

# 稻瘟病菌中假定三羧酸循环相关蛋白的生物信息学分析

石磊<sup>1</sup>, 骆琼<sup>1</sup>, 王建华<sup>1</sup>, 衣雨娇<sup>1</sup>, 郑文辉<sup>1</sup>, 周洁<sup>1</sup>, 李光普<sup>1,2</sup>

5 (1. 福建农林大学功能基因组研究中心, 福建, 福州, 350002;  
2. 美国俄克拉荷马大学健康科学中心)

**摘要:** 三羧酸循环是需氧生物体物质和能量的代谢枢纽, 在生物体的各项生命活动中发挥重要的功能。但是, 迄今未见对稻瘟病菌中的三羧酸 (TCA) 循环相关蛋白进行鉴定和开展功能研究。本研究从酿酒酵母的 17 个 TCA 循环相关蛋白中比对得到稻瘟病菌的 18 个假定的 TCA 循环相关蛋白; 功能结构域和系统进化分析结果表明, 稻瘟病菌与酿酒酵母中 TCA 循环关键酶相应同源蛋白含相同的结构域, 与禾谷镰刀菌、粗糙脉孢菌及构巢曲霉的进化关系最相近; 在同源比对的基础上, 进一步绘制了稻瘟病菌中 TCA 循环的代谢网络图, 为系统分析稻瘟病菌 TCA 循环的网络系统提供参考。基因表达数据库结果表明, 稻瘟病菌 TCA 循环相关基因在菌丝和附着胞形成阶段的均有表达, 但基因间表达量差异很大

15 **关键词:** 稻瘟病菌; 三羧酸循环; 生物信息学; 柠檬酸合酶; 异柠檬酸脱氢酶;  $\alpha$ -酮戊二酸脱氢酶

**中图分类号:** S432.4+4

## 20 The Bioinformatics Analysis of Putative Tricarboxylic Acid Cycle-Related Proteins in *Magnaporthe oryzae*

SHI Lei<sup>1</sup>, LUO Qiong<sup>1</sup>, WANG Jianhua<sup>1</sup>, YI Yujiao<sup>1</sup>, ZHENG Wenhui<sup>1</sup>, ZHOU Jie<sup>1</sup>,  
LI Guangpu<sup>1,2</sup>

25 (1. The functional genomics center, Fujian Agriculture and Forestry University, Fujian Fuzhou, 350002;

2. University of Oklahoma Health Science Center)

**Abstract:** Tricarboxylic acid cycle is hub material and energy metabolism and plays an important function in the life activities of aerobic organisms. However, the TCA cycle-related proteins in the plant pathogenic fungi, *Magnaporthe oryzae*, have not yet been identified and studied for their functions. In this study, we identified 18 putative TCA cycle-related proteins in *M.oryzae* by using 17 *Saccharomyces cerevisiae* proteins through BLAST search. The corresponding homologous protein contains the same functional domains in *S. cerevisiae* and *M.oryzae*., the key proteins of TCA cycle in *M. oryzae* have the closest evolutionary relationship with *Fusarium graminearum*, *Neurospora crassa* and *Aspergillus nidulans*.by phylogenetic analysis. Based on the above results the TCA cycle metabolic network is drawn providing reference for further system analysis of rice blast fungus TCA cycle. In addition, all genes in TCA cycle expressed between appressorium and hyphal different stages, but there is obvious difference among genes.

35 **Key words:** *Magnaporthe oryzae*, Tricarboxylic acid cycle, Bioinformatics, Citrate synthase, Isocitrate dehydrogenase,  $\alpha$ -ketoglutarate dehydrogenase

## 40 0 引言

新陈代谢是最基本的生命现象之一, 生物体通过新陈代谢获得能量和营养物质, 进而维系生命活动。三羧酸循环 (tricarboxylic acid cycle, TCA 循环) 又称柠檬酸循环、克氏循环 (Krebs cycle), 是需氧生物普遍存在的新陈代谢途径。TCA 循环不仅是三大营养化合物(糖

基金项目: 国家自然科学基金 (31270179)

作者简介: 石磊 (1989 年-), 男, 硕士研究生, 主要研究方向: 分子植物病理学

通信联系人: 李光普 (1963 年-), 男, 教授, 主要研究方向: 细胞生物学, 生物化学与分子生物学. E-mail: guangpu-li@ouhsc.edu

类、脂类和氨基酸)生物氧化的共同通路,同时也是生物代谢的枢纽,为生物的各种合成代谢提供前体物质和能量。

由稻瘟病菌引起的稻瘟病对世界粮食生产影响重大<sup>[1,2]</sup>。稻瘟病菌分生孢子通过特异的侵染结构——附着胞,侵入宿主,附着胞膨压的形成和黑色素化依赖于分生孢子中糖原和脂肪粒的降解和转运<sup>[3,4]</sup>。已有的研究表明:过氧化物酶体中的 $\beta$ 氧化多功能蛋白 Mfp1 的突变,影响脂肪的降解和转运以及附着胞的功能,使突变体的致病性降低<sup>[5]</sup>;线粒体脂肪酸的 $\beta$ 氧化酶 Ech1 缺失后,附着胞黑色素减少,且不能利用短链脂肪酸<sup>[6]</sup>;稻瘟病菌糖原分解代谢相关基因 *AGL1* 和 *GPH1* 的缺失突变体中糖原代谢受到影响,导致突变体致病性降低<sup>[7]</sup>;此外,肉毒碱乙酰转移酶 Pth2、Crat1 通过影响乙酰 CoA 在细胞内的转运,从而影响脂肪的代谢和糖异生等途径,使得附着胞黑色素的形成减少,最终导致突变体致病性丧失<sup>[8,9]</sup>;肉毒碱-酰基肉毒碱载体蛋白 MoCRC1 缺失后附着胞膨压降低,侵染菌丝的生长速度减慢,致病性丧失<sup>[10]</sup>。由此可见,稻瘟病菌的能量代谢,特别是在附着胞形成阶段的能量代谢对稻瘟病菌的侵染及致病极为重要。

然而,不管是线粒体和过氧化物酶体中脂肪酸的 $\beta$ 氧化、糖原的水解代谢,还是肉毒碱介导乙酰 CoA 的转运都直接或间接的影响细胞内的三羧酸循环过程,因而我们推测 TCA 循环相关蛋白对稻瘟病菌附着胞的发育及成熟起重要作用。由于 TCA 循环相关蛋白在稻瘟病菌中的组成及功能还未被系统研究。因此,本文利用生物信息学的方法,对稻瘟病菌的 TCA 循环相关蛋白进行了鉴定和分析,为进一步研究 TCA 循环在稻瘟病菌生理和致病过程中的功能奠定了基础。

## 1 方法

### 1.1 稻瘟病菌 TCA 循环相关蛋白氨基酸序列的获得

从酵母生化通路数据库 (<http://pathway.yeastgenome.org/>) 中找出酿酒酵母 TCA 循环通路中的相关蛋白。将上述蛋白的基因号放入酵母数据库 (<http://www.yeastgenome.org/>) 进行检索获得 TCA 循环相关蛋白的各个氨基酸序列,进而在 BLASTP vs. fungi 中选择 *Magnaporthe oryzae* (<http://www.yeastgenome.org/cgi-bin/blast-fungal.pl?name=YPR001W&suffix=prot>) 运行 WU-BLAST 得到稻瘟病菌中同源基因。本研究选择 Score(bits)>50 且 E-Value < $e^{-6}$  的基因作为候选基因。为确定候选基因是否为真正的酿酒酵母的同源基因,将上述获得的候选基因检索 NCBI HomoloGene (<http://www.ncbi.nlm.nih.gov/homologene/>) 数据库选择 *S.cerevisiae* 得到稻瘟病菌候选基因在酿酒酵母中的同源基因,初步确定稻瘟病菌中编码 TCA 循环相关蛋白的候选基因。

### 1.2 蛋白基本信息分析

根据已获得候选基因的氨基酸序列,利用在线分析软件 ExPASy ProtParam<sup>[11]</sup> (<http://web.expasy.org/protparam/>) 工具分析其理化性质;利用 CBS 预测服务器 (CBS Prediction Servers) 中的 TMHMM Server v.2.0<sup>[12]</sup> (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>)、SignalP 4.1 Server<sup>[13]</sup> (<http://www.cbs.dtu.dk/services/SignalP/>) 和 TargetP1.1Server<sup>[14]</sup> (<http://www.cbs.dtu.dk/services/TargetP/>) 分别对候选基因的跨膜结构域、信号肽和蛋白亚细胞定位进行预测和分析;同时利用 MitoProt (<http://ihg.gsf.de/ihg/mitoprot.html>) 对候选蛋白是否含有线粒体转运肽进行预测<sup>[15]</sup>。

### 1.3 TCA 循环相关蛋白的结构域预测

利用在线蛋白结构域分析软件 SMART<sup>[16]</sup> (<http://smart.embl-heidelberg.de/>) 对酿酒酵母和稻瘟病菌中的 TCA 循环相关蛋白进行结构域分析, 进一步将分析结果绘制成 TCA 循环关键酶的结构域模式图。

### 1.4 TCA 循环关键酶的系统进化树分析

根据酵母菌中 TCA 循环关键酶的氨基酸序列, 在 NCBI 中找到动物: *Homo sapiens* (智人) 和 *Mus musculus* (家鼠); 植物: *Arabidopsis thaliana* (拟南芥) 和 *Oryza sativa japonica* (水稻) 和真菌: *Aspergillus nidulans* (构巢曲霉)、*Candida albicans* (白色念珠菌)、*Fusarium graminearum* (禾谷镰刀菌)、*Neurospora crassa* (粗糙脉孢菌)、*Magnaporthe oryzae* (稻瘟病菌) 和 *Saccharomyces cerevisiae* (酿酒酵母) 中参与 TCA 循环的关键酶的氨基酸序列及 NCBI 登录号。本研究用 ClustalX2 软件进行多序列比对, 采用默认参数生成 aln 文件后用 MEGA6 软件转化为 meg 文件, 并用邻接法 (Neighbor-Joining method) 运行 Bootstrap method 进行 1000 次重复抽样检验分析, 构建成系统发育树。

### 1.5 基因表达量分析

稻瘟病菌 TCA 循环候选基因菌丝阶段 (液体的 CM 培养基和 MM 培养基) 和附着胞形成阶段的基因表达量数据来源于稻瘟病菌公共基因表达量数据库 (<http://cogeme.ex.ac.uk/supersage/>)<sup>[17]</sup>。

## 2 结果

### 2.1 酿酒酵母和稻瘟病菌中参与 TCA 循环蛋白序列的获得

本研究中酿酒酵母的 17 个 TCA 循环相关蛋白来自酵母生化通路数据库, 与 Przybyla-Zawislak B<sup>[18]</sup> 报导的结果相一致 (图 1A)。其中乙酰 CoA 和草酰乙酸经柠檬酸合酶 (CIT1、CIT3) 催化合成柠檬酸, 顺乌头酸酶 (ACO1、ACO2) 将柠檬酸转化为异柠檬酸, 异柠檬酸脱氢酶 (IDH1、IDH2) 在 NAD<sup>+</sup> 存在下将异柠檬酸氧化脱羧为  $\alpha$ -酮戊二酸并产生 NADH 和 CO<sub>2</sub>; 在 CoA 和 NAD<sup>+</sup> 存在下,  $\alpha$ -酮戊二酸被  $\alpha$ -酮戊二酸脱氢酶复合体 (KGD1、KGD2 和 LPD1) 氧化脱羧成琥珀酰 CoA 同时产生一个 NADH 和 CO<sub>2</sub>; 琥珀酰 CoA 被琥珀酰辅酶 A 合成酶 (LSC1、LSC2) 催化生成琥珀酸, 经琥珀酸脱氢酶复合体 (SDH1、SDH2、SDH3 和 SDH4) 催化生成延胡索酸, 并将电子转移到呼吸链; 延胡索酸被延胡索酸酶 (FUM1) 催化成苹果酸, 苹果酸脱氢酶 (MDH1) 能够利用苹果酸生成草酰乙酸提供下一轮 TCA 循环的底物。 (图 1A)

利用同源比对方法, 在稻瘟病菌中获得 46 个候选蛋白的基因。由于部分基因在序列具有同源性, 而造成多次计数, 去除重复计数的基因后, 最终获得 32 个候选基因。并找到这些基因在 *S.cerevisiae* 中的同源基因 (表 1)。结果显示: 同源比对得到稻瘟病菌中首个候选基因均为酵母中参与 TCA 循环的同源基因, 除 SDH3 和 SHD4 外, 所有的 Score(bits) 值均大于 300, E-Value 均小于 e<sup>-80</sup>, 这些结果表明在稻瘟病菌和酵母菌中参与 TCA 循环的相关蛋白具有较高的保守性。从 32 个候选基因中, 去除 8 个酵母同源基因为非 TCA 相关的基因 (MGG\_04842、MGG\_01553、MGG\_01566、MGG\_05223、MGG\_09878、MGG\_12749、MGG\_03619、MGG\_08835) 以及在酵母中不含同源基因且 Score(bits) 值低于 100, E-Value 大于 e<sup>-25</sup> 的 4 个候选基因 (MGG\_09814、MGG\_09813、MGG\_06719、MGG\_10453)。初

120 步确定 20 个稻瘟病菌 TCA 相关基因, 包括 17 个酵母 TCA 循环的同源基因和 3 个 (MGG\_06547、MGG\_04449、MGG\_15167) 在酵母中不含同源基因的基因。

进一步用 MGG\_06547、MGG\_04449 和 MGG\_15167 基因编码的氨基酸序列在酵母数据库中进行同源比对。比对结果显示其总分最高的同源蛋白分别为 LYS12(homoisocitrate dehydrogenase)、KGD2 和 LSC1, 且 E value 均低于  $e^{-20}$ , Identities 均高于 30%。表明 MGG\_04449

125 和 MGG\_15167 可能编码酵母的同源蛋白 KGD2 和 LSC1 参与稻瘟病菌的 TCA 循环。

表 1: 酵母 TCA 基因同源比对结果

Table 1. The results of yeast TCA homologue BLAST

No.	<i>S.cerevisiae</i>	<i>M.oryzae</i>	Description	Score(bits)	E value	Identities	<i>S.cerevisiae</i> homologene
1	CIT1	MGG_07202	citrate synthase	526.4	1.20E-152	64%	CIT1
2		MGG_02617	2-methylcitrate synthase	408.8	3.00E-117	49%	CIT3
3	CIT3	MGG_02617	2-methylcitrate synthase	423.3	1.40E-121	50%	CIT3
4		MGG_07202	citrate synthase	291.6	5.90E-82	50%	CIT1
5	ACO1	MGG_03521	aconitate hydratase	1029.8	3.60E-304	70%	ACO1
6		MGG_08584	aconitate hydratase	676.4	8.90E-198	51%	ACO2
7		MGG_04842	Homoaconitase	135.3	4.80E-32	26%	LYS4
8		MGG_01553	3-isopropylmalate dehydratase	59.3	4.90E-24	23%	LEU1
9		MGG_09814	3-isopropylmalate dehydratase large subunit 2	89.2	3.80E-21	25%	
10	ACO2	MGG_08584	aconitate hydratase	807.7	2.60E-237	58%	Aco2
11		MGG_03521	aconitate hydratase	709.8	7.60E-208	52%	Aco1
12		MGG_04842	homoaconitase	140.2	1.80E-33	27%	LYS4
13		MGG_01553	3-isopropylmalate dehydratase	72.6	2.20E-25	23%	LEU1
14		MGG_09814	3-isopropylmalate dehydratase large subunit 2	87.1	9.30E-23	24%	
15	IDH1	MGG_01995	isocitrate dehydrogenase subunit 1	364.8	5.40E-104	54%	IDH1
16		MGG_01712	isocitrate dehydrogenase subunit 2	247.2	1.30E-68	41%	IDH2
17		MGG_06547	tartrate dehydrogenase/decarboxylase ttuC	115.6	5.70E-29	27%	
18		MGG_01566	homoisocitrate dehydrogenase	95.2	7.90E-23	29%	LYS12
19		MGG_09813	3-isopropylmalate dehydrogenase	75.5	1.40E-14	25%	
20		MGG_05223	3-isopropylmalate dehydrogenase	72.3	1.60E-13	26%	LEU2
21	IDH2	MGG_01712	isocitrate dehydrogenase subunit 2	376.1	2.20E-107	59%	IDH2
22		MGG_01995	isocitrate dehydrogenase subunit 1	255.3	4.80E-71	43%	IDH1
23		MGG_06547	tartrate dehydrogenase/decarboxylase ttuC	124	1.60E-31	32%	
24		MGG_01566	homoisocitrate dehydrogenase	105.7	5.30E-26	33%	LYS12
25		MGG_05223	3-isopropylmalate dehydrogenase	92.7	8.10E-26	32%	LEU2
26		MGG_09813	3-isopropylmalate dehydrogenase	85.3	4.10E-18	30%	
27	KGD1	MGG_12767	2-oxoglutarate dehydrogenase	1190.7	0.00E+00	63%	KGD1
28	KGD2	MGG_03149	dihydrolypoyllysine-residue succinyltransferase	318	2.20E-115	70%	KGD2
29		MGG_09878	dihydrolypoyllysine-residue acetyltransferase component-pyruvate dehydrogenase complex	102.9	2.40E-29	51%	LAT1
30		MGG_04449	branched-chain alpha-keto acid lipoamide acyltransferase	100.5	1.90E-26	28%	
31	LPD1	MGG_17072	dihydrolypoyl dehydrogenase	531.7	6.60E-154	61%	LPD1
32		MGG_12749	glutathione reductase	167.7	1.20E-44	26%	GLR1
33	LSC1	MGG_05480	succinyl-CoA ligase subunit alpha	305.3	4.30E-86	57%	LSC1
34		MGG_15167	succinyl-CoA ligase subunit alpha	212.4	4.10E-58	39%	

35		MGG_06719	ATP-citrate synthase subunit 1	78.3	2.60E-15	27%	
36	LSC2	MGG_16230	succinyl-CoA ligase subunit beta	403.5	1.20E-115	55%	LSC2
37		MGG_15167	succinyl-CoA ligase subunit alpha	81.8	1.30E-15	39%	
38	SDH1	MGG_00168	hypothetical protein	864.3	2.30E-254	72%	SDH1
39		MGG_03619	fumarate reductase	55	4.10E-14	46%	OSM1
40	SDH2	MGG_00167	hypothetical protein	343	2.00E-97	71%	SDH2
41	SDH3	MGG_04876	succinate dehydrogenase cytochrome B subunit	51.2	1.40E-09	29%	SDH3
42	SDH4	MGG_00666	hypothetical protein	76.9	2.60E-17	34%	SDH4
43	FUM1	MGG_17962	fumarate hydratase	578.8	2.00E-168	70%	FUM1
44	MDH1	MGG_09367	malate dehydrogenase	380.3	1.20E-108	64%	MDH1
45		MGG_08835	malate dehydrogenase	305.3	4.30E-86	55%	MDH3
46		MGG_10453	L-lactate dehydrogenase	51.9	8.80E-07	25%	

## 2.2 稻瘟病菌 TCA 循环候选蛋白的基本信息分析

蛋白质理化性质是蛋白质序列分析的内容之一，分析蛋白质的理化特征可以为蛋白质的研究提供重要的参考依据。通过 ProtParam 对初步确定的 20 个稻瘟病菌 TCA 相关蛋白氨基酸序列的理论等电点 PI、分子质量、负电荷残基总数 (Asp + Glu)、正电荷残基总数 (Arg + Lys)、总平均亲水性(Grand average of hydropathicity, GRAVY)、稳定性等进行分析(表 2)。比较发现：MGG\_04876 和 MGG\_00666 带负电荷的氨基酸含量很低，仅为 3.80% (7/184) 和 4.68% (8/171)，而其他蛋白负电荷 (或正电荷) 氨基酸残基这一比例基本在 9%-13% 之间，并且其理论等电点高达 10.96 和 10.35。此外 MGG\_12767、MGG\_03149、MGG\_04449、MGG\_15167、MGG\_00167 和 MGG\_00666 理论不稳定性系数较高，呈不稳定状态。

## 2.3 稻瘟病菌 TCA 循环候选蛋白的信号肽、转运肽及跨膜和定位分析

参与 TCA 循环的蛋白由细胞核基因编码，翻译后产生的新生蛋白 N 端含约 20~80 个氨基酸的转运肽将其从细胞质转运至线粒体。因此对候选蛋白的跨膜结构域、信号肽、线粒体转运肽和亚细胞定位进行分析，有利于进一步分析和确定候选蛋白是否参与 TCA 循环。

分析发现这 20 个候选蛋白均不含信号肽，除 MGG\_06547 外的 19 个候选蛋白均定位于线粒体中，且含线粒体转运肽的概率均高于 95% (MGG\_06547 仅为 2.62%)。表明 MGG\_06547 定位于细胞内线粒体外，不参与 TCA 循环。对跨膜结构域的预测结果显示：MGG\_04876 含有 3 个跨膜结构域，MGG\_00666 编码的蛋白虽不跨膜但其 C 端具有可结合在膜表面的结构域(表 2)，表明 MGG\_04876 和 MGG\_00666 可能作为琥珀酸脱氢酶的亚基定位于线粒体内膜将 TCA 循环产生的电子传递到呼吸链。

## 2.4 蛋白的结构域预测和稻瘟病菌 TCA 循环相关蛋白的确定

TCA 循环作为真核生物的基本代谢通路，不同物种间存在一定的保守性，因此对 TCA 循环相关蛋白结构域进行预测，并将稻瘟病菌和酿酒酵母中相应的同源蛋白进行比较有助于进一步对稻瘟病菌 TCA 循环相关蛋白的确定。其中对三个在酿酒酵母中未找到同源蛋白的 MGG\_06547、MGG\_04449 和 MGG\_15167 分析发现：MGG\_06547 具有和 IDH2 相同的 Iso\_dh 结构域；MGG\_04449 不具 KGD2 中包含的 Biotin\_lipoyl 和 2-oxoacid\_dh 结构域；酿酒酵母 LSC1 具有一个 CoA-binding 结构域和一个 Ligase\_CoA 结构域，在稻瘟病菌中 MGG\_05480 和 MGG\_15167 都具有相同的结构域，但 MGG\_15167 具两个 Ligase\_CoA 结构域。表明 MGG\_04449 不参与 TCA 循环，MGG\_15167 作为酿酒酵母中 LSC1 的另一个拷贝参与 TCA

循环。

表 2: 稻瘟病菌中 TCA 循环候选蛋白的在线分析

Table 2. *M.oryzae* in the TCA cycle on-line analysis of candidate genes

No.	<i>M.oryzae</i>	aa	PI	Mw	-	+	GRAVY	Instability index	Signal peptide	Transmembrane	Localization	Mitochondrial transit peptide	
1	MGG_07202	CIT1	472	8.86	52185.8	51	57	-0.244	30.57	no	no	M	0.9864
2	MGG_02617	CIT3	477	9.14	52306.8	47	55	-0.257	31.34	no	no	M	0.9977
3	MGG_03521	ACO1	784	6.57	85073.3	92	88	-0.36	35.4	no	no	M	0.9938
4	MGG_08584	ACO2	801	5.73	85998.5	88	76	-0.238	31.3	no	no	M	0.9996
5	MGG_01995	IDH1	386	9.09	42044.2	40	48	-0.199	35.72	no	no	M	0.9857
6	MGG_01712	IDH2	385	7.67	41373.6	44	45	-0.028	39.1	no	no	M	0.9953
7	MGG_06547	LYS12	360	5.47	39204.6	46	38	-0.124	26.46	no	no	-	0.0262
8	MGG_12767	KGD1	1008	6.53	114183	127	118	-0.524	41.04	no	no	M	0.9998
9	MGG_03149	KGD2	421	8.42	46558.5	60	62	-0.422	54.01	no	no	M	0.9913
10	MGG_04449	KGD2	523	8.96	56451.1	55	62	-0.156	44.13	no	no	M	0.9986
11	MGG_17072	LPD1	508	6.93	54151.9	58	57	-0.151	35.36	no	no	M	0.9957
12	MGG_05480	LSC1	333	9.04	34467.4	26	31	0.036	39.82	no	no	M	0.9647
13	MGG_15167	LSC1	733	6.89	78021.5	82	80	-0.15	41.02	no	no	M	0.9592
14	MGG_16230	LSC2	446	5.67	48410.3	59	62	-0.145	38.02	no	no	M	0.992
15	MGG_00168	SDH1	646	6.35	70486.3	76	69	-0.348	36.44	no	no	M	0.9992
16	MGG_00167	SDH2	274	9.18	31305.1	29	42	-0.479	66.62	no	no	M	0.9666
17	MGG_04876	SDH3	184	10.96	20155.4	7	18	0.205	33.56	no	yes	M	0.9954
18	MGG_00666	SDH4	171	10.35	18499.5	8	16	0.081	57.37	no	no	M	0.9895
19	MGG_17962	FUM1	546	8.38	58818.3	54	57	-0.177	36.1	no	no	M	0.9974
20	MGG_09367	MDH1	336	8.26	35191.3	35	37	-0.008	33.26	no	no	M	0.9723

注: PI: 理论等电点, Mw: 分子量; -: 负电荷残基总数 (Asp + Glu); +: 正电荷残基总数 (Arg + Lys); GRAVY: Grand average of hydropathicity, 总平均亲水性; Instability index: 不稳定性系数, 系数越高越不稳定; M: Mitochondrion, 线粒体。

综上所述, 利用生物信息学分析的方法从酿酒酵母 17 个 TCA 循环相关蛋白比对找到稻瘟病菌 20 个的候选蛋白。用在线生物信息学方法分析并排除了 MGG\_06547 和 MGG\_06547, 最终确定稻瘟病菌中 18 个 TCA 循环相关蛋白, 其中, LSC1 基因在稻瘟病菌中存在两个拷贝 (MGG\_05480 和 MGG\_15167) (图 1B)。

165

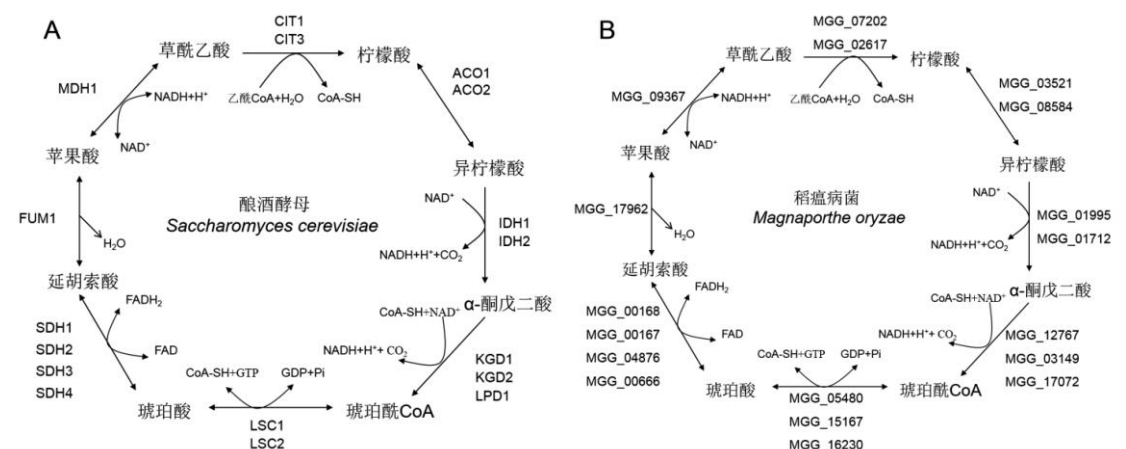
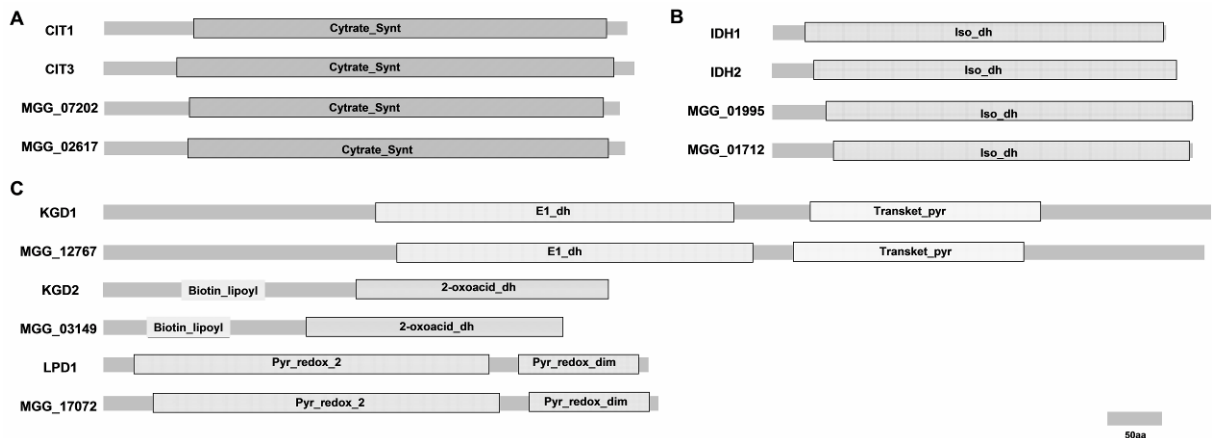


图 1: 酿酒酵母 (A) 和稻瘟病菌 (B) 中的 TCA 循环

Fig1. TCA cycle in *S.cerevisiae* (A) and *M.oryzae* (B).

170 此外，酿酒酵母 TCA 循环相关蛋白和稻瘟病菌中相应候选蛋白均具有相同的结构域。  
 以 TCA 循环中起关键作用的柠檬酸合酶、异柠檬酸脱氢酶和 $\alpha$ -酮戊二酸脱氢酶为例：酿酒  
 酵母中 CIT1、CIT3 和稻瘟病菌柠檬酸合酶的两个候选蛋白 MoCIT(MGG\_07202)、  
 MoMCS(MGG\_02617)都有相同的 Citrate\_Synt 柠檬酸合酶的催化结构域（图 2A）；酿酒酵  
 母中 IDH1、IDH2 和稻瘟病菌相应的两个异柠檬酸脱氢酶候选蛋白 MoIDH1(MGG\_01995)  
 175 和 MoIDH2(MGG\_01712)都有相同的 Iso\_dh 异柠檬酸脱氢酶催化结构域（图 2B）；酿酒酵  
 母中 KGD1、KGD2、LPD1 和稻瘟病菌中相应的 $\alpha$ -酮戊二酸脱氢酶复合体三个候选蛋白  
 MoKGD1(MGG\_12767)、MoKGD2(MGG\_03149)、MoLPD1(MGG\_17072) 都有相同的催化  
 结构域（图 2）。这些结果表明稻瘟病菌和酿酒酵母的 TCA 循环相关蛋白具有较高的保守性。



180

图 2: TCA 循环关键酶的结构域分析。

Fig2. Domain analysis of key enzymes in the TCA cycle

注：A：柠檬酸合酶（*S.cerevisiae*: CIT1, CIT2, CIT3; *M.oryzae*: MGG\_07202, MGG\_02617）的 Citrate\_Synt  
 催化结构域；B：异柠檬酸脱氢酶（*S.cerevisiae*: IDH1、DH2; *M.oryzae*: MGG\_01995, MGG\_01712）的 Iso\_dh  
 185 催化结构域；C： $\alpha$ -酮戊二酸脱氢酶复合体的三个亚基（*S.cerevisiae*: KGD1, KGD2, LPD1; *M.oryzae*:  
 MGG\_12767, MGG\_03149, MGG\_17072）及其相应催化结构域：KGD1, MGG\_12767 具有相同的 E1\_dh,  
 Transket\_pyr 结构域；KGD2, MGG\_03149 具有相同的 Biltin\_Lipoyl 和 2-oxoacid\_dh 结构域；LPD1,  
 MGG\_17072 具有相同的 Pyr\_redox\_2 和 Pyr\_redox\_dim 结构域。

### 2.5 TCA 循环关键酶的系统进化分析

190 TCA 循环作为生物体重要的代谢通路在不同物种间具一定保守性，但伴随着进化关系  
 的不同，物种间亦可能存在差异。因此，本研究构建成柠檬酸合酶、异柠檬酸脱氢酶和 $\alpha$ -  
 酮戊二酸脱氢酶复合体的系统发育树（图 3），并对不同物种 TCA 循环关键酶进行进化分  
 析。从图中可以看出，不同物种间 TCA 循环的关键酶具有较高的保守性，但和真菌相比，  
 动植物间的亲缘关系更近，并且在植物中存在更多的拷贝。

195 由图 3A 柠檬酸合酶的系统演化关系可以发现：CIT 在真菌、植物（水稻和拟南芥）和  
 动物（人和家鼠）内具有高度的同源性，同时在真菌内还具一个 MCS（CIT3 同源）的二甲  
 基柠檬酸合酶基因，其同样具有柠檬酸合酶的催化功能。MoCIT 和 MoMCS 与禾谷镰刀菌、  
 粗糙脉孢菌的同源性最高，其次是构巢曲霉和动植物。图 3B 显示：动物和植物中异柠檬酸  
 脱氢酶的演化关系较近，同时离真菌 IDH1 和 IDH2 较远。IDH1 和 IDH2 在六种真菌中都具  
 200 有较高的同源性，并且稻瘟病菌中异柠檬酸脱氢酶离禾谷镰刀菌、粗糙脉孢菌和构巢曲霉演  
 化关系较近，其次是白色念珠菌和酿酒酵母。由图 3C 可知， $\alpha$ -酮戊二酸脱氢酶复合体的三

个亚基在不同物种中都具有较高的同源性,但 AnLPD1 和 LPD1 与动植物有更近的演化关系, KGD1 和 KGD2 在分类系统内更具保守性。稻瘟病菌中 $\alpha$ -酮戊二酸脱氢酶离禾谷镰刀菌、粗糙脉孢菌和构巢曲霉较近, 其次是白色念珠菌和酿酒酵母。

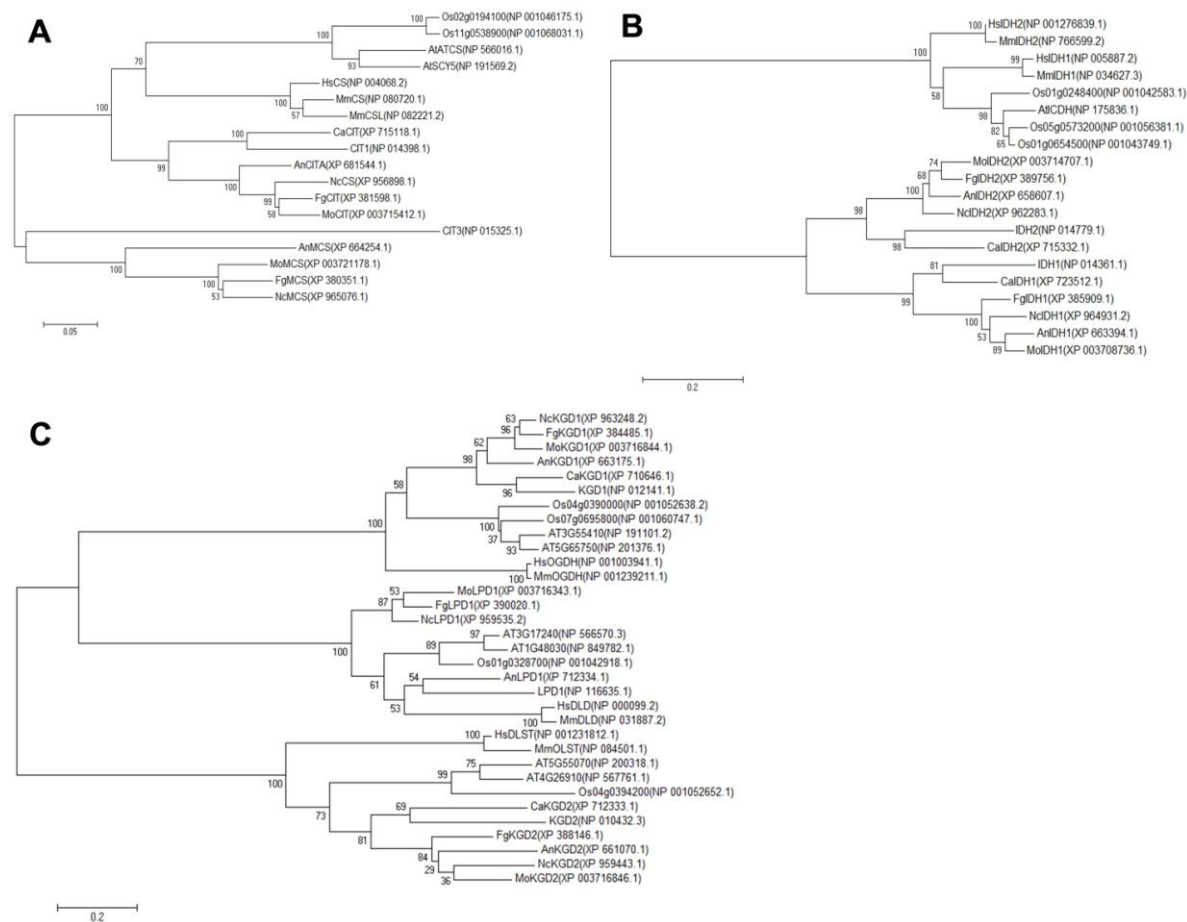


图 3: TCA 循环关键酶的系统演化关系 (括号内为 NCBI 登录号)

Fig3. Phylogenetic relationship of the key enzymes of the TCA cycle

A: 柠檬酸合酶; B: 异柠檬酸脱氢酶; C:  $\alpha$ -酮戊二酸脱氢酶复合体

注: Hs: *H.sapiens*(智人); Mm: *M.musculus*(家鼠); At: *A.thaliana* (拟南芥); Os: *O.sativa japonica* (水稻); An: *A.nidulans* (构巢曲霉); Ca: *C.albicans* (白色念珠菌); Fg: *F.graminearum*(禾谷镰刀菌); Mo: *M.oryzae* (稻瘟病菌); Nc: *N.crassa* (粗糙脉孢菌); CIT1, CIT2, CIT3: *S.cerevisiae* (酿酒酵母柠檬酸合酶); IDH1, IDH2: *S.cerevisiae* (酿酒酵母异柠檬酸脱氢酶); KGD1, KGD2, LPD1: *S.cerevisiae* (酿酒酵母 $\alpha$ -酮戊二酸脱氢酶复合体的三个亚基)。

### 2.6 基因表达量分析

由于基因的时空表达体现了蛋白在特异阶段的功能,所以本研究利用公共数据库分析了稻瘟病菌 TCA 循环相关基因在不同阶段的表达量, 其中 MGG\_17072、MGG\_16230、MGG\_17962 未检测到信号, 此外的 15 个基因在不同时期的表达量见图 4。如图所示, 在菌丝生长阶段和附着胞形成阶段 TCA 循环不同基因的表达量存在明显差异。在整个菌丝阶段和附着胞形成过程中 MGG\_03521、MGG\_00666、MGG\_01712、MGG\_00168、MGG\_07202 的表达量较高, MGG\_04876、MGG\_09367、MGG\_15167、MGG\_01995 的表达量较低。

由酿酒酵母和稻瘟病菌中 TCA 循环通路(图一)可以看出, 在整个 TCA 循环过程中, 每一步反应有一个或多个拷贝蛋白共同参与。从表达量数据中可以发现, 多个基因参与的同一路径中各个基因的表达量并不相同, 如柠檬酸合酶(MGG\_07202、MGG\_02617), 顺乌

225 头酸酶 (MGG\_03521、MGG\_08584)，异柠檬酸脱氢酶 (MGG\_01995、MGG\_01712)，琥珀酰辅酶 A 合成酶 (MGG\_05480、MGG\_15167)，琥珀酰辅酶脱氢酶 (MGG\_00168、MGG\_00167、MGG\_04876、MGG\_00666) 中不同拷贝的表达量存在明显差异。表明在 TCA 循环的代谢和调控过程中，多拷贝的基因中可能存在主效基因和辅助基因。其中高表达基因 MGG\_07202、MGG\_03521、MGG\_01712、MGG\_05480、MGG\_00168 和 MGG\_00666 在各自的反应中起主效作用；而表达量较低的基因 MGG\_02617、MGG\_08584、MGG\_01995、  
230 MGG\_15167 和 MGG\_04876 则起辅助作用。

TCA循环相关基因菌丝阶段和附着胞形成阶段的表达量

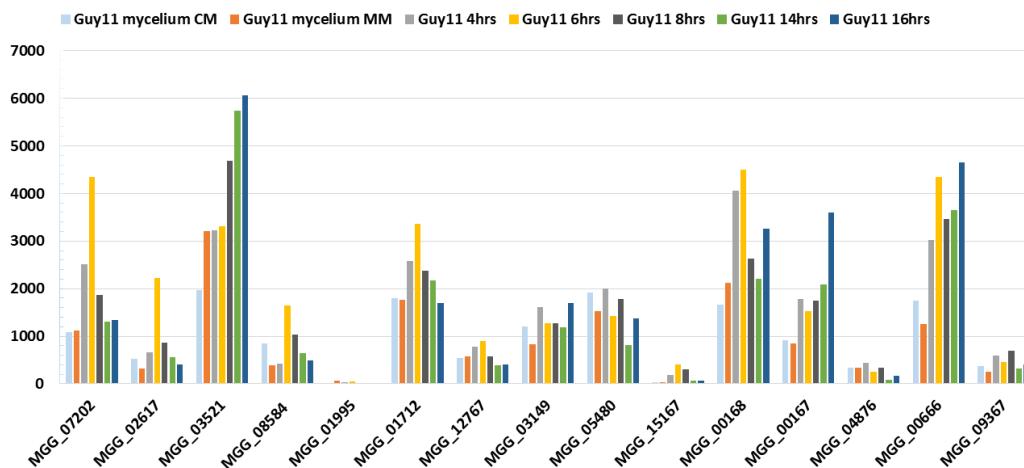


图 4: 稻瘟病菌 TCA 循环相关基因菌丝阶段和附着胞形成阶段的表达量

Fig4. The expression of *M.oryzae* TCA cycle genes

在菌丝阶段，同一基因在 CM 培养基中和 MM 培养基中总体差异不显著。MGG\_03521、  
235 MGG\_01995 和 MGG\_00168 基因在 MM 培养基中表达量上调；MGG\_08584、MGG\_03149 和 MGG\_00666 基因在 MM 培养基中表达量下调。同时 MGG\_07202、MGG\_03521、MGG\_01712、MGG\_03149、MGG\_05480、MGG\_00168、MGG\_00666 七个基因表达量较高 (>1000)。

在附着胞形成的不同阶段，TCA 循环不同基因的表达量有很大的变化：在孢子萌发和  
240 附着胞形成的早期(4h 到 6h)和菌丝阶段相比表达量皆有不同程度的上调，MGG\_07202、MGG\_02617、MGG\_01712、MGG\_00168、MGG\_00666 等的上调较为明显，表明附着胞形成早期可能需要 TCA 循环提供更多的中间代谢底物和能量；但随着附着胞的形成及成熟 (8-16 小时) 多数基因 (如：MGG\_07202、MGG\_02617、MGG\_08584、MGG\_01712 等) 表达量开始降低，但在附着胞成熟后期 (16h) MGG\_03521、MGG\_03149、MGG\_05480、  
245 MGG\_00168、MGG\_00167、MGG\_00666 表达量上调，MGG\_03521 在整个孢子萌发以及附着胞成熟阶段均表现为上调，表明这些基因可能在附着胞成熟后期对稻瘟病菌的侵染和致病过程起重要功能。

### 3 讨论

250 TCA 循环是需氧生物普遍存在的新陈代谢通路，该循环利用乙酰 CoA 和草酰乙酸经柠檬酸合酶产生柠檬酸，之后又经过七种 TCA 相关蛋白依次产生异柠檬酸、 $\alpha$ -酮戊二酸、琥珀酰 CoA、琥珀酸、延胡索酸和苹果酸再生成草酰乙酸，完成一个循环。TCA 循环相关蛋白通过物质之间的转换，最终将乙酰 CoA 彻底氧化分解成 CO<sub>2</sub> 和水，同时将产生的质子和

电子传递给辅酶  $\text{NAD}^+$  和  $\text{FAD}$  并连接呼吸链，最终生成生命活动所需要的能量物质  $\text{ATP}$ 。稻瘟病菌中乙酰 CoA 可以通过糖酵解代谢产物丙酮酸转换而成<sup>[17]</sup>，也可以通过脂肪的降解产生<sup>[5,6]</sup>。同时，乙酰 CoA 还可以通过肉毒咸-乙酰肉毒咸运输途径在细胞内跨膜转运<sup>[8-10]</sup>。此外，TCA 循环的中间产物还能参与氨基酸等进行转换。然而，线粒体 TCA 循环作为生物氧化的最终通路在稻瘟病菌中被报道的却很少。

本研究从酿酒酵母中 17 个 TCA 循环相关蛋白同源比对得到稻瘟病菌中 32 个候选蛋白，同时反向比对得到酵母中相应的同源基因。酵母同源比对得到的稻瘟病菌中首个候选基因均为酵母中参与 TCA 循环的同源基因，表明在酵母和稻瘟病菌中 TCA 循环具一定的保守性。同时也初步排除了 12 个可能不参与 TCA 循环的候选蛋白。

利用在线生物信息学分析软件，对 20 个 TCA 循环相关候选蛋白序列组成成分的理化性质、跨膜结构域、信号肽、线粒体转运肽以及亚细胞定位进行了预测和分析（表 3），结合结构域预测最终确定了稻瘟病菌中 18 个 TCA 循环相关蛋白，并绘制稻瘟病菌 TCA 循环图谱（图 1）。对比酿酒酵母中 TCA 循环相关蛋白（数据未显示），其与以上生物学的分析和预测基本一致。值得注意的是，琥珀酰辅酶 A 合成酶在酵母中有两个拷贝（LSC1、LSC2），而在稻瘟病菌中有三个拷贝（MGG\_05480、MGG\_15167、MGG\_16230），同时表达量分析数据显示 MGG\_05480 和 MGG\_15167 同时作为 LSC1 的同源蛋白，MGG\_15167 在稻瘟病菌中可能起辅助作用。

稻瘟病菌 TCA 候选蛋白均无信号肽，但是有线粒体转运肽，定位于线粒体，且多数不含跨膜结构域，但 MGG\_04876 (MoSDH3) 蛋白具三个跨膜结构域，MGG\_00666 (MoSDH4) 蛋白 C 端锚定在膜上，同时它们的理论等电点很高，疏水性氨基酸多于亲水性氨基酸，GRAVY 均为正值。表明 MGG\_04876 和 MGG\_00666 作为 SDH3 和 SDH4 的同源基因在稻瘟病菌中编码 MoSDH3 和 MoSDH4 为琥珀酸脱氢酶的亚基，定位于线粒体内膜将 TCA 循环产生的电子传递到呼吸链。

对 TCA 循环关键酶的结构域系统演化分析显示，稻瘟病菌和酿酒酵母中相应蛋白都具有相同的结构域（图 3），表明作为生物体主要的代谢通路在不同物种间具一定保守性，同时也存在一定的差异：稻瘟病菌 TCA 循环三种关键酶与禾谷镰刀菌、粗糙脉孢菌和构巢曲霉中相应的同源蛋白亲缘关系最近，而与白色念珠菌和酿酒酵母中相应同源蛋白的关系较近，与植物和动物的进化关系较远（图 4）；此外，TCA 循环关键酶的进化反映出动物界、植物界和真菌界的分化，说明真核生物 TCA 循环随物种进化而进化，生物学功能可能也随着物种亲缘关系的远近而有所差异，稻瘟病菌中 TCA 循环及各基因的功能尚需要进一步研究。

来自稻瘟病菌公共基因表达量数据库中的数据显示：MGG\_17072、MGG\_16230 和 MGG\_17962 未检测到信号。其中： $\alpha$ -酮戊二酸脱氢酶和琥珀酰 CoA 合成酶在稻瘟病菌中都有三个拷贝，MGG\_17072 和 MGG\_16230 分别作为其中的一个拷贝可能都起辅助作用一般情况不表达或表达量很低而未检测到。该数据库是由 Darren M. Soanes 用 RNA-Seq 和 High Throughput (HT)-SuperSAGE 技术获得的稻瘟病菌附着胞形态发生过程中 10591 个基因（含全基因组 96% 的基因）的转录组数据构建而成<sup>[17]</sup>。本研究 FUM1 在稻瘟病菌中同源基因 MGG\_17962 作为一个假定延胡索酸酶，存在于基因表达数据库中的 4% 未检测到的信号很可能是由实验设计和实验进程中的一些不足和错误导致。

此外，从图 4 中数据可以发现，TCA 循环相关基因在孢子萌发和附着胞形成的早期(4h 到 6h)与菌丝阶段相比表达量皆有不同程度的上调，可能与附着胞形成时需要更多的能量和

中间代谢底物有关, 同时也证明 TCA 循环可能参与稻瘟病菌附着胞的形成。但随着附着胞的形成及成熟(8-16 小时)多数基因表达量降低, 但在附着胞成熟后期(16h) MGG\_03521 (MoACO1)、MGG\_03149 (MoKGD2)、MGG\_05480 (MoLSC1)、MGG\_00168 (MoSDH1)、MGG\_00167 (MoSDH2)、MGG\_00666 (MoSDH4) 表达量上调(图 4), 表明这些基因可能通过调控代谢参与附着胞成熟和稻瘟病菌的致病性。总之, TCA 循环可能在整个稻瘟病菌的致病过程都起重要的作用。

病原与其寄主互作过程中表现出代谢通路的多样性和代谢调控途径的复杂性, 因而具有重要的研究价值。本研究采用不同的在线生物信息学分析软件对稻瘟病菌 TCA 循环相关基因进行了预测和分析, 对稻瘟病菌假定 TCA 循环相关蛋白的生物信息学分析为深入研究稻瘟病菌 TCA 循环相关基因及其功能奠定了基础, 同时也为今后进一步研究稻瘟病菌在侵染致病中的能量和物质代谢及其调控提供了参考。

### 305 [参考文献] (References)

- [1] Skamnioti P, Gurr SJ (2009) Against the grain: safeguarding rice from rice blast disease. *Trends in Biotechnology* 27: 141-150.
- [2] Wilson RA, Talbot NJ (2009) Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nature Reviews Microbiology* 7: 185-195.
- 310 [3] Howard RJ, Ferrari MA, Roach DH, Money NP (1991) Penetration of hard substrates by a fungus employing enormous turgor pressures. *Proceedings of the National Academy of Sciences of the United States of America* 88: 11281-11284.
- [4] Thines E, Weber RW, Talbot NJ (2000) MAP kinase and protein kinase A-dependent mobilization of triacylglycerol and glycogen during appressorium turgor generation by *Magnaporthe grisea*. *The Plant cell* 12: 1703-1718.
- 315 [5] Wang Z-Y, Soanes DM, Kershaw MJ, Talbot NJ (2007) Functional analysis of lipid metabolism in *Magnaporthe grisea* reveals a requirement for peroxisomal fatty acid beta-oxidation during appressorium-mediated plant infection. *Molecular Plant-Microbe Interactions* 20: 475-491.
- [6] Patkar RN, Ramos-Pamplona M, Gupta AP, Fan Y, Naqvi NI (2012) Mitochondrial beta-oxidation regulates organellar integrity and is necessary for conidial germination and invasive growth in *Magnaporthe oryzae*. *Molecular Microbiology* 86: 1345-1363.
- 320 [7] Badaruddin M, Holcombe LJ, Wilson RA, Wang Z-Y, Kershaw MJ, et al. (2013) Glycogen Metabolic Genes Are Involved in Trehalose-6-Phosphate Synthase-Mediated Regulation of Pathogenicity by the Rice Blast Fungus *Magnaporthe oryzae*. *Plos Pathogens* 9.
- [8] Bhambra GK, Wang ZY, Soanes DM, Wakley GE, Talbot NJ (2006) Peroxisomal carnitine acetyl transferase is required for elaboration of penetration hyphae during plant infection by *Magnaporthe grisea*. *Molecular Microbiology* 61: 46-60.
- 325 [9] Ramos-Pamplona M, Naqvi NI (2006) Host invasion during rice-blast disease requires carnitine-dependent transport of peroxisomal acetyl-CoA. *Molecular Microbiology* 61: 61-75.
- [10] Yang J, Kong L, Chen X, Wang D, Qi L, et al. (2012) A carnitine-acylcarnitine carrier protein, MoCrc1, is essential for pathogenicity in *Magnaporthe oryzae*. *Current Genetics* 58: 139-148.
- [11] Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *Journal of molecular biology* 157: 105-132.
- 335 [12] Ikeda M, Arai M, Lao DM, Shimizu T (2002) Transmembrane topology prediction methods: a re-assessment and improvement by a consensus method using a dataset of experimentally-characterized transmembrane topologies. *In silico biology* 2: 19-33.
- [13] Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 8: 785-786.
- 340 [14] Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of molecular biology* 300: 1005-1016.
- [15] Guda C, Guda P, Fahy E, Subramaniam S (2004) MITOPRED: a web server for the prediction of mitochondrial proteins. *Nucleic Acids Research* 32: W372-W374.
- [16] Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America* 95: 5857-5864.
- 345 [17] Soanes DM, Chakrabarti A, Paszkiewicz KH, Dawe AL, Talbot NJ (2012) Genome-wide Transcriptional Profiling of Appressorium Development by the Rice Blast Fungus *Magnaporthe oryzae*. *Plos Pathogens* 8.
- [18] Przybyla-Zawislak B, Gadde DM, Ducharme K, McCammon MