

# Mining and bioinformatics analysis of Cinnamyl CoA Reductase gene in Emei strawberry

Fawang Liu<sup>1,2</sup> Aihua Wang<sup>1</sup> Yin Zhang<sup>1</sup> Xinru Lv<sup>1</sup>

1. School of Biological and Food Engineering, Suzhou University, Suzhou, Anhui, 234000, China

2. Engineering Research Center for Development and High Value Utilization of Genuine Medicinal Materials in North Anhui Province, Suzhou, Anhui, 234000, China

## Abstract

To investigate the genes responsible for lignin biosynthesis in Emei strawberry stems, we searched the transcriptome database of Emei strawberry for the rate-limiting enzyme gene (cinnamoyl CoA reductase) in the lignin biosynthesis pathway. We utilized bioinformatics software DNAMAN, as well as online programs such as MEME and SOPMA, to predict the biological characteristics of cinnamoyl CoA reductase. Our findings revealed the presence of three cinnamoyl CoA reductase genes with complete open reading frames in Emei strawberries, sharing a sequence similarity of 76.82%. These genes were found to possess substrate binding and catalytic sites, and showed higher similarity with dihydroflavonol reductase and cinnamyl alcohol dehydrogenase in protein tertiary structure, suggesting their potential for catalytic activity and involvement in the lignification process of Emei strawberry stems.

## Keywords

Emei strawberry; cinnamoyl CoA reductase; lignification

# 峨眉草莓肉桂酰辅酶 A 还原酶基因的挖掘和生物信息学分析

刘发旺<sup>1,2</sup> 王爱华<sup>1</sup> 张银<sup>1</sup> 吕欣茹<sup>1</sup>

1. 宿州学院生物与食品工程学院, 中国·安徽 宿州 234000

2. 安徽省皖北道地药材开发与高值化利用工程研究中心, 中国·安徽 宿州 234000

## 摘要

为探究参与峨眉草莓茎秆木质化的基因, 从峨眉草莓转录组数据库中挖掘木质素生物合成途径上的限速酶基因肉桂酰辅酶 A 还原酶基因, 采用生物信息学软件 DNAMAN, 以及 MEME、SOPMA 等在线程序预测肉桂酰辅酶 A 还原酶生物学特性。结果表明从峨眉草莓中挖掘到 3 个具有完整开放阅读框的肉桂酰辅酶 A 还原酶基因, 其序列相似度为 76.82%, 且都具有底物结合和催化位点, 蛋白三级结构预测显示与二氢黄酮醇还原酶和肉桂醇脱氢酶空间结构相似, 推测其具有催化活性, 可能参与峨眉草莓茎秆木质化的形成。

## 关键词

峨眉草莓; 肉桂酰辅酶 A 还原酶; 木质化

【基金项目】宿州市科技计划重大专项(项目编号: SZKJXM202423); 宿州学院非财政资金科研项目(项目编号: 2024xhx196); 安徽省教育厅自然科学基金项目(项目编号: 2024AH051806); 宿州学院博士科研启动基金项目(项目编号: 2022BSK033); 宿州学院大学生创新创业训练计划项目(项目编号: ZCXM24-167、ZCXM24-201); 宿州学院教学研究和质量工程项目(项目编号: szxy2024jyxm25、szxy2024zybj01)。

【作者简介】刘发旺(1990-), 男, 中国安徽宿州人, 博士, 讲师, 从事基因功能研究。

## 1 引言

草莓是我国的一种主要水果, 我国草莓产量占据全球草莓产量的 32.02%。安徽是草莓的种植大省, 其中红颜草莓、峨眉草莓、奶油草莓等深受消费者的喜爱, 具有重要的研究价值 [1]。

肉桂酰辅酶 A 还原酶基因是植物中普遍存在的基因家族, 其参与植物木质素的合成 [2,3]。木质素在保持植物茎秆的硬度、增强植物抗倒伏性能上具有重要的作用, 在提升植物的生物质能方面也具有广泛的研究价值, 例如黑麦草作为动物饲料的重要来源, 提升黑麦草木质素含量可增加黑麦草的产量, 有力促进畜牧业发展 [4]。肉桂酰辅酶 A 还原酶基因在植物代谢通路上处于苯丙烷类生物合成通路, 该通路上参与木质素生物合成的酶除了肉桂酰辅酶 A 还原酶之外,

还有肉桂醇脱氢酶、儿茶酚氧甲基转移酶、苯丙氨酸解氨酶等，它们共同参与调控木质素的合成。在木质素合成的路径中，肉桂酰辅酶A还原酶所处的位置在碳代谢流的分支点，该位置对于调节碳代谢流向木质素或者黄酮等成分至关重要，因此，肉桂酰辅酶A还原酶是调控木质素合成的限速酶。基于肉桂酰辅酶A还原酶在植物碳代谢流调控中的重要性，越来越多的研究者对其开展基因克隆和功能研究，进一步发现其新功能及转录调控机制。

草莓的茎秆中木质素的含量对于草莓的生长、水分运输、草莓产量至关重要，直接影响草莓的品质。本文借助高通量转录组测序技术和生物信息学分析，首次对草莓新品种峨眉草莓中参与木质素生成的基因肉桂酰辅酶A还原酶基因进行挖掘和生物信息学分析，在此基础上，对草莓肉桂酰辅酶A还原酶基因的理化性质和生物学功能做了初步预测。研究结果为进一步的肉桂酰辅酶A还原酶基因克隆、功能研究，以及通过分子育种技术培育高木质素草莓新品种提供重要基础数据。

## 2 实验材料

峨眉草莓种植于安徽省皖北道地药材开发与高值化利用工程研究中心实验室温室，生长温度为24℃，相对湿度55%，光照/黑暗(16h/8h)。新鲜峨眉草莓在液氮中速冻，随后进行转录组测序。峨眉草莓转录组数据保存于本实验室。

## 3 方法

### 3.1 基因功能注释与挖掘

使用NR (<https://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr>)和KEGG (<https://www.genome.jp/kegg/>)数据库对峨眉草莓转录组数据进行功能注释。使用FPKM (Fragments Per Kilobase Million, Fragments Per Kilobase of exon model per Million mapped fragments)计算基因表达量。其中基因表达量大于100且注释为肉桂酰辅酶A还原酶(cinnamoyl-CoA reductase)的基因用于生物信息学分析。

### 3.2 生物信息学分析

使用NCBI的ORF工具将峨眉草莓的肉桂酰辅酶A还原酶基因序列转换为氨基酸序列，然后使用DNAMAN软件比对氨基酸的相似度。使用MEME在线工具分析肉桂酰辅酶A还原酶的保守序列，设置Motif个数为3个。SOPMA在线工具分析肉桂酰辅酶A还原酶的分子式、分子量、亲水性、疏水性、理论等电点，以及二级结构和蛋白的三级空间结构等，通过比较与之相似的蛋白结构，预测峨眉草莓中肉桂酰辅酶A还原酶的生物活性。

## 4 结果与讨论

### 4.1 峨眉草莓肉桂酰辅酶A还原酶基因挖掘

峨眉草莓转录组数据库中的基因经NR和KEGG数据

库注释，共有超过2万个基因获得注释，对这些注释的基因进行筛选，其中有5个基因注释为肉桂酰辅酶A还原酶基因，3个具有完整的编码区，其基因名分别为LOC101295856、LOC101290955和LOC105352966。进一步将这三个基因提交至美国国家生物技术信息中心的NCBI数据库进行检索确证，结果显示其功能注释都为cinnamoyl-CoA reductase，表明成功挖掘到了3个峨眉草莓肉桂酰辅酶A还原酶基因，命名为FaCCR1、FaCCR2和FaCCR3。

### 4.2 峨眉草莓肉桂酰辅酶A还原酶氨基酸序列分析

使用DNAMAN软件对3个峨眉草莓肉桂酰辅酶A还原酶氨基酸序列比对，其序列相似度69.46%，3个肉桂酰辅酶A还原酶序列中都含有底物结和位点，推测可能具有酶催化活性(图1A)。使用MEME对峨眉肉桂酰辅酶A还原酶的保守域分析，共获得3个保守域，分别是保守域1(GACCCHSARGCAGAAKT AHTKGAMCCWGCTKTG ARGGGWACDCTTAATGT)、保守域2(GTGTGYG TR ACAGGAGCWSWGGRTTYRYWGCHTCMTGGSTSGTC AAGCT)和、保守域3(GTKGATGTTMGWGATKTWGC YRAWGCACWBMTTCTWGCMTWTRA GAAMSC)(图1B)。肉桂酰辅酶A还原酶一级结构分析显示，FaCCR1的分子式为 $C_{1610}H_{2550}N_{422}O_{477}S_{12}$ ，分子量为35835.22Da，氨基酸数326，理论等电点为6.26，脂肪族指数92.64，疏水指数为-0.098，表明为亲水性蛋白；FaCCR2的分子式为 $C_{1608}H_{2539}N_{415}O_{496}S_{13}$ ，分子量为36038.11Da，氨基酸数326，理论等电点为5.47，脂肪族指数87.85，疏水指数为-0.206，表明为亲水性蛋白；FaCCR3的分子式为 $C_{1598}H_{2527}N_{409}O_{479}S_{13}$ ，分子量为35549.88Da，氨基酸数323，理论等电点为5.28，脂肪族指数96.22，疏水指数为0.021，表明为疏水性蛋白。肉桂酰辅酶A还原酶二级结构分析显示，3个肉桂酰辅酶A还原酶都以 $\alpha$ -螺旋和无规则卷曲为主，FaCCR1中两种二级结构占比分别是39.57%和44.79%，FaCCR2中其占比都是42.64%，FaCCR3中其占比分别是39.01%和45.20%(图1C)。蛋白三级结构显示，FaCCR1与蔷薇科玫瑰的肉桂酰辅酶A还原酶(蛋白序列号A0A2P6S7Y0.1.A)空间结构最相似，相似度达84.57%，FaCCR2与蔷薇科杏的二氢黄酮醇还原酶(蛋白序列号A0A6J5X640.1.A)空间结构最相似，相似度达73.13%，FaCCR3与蔷薇科玫瑰的肉桂醇脱氢酶(蛋白序列号A0A2P6S7Y2.1.A)空间结构最相似，相似度达83.28%。蛋白空间结构预测表明峨眉草莓肉桂酰辅酶A还原酶表现出还原活性，参与黄酮、肉桂醛等苯丙烷类的还原，而这些苯丙烷类是木质素的前体物质。3个峨眉草莓肉桂酰辅酶A还原酶的生物信息学分析初步表明挖掘到的这2个基因可能参与峨眉草莓木质素的合成(图1D)。

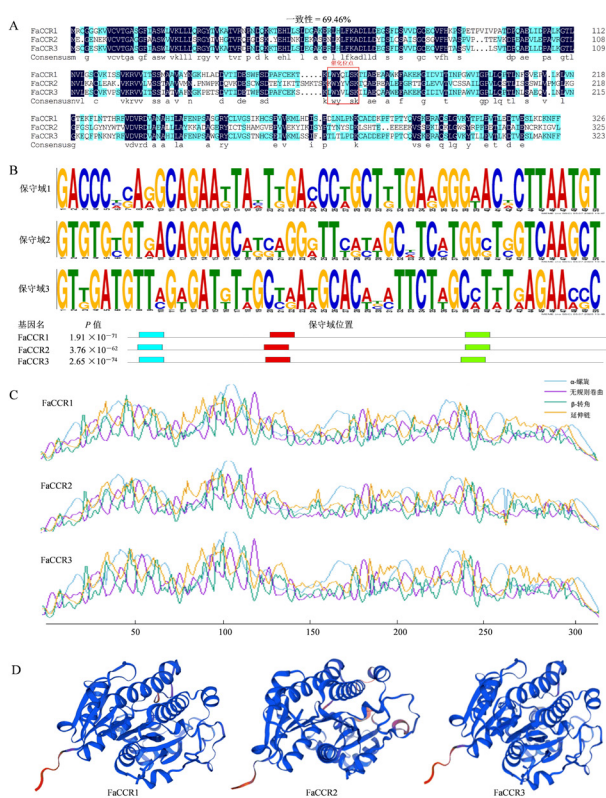


图1 峨眉草莓的肉桂酰辅酶A还原酶的生物信息学分析

(A) 肉桂酰辅酶A还原酶氨基酸序列比对；(B) 肉桂酰辅酶A还原酶保守结构域分析；(C) 肉桂酰辅酶A还原酶二级结构分析；(D) 肉桂酰辅酶A还原酶蛋白三级结构分析。

### 4.3 峨眉草莓肉桂酰辅酶A还原酶基因序列分析与功能预测

峨眉草莓肉桂酰辅酶A还原酶基因 *FaCCR1*、*FaCCR2* 和 *FaCCR3* 都具有完整的开放阅读框，基因长度分析显示 *FaCCR1* 基因长度 982bp、*FaCCR2* 基因长度 982bp、*FaCCR3* 基因长度 972bp。峨眉草莓肉桂酰辅酶A还原酶基因长度在 1000bp 附近，符合植物肉桂酰辅酶A还原酶基因长度特征，例如辣椒肉桂酰辅酶A还原酶基因（基因序列号 NM\_001324825.1）长度为 1005bp，番果肉桂酰辅酶A还原酶基因（基因序列号 NM\_001247368.1）长度为 999bp，高粱肉桂酰辅酶A还原酶基因（基因序列号 OM751894.1）长度为 909bp。因此，从基因序列长度来看，本文挖掘到的 3 个肉桂酰辅酶A还原酶基因可能具有该酶的生物活性。

对峨眉草莓肉桂酰辅酶A还原酶基因编码的氨基酸序列分析显示，它们都具有酶的底物结合位点和催化位点“NWYCYGK”，研究表明该位点是判断是否是肉桂酰辅酶A还原酶的特征指标。本研究中 *FaCCR1* 该位点突变为

“LWYQLSK”、*FaCCR2* 该位点突变为“KWYYVSK”、*FaCCR3* 该位点突变为“LWYVLSK”，这可能与植物的进化和植物适应环境导致的基因突变有关。峨眉草莓肉桂酰辅酶A还原酶的分子量都在 30000Da（30kDa）以上，这与其他研究报道的一致，例如苎麻的肉桂酰辅酶A还原酶分子量 30.67kDa。综合基因序列、氨基酸序列以及其他生物信息学分析初步表明峨眉草莓肉桂酰辅酶A还原酶基因 *FaCCR1*、*FaCCR2* 和 *FaCCR3* 可能具有生物活性。

作为植物中的一类家族基因，肉桂酰辅酶A还原酶基因数量较多且功能多样，例如水稻中有 33 个肉桂酰辅酶A还原酶基因，白杨中有 8 个肉桂酰辅酶A还原酶基因。肉桂酰辅酶A还原酶基因不仅参与木质素的合成，还参与植物应激响应，例如在模式植物拟南芥中 *AtCCR1* 参与木质素的合成，而 *AtCCR2* 则参与植物对外界环境胁迫如干旱、病虫害的响应 [5]。本文我们挖掘到了峨眉草莓 *FaCCR1*、*FaCCR2* 和 *FaCCR3* 基因，经生物信息分析预测了其特性，未来可对其进行基因克隆，并进行体外蛋白的诱导表达和酶的催化活性验证，或者通过转基因过表达、RNAi 干扰、CRISPR-Cas9 基因敲除来探究其在峨眉草莓生长发育或者木质素合成中的功能。

## 5 结语

本文对峨眉草莓肉桂酰辅酶A还原酶进行了系统的挖掘，获得了 3 个完整的肉桂酰辅酶A还原酶基因 *FaCCR1*、*FaCCR2* 和 *FaCCR3*，采用多种生物信息学工具预测了其理化性质。研究表明 3 个肉桂酰辅酶A还原酶二级结构都以  $\alpha$ -螺旋和无规则卷曲为主，*FaCCR1* 和 *FaCCR2* 为亲水性蛋白，*FaCCR3* 疏水性蛋白。三级结构表明 3 个肉桂酰辅酶A还原酶具有还原性，可能参与肉桂醇、肉桂醛等木质素前体物质的合成，进而调控峨眉草莓木质素的生成。研究结果为峨眉草莓的种质改良提供了基因数据，指导峨眉草莓的育种。

## 参考文献

- [1] 李薇,王哲,刘妍,等.我国草莓产业发展新形势、存在的问题及建议[J].中国南方果树,2025,54(2):240-248.
- [2] 王建国.植物肉桂酰辅酶A还原酶的研究进展[J].辽宁林业科技,2023(2):43-46.
- [3] 杨硕知.木质素生物合成途径中关键酶肉桂酰辅酶A还原酶研究进展综述[J].安徽农学通报,2010,16(12):44-45.
- [4] 张昊,陈亚娟,姜廷波,等.木质素生物合成翻译后修饰调控研究进展[J].生物技术进展,2024,14(4):519-528.
- [5] 刘晓晶,崔浪军,夏飞,等.拟南芥肉桂酰辅酶A还原酶(*AtCCR1/2*)基因的生物信息学分析[J].陕西师范大学学报(自然科学版),2011,39(03):67-72.