 motif3, 且保守基序排布顺序十分相似。同时对水稻 PEPCK 家族基因的基因结构进行分析, 观察到水稻 PEPCK 基因家族均含有内含子及外显子, 但不同基因的内含子和外显子差异较大 (图 1)。



2.3 水稻 PEPCK 基因家族的染色体定位

利用水稻注释文件获得基因在染色体上的位置信息并进行可视化, 结果 (图 2) 显示, 水稻中 4 个 PEPCK 基因分布于 3 条染色体上, 其中 *Os03g0255500* 分布在 3 号染色体, *Os04g0592500* 分布在 4 号染色体, *Os10g0204300* 以及 *Os10g0204400* 分布在 10 号染色体。

2.4 水稻 PEPCK 基因家族的启动子顺势作用元件预测分析

对水稻 PEPCK 基因家族进行启动子顺式作用元件预测, 结果 (图 3) 显示激素调控相关元件数目最多, 基因家族全体成员均含有激素调控相关元件, 且集中分布在脱落酸响应与茉莉酸甲酯响应两处, 表明水稻 PEPCK 基因家族成员的表达水平, 较多地受到脱落酸与茉莉酸甲酯的调控。其次是光

响应元件, 分析显示水稻 PEPCK 基因家族均含有光响应相关元件, *Os10g0204300* 启动子区域的光响应元件数目最多, 暗示着其表达或活性可能受到光诱导。在非生物胁迫方面, 除 *Os10g0204300* 的启动子区域没有干旱响应元件以外, 其他家族成员均含有。另外, 在植物生长方面, 除 *Os03g0255500.1* 外其他家族成员的启动子区域均含有植物生长相关元件。总的来看, 水稻 PEPCK 基因家族主要参与植物应对非生物胁迫、激素调控以及光响应等方面。

2.5 水稻 PEPCK 基因家族的表达分析

为了解水稻 PEPCK 基因家族的组织特异性表达, 利用水稻胚乳、胚、种子、芽、叶片、花序、雌蕊、花药等部位的表达量进行分析并绘制热图, 结果 (图 4) 显示 PEPCK 基因家族在水稻各组织不同阶段均有表达, 且在芽、胚、叶片、种子中表达量较高。

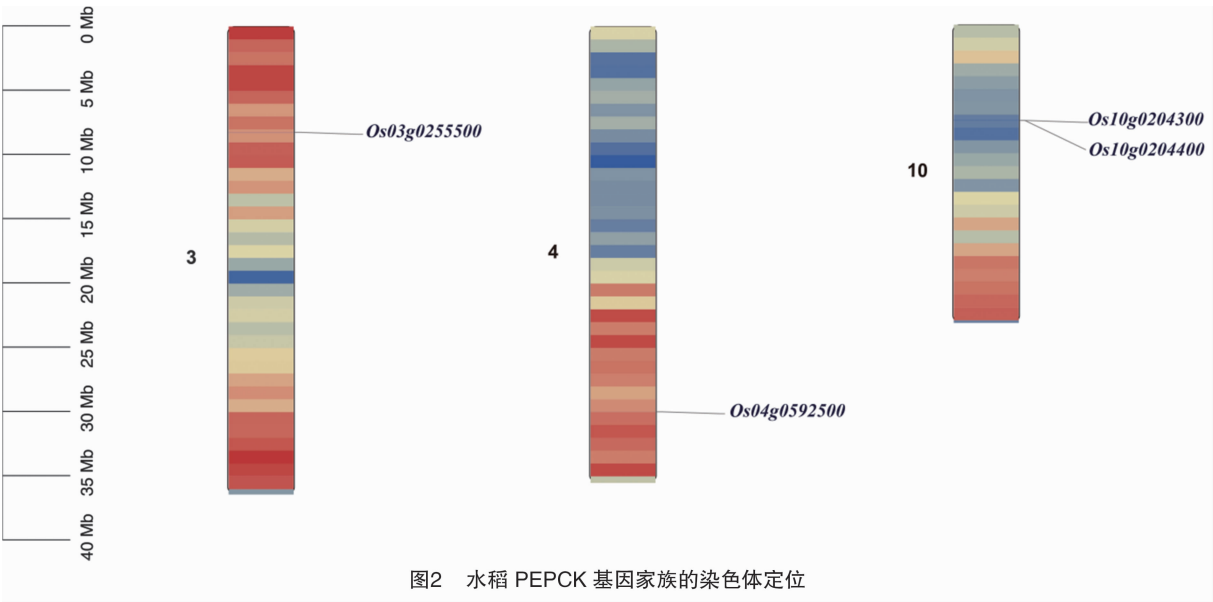


图2 水稻 PEPCK 基因家族的染色体定位

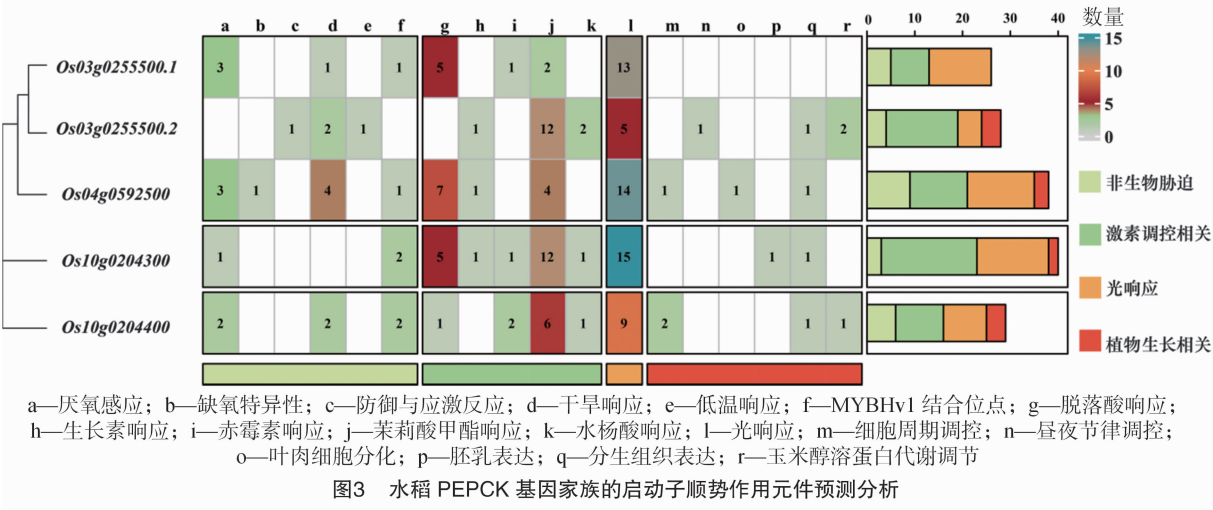


图3 水稻 PEPCK 基因家族的启动子顺势作用元件预测分析

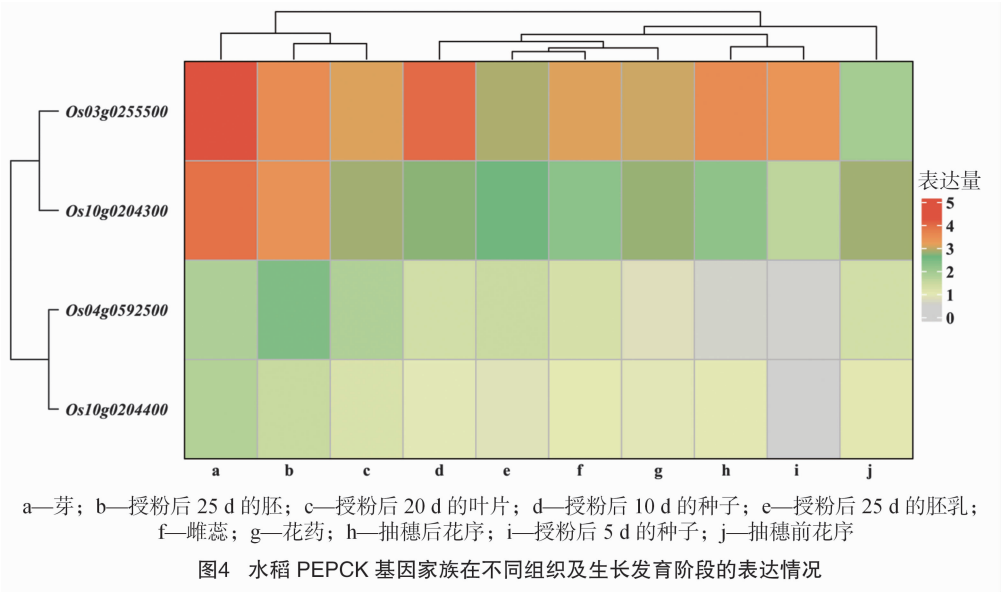


图4 水稻 PEPCK 基因家族在不同组织及生长发育阶段的表达情况

程中均呈上调趋势。此外,利用数据库中的表达谱数据,对在各种胁迫下 PEPCK 家族基因的表达水平进行了分析,发现 *Os03g0255500* 与 *Os10g0204400* 在糊粉层中的表达量分别在受到赤霉素与脱落酸诱导后显著上调(图 5)。

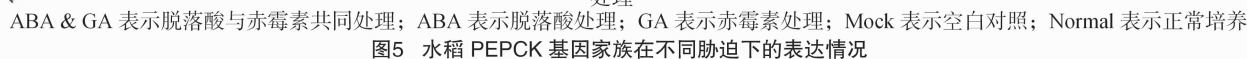


图5 水稻 PEPCK 基因家族在不同胁迫下的表达情况

3 讨论与结论

植物 PEPCK 广泛参与并调控植物的生长发育, 对其生命活动起到关键性作用^[16]。本研究通过生物信息学的方法, 筛选得到 4 个水稻 PEPCK 基因家族成员。对水稻 PEPCK 基因家族成员进行蛋白理化性质分析可知, 除 *Os04g0592500* 为疏水蛋白外, 其他家族成员均为亲水蛋白。同时, 水稻 PEPCK 基因家族成员蛋白的不稳定指数位于 8.21 ~ 38.80 之间, 表明水稻 PEPCK 基因家族蛋白均具有较好的稳定性。蛋白保守基序分析结果显示, 水稻 PEPCK 基因家族蛋白的保守基序的位置与排布顺序均十分相似, 表明 PEPCK 基因家族在水稻中具有较高的保守性。

为了进一步分析水稻 PEPCK 家族基因的进化关系,利用禾本科植物小麦、水稻和玉米 PEPCK 家族基因进行共线性分析,发现有 2 个水稻 PEPCK 基因与 6 个小麦 PEPCK 基因存在共线性,水稻 PEPCK 基因与 4 个玉米 PEPCK 基因存在共线性(图 7)。表明 PEPCK 基因的进化在禾本科不同物种间经历相似途径。

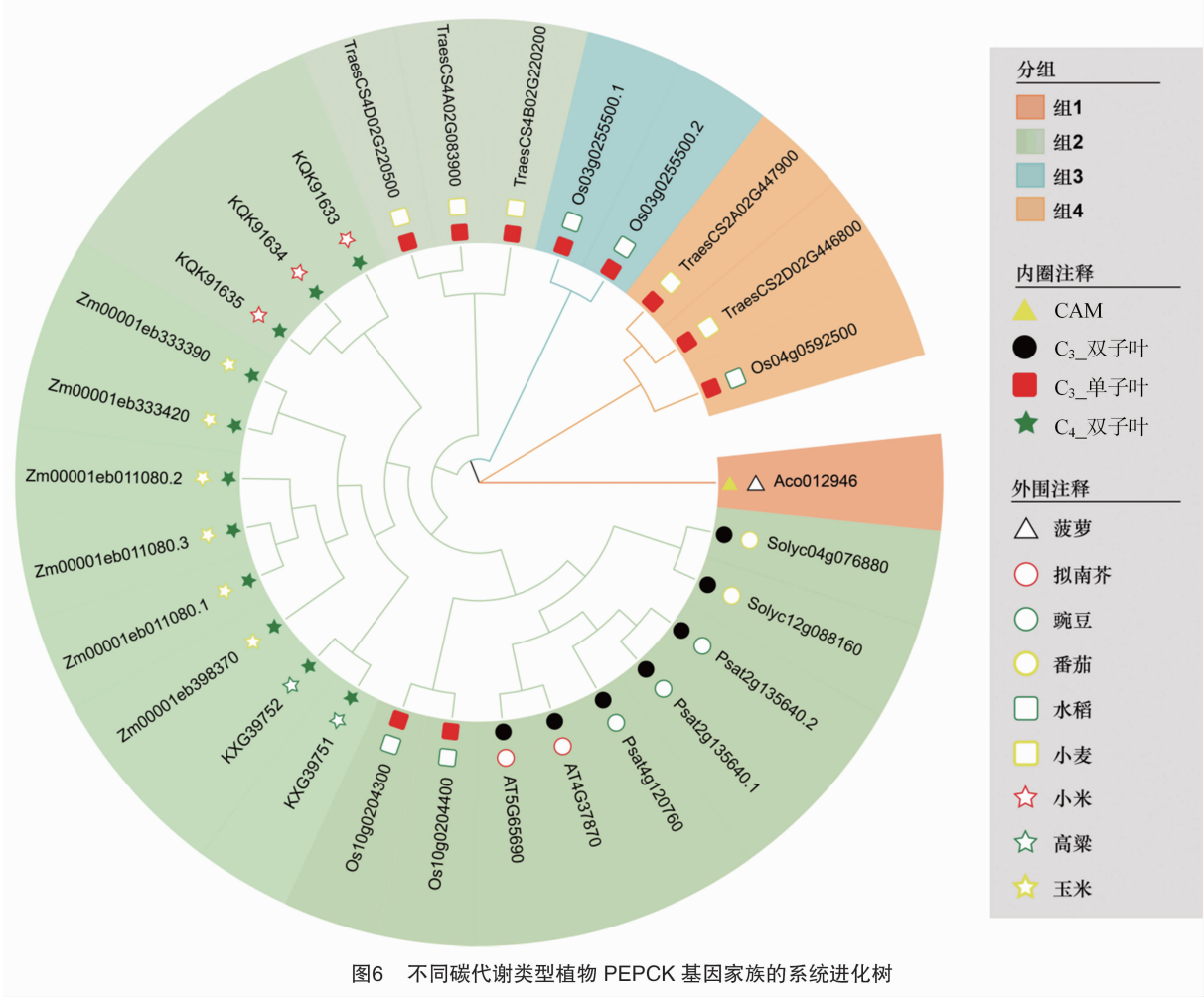


图6 不同碳代谢类型植物 PEPCK 基因家族的系统进化树

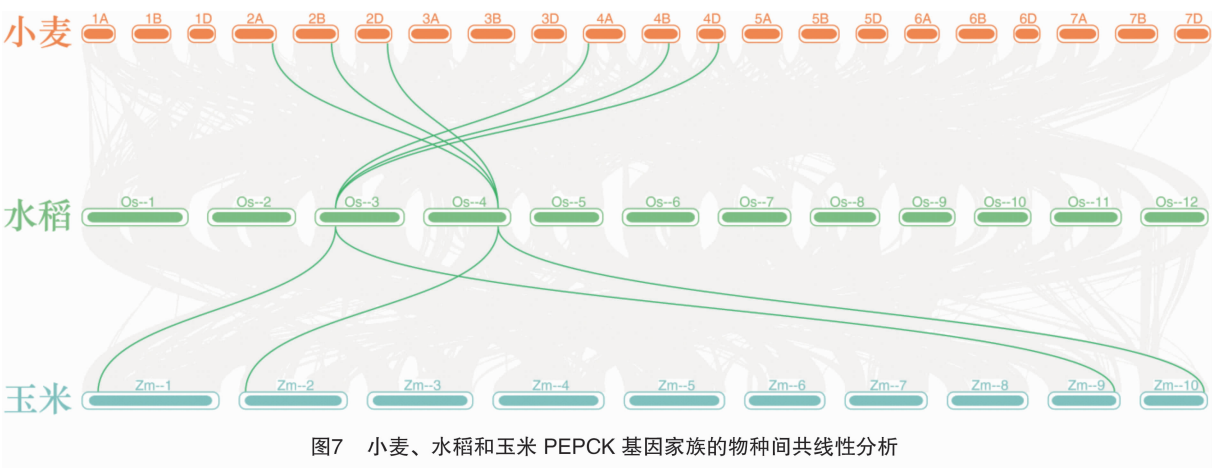


图7 小麦、水稻和玉米 PEPCK 基因家族的物种间共线性分析

时,水稻 PEPCK 基因家族的基因表达分析结果显示 PEPCK 基因家族在芽、胚、叶片与种子中的表达量较高,且在糊粉层中的表达量在受到赤霉素与脱落酸诱导后显著上调。有研究表明,脱落酸对于植物生长、发育以及应激反应具有重要调控作用^[17];脱落酸与赤霉素的比值对于水稻种子萌发以及后续生

长具有重要影响^[18];在水稻受到机械损伤后,茉莉酸甲酯通过调节免疫相关蛋白,可以保护植物免受由伤口引起的氧化应激和光合作用的损害^[19]。这些结果提示着水稻 PEPCK 基因家族不仅是光合作用碳反应中重要的一环^[20],在水稻种子萌发、防御与应激、抵御干旱胁迫等方面也具有重要意义,同

时其表达或活性可能受到光调控。

近十年来,植物 PEPCK 在高等植物中的功能与调控机制得到了广泛而深入的研究,为理解其在高等植物生命活动中的关键作用提供了深入的见解。同时,也有研究表明 PEPCK 在低等植物莱茵衣藻的糖异生过程中同样发挥着重要作用^[21]。基于这些背景信息,本研究旨在进一步分析水稻 PEPCK 基因家族在高等植物中的进化地位,从而更全面地揭示 PEPCK 在植物界的功能与演化历程。在构建不同碳代谢类型植物的 PEPCK 基因家族蛋白系统发育树后发现,水稻作为 C₃ 单子叶植物,其 PEPCK 基因家族中 *Os10g0204300*、*Os10g0204400* 与 C₃ 双子叶植物亲缘关系最近,而 *Os03g0255500.1* 及 *Os03g0255500.2* 与 C₄ 植物 PEPCK 亲缘关系较近。利用小麦、水稻、玉米进行共线性分析后发现,PEPCK 基因的进化在禾本科不同物种间经历相似途径。推测 PEPCK 基因在不同碳代谢类型植物以及单双子叶植物的分化中可能较为保守,同时也推测禾本科植物 PEPCK 基因发生进化的时间早于物种分化。

本研究利用生物信息学方法对水稻 PEPCK 基因家族成员进行了特征分析、表达分析以及进化关系分析,为今后进一步开展水稻中该家族基因成员的功能研究奠定了基础,而通过对水稻 PEPCK 的作用功能进行研究,可为利用该基因进行作物改良提供理论依据,具有重要的理论意义与经济价值。

参考文献:

- [1] Martín M, Plaxton W C, Podestá F E. Activity and concentration of non - proteolyzed phosphoenolpyruvate carboxykinase in the endosperm of germinating castor oil seeds.; effects of anoxia on its activity[J]. *Physiologia Plantarum*, 2007, 130(4): 484 - 494.
- [2] Huber S C, Kanai R, Edwards G E. Decarboxylation of malate by isolated bundle - sheath cells of certain plants having the C₄ - dicarboxylic acid cycle of photosynthesis[J]. *Planta*, 1973, 113(1): 53 - 66.
- [3] Ditttrich P, Campbell W H, Black C C. Phosphoenolpyruvate carboxykinase in plants exhibiting crassulacean acid metabolism[J]. *Plant Physiology*, 1973, 52(4): 357 - 361.
- [4] Jiang D, Zhang H, Cai H, et al. Overexpression of *ZmPCK2*, a phosphoenolpyruvate carboxykinase gene from maize confers enhanced tolerance to water deficit stress in rice[J]. *Plant Science*, 2022, 317: 111195.
- [5] Benedict C R, Beevers H. Formation of sucrose from malate in germinating castor beans. I. Conversion of malate to phosphoenol - pyruvate[J]. *Plant Physiology*, 1961, 36(5): 540 - 544.
- [6] Leegood R C, ap Rees T. Identification of the regulatory steps in gluconeogenesis in cotyledons of *Cucurbita pepo*[J]. *Biochimica et Biophysica Acta*, 1978, 542(1): 1 - 11.
- [7] Bailey K J, Leegood R C. Nitrogen recycling from the xylem in rice leaves; dependence upon metabolism and associated changes in xylem hydraulics[J]. *Journal of Experimental Botany*, 2016, 67(9): 2901 - 2911.
- [8] Mistry J, Chuguransky S, Williams L, et al. Pfam: The protein families database in 2021[J]. *Nucleic Acids Research*, 2021, 49(D1): D412 - D419.
- [9] Wilkins M R, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server[J]. *Methods in Molecular Biology*, 1999, 112: 531 - 552.
- [10] Horton P, Park K J, Obayashi T, et al. WoLF PSORT: protein localization predictor[J]. *Nucleic Acids Research*, 2007, 35(Suppl 2): W585 - W587.
- [11] Bailey T L, Johnson J, Grant C E, et al. The MEME suite[J]. *Nucleic Acids Research*, 2015, 43(W1): W39 - W49.
- [12] Chen C, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data[J]. *Molecular Plant*, 2020, 13(8): 1194 - 1202.
- [13] Xia L, Zou D, Sang J, et al. Rice expression database (RED): an integrated RNA - Seq - derived gene expression database for rice[J]. *Journal of Genetics and Genomics*, 2017, 44(5): 235 - 241.
- [14] Tamura K, Stecher G, Kumar S. MEGA 11: molecular evolutionary genetics analysis version 11[J]. *Molecular Biology and Evolution*, 2021, 38(7): 3022 - 3027.
- [15] Wang Y, Tang H, Debarry J D, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity[J]. *Nucleic Acids Research*, 2012, 40(7): e49.
- [16] Leegood R C, Walker R P. Regulation and roles of phosphoenolpyruvate carboxykinase in plants[J]. *Archives of Biochemistry and Biophysics*, 2003, 414(2): 204 - 210.
- [17] Chen K, Li G J, Bressan R A, et al. Absciscic acid dynamics, signaling, and functions in plants[J]. *Journal of Integrative Plant Biology*, 2020, 62(1): 25 - 54.
- [18] Hu Y, Song S, Weng X Y, et al. The heading - date gene inhibits seed germination by modulating the balance between abscisic acid and gibberellins[J]. *Crop Journal*, 2021, 9(2): 297 - 304.
- [19] Bertini L, Palazzi L, Proietti S, et al. Proteomic analysis of MeJa - induced defense responses in rice against wounding[J]. *International Journal of Molecular Sciences*, 2019, 20(10): 2525.
- [20] Shen Z, Dong X M, Gao Z F, et al. Phylogenic and phosphorylation regulation difference of phosphoenolpyruvate carboxykinase of C₃ and C₄ plants[J]. *Journal of Plant Physiology*, 2017, 213: 16 - 22.
- [21] Torresi F, Rodriguez F M, Gomez - Casati D F, et al. Two phosphoenolpyruvate carboxykinases with differing biochemical properties in *Chlamydomonas reinhardtii*[J]. *FEBS Letters*, 2023, 597(4): 585 - 597.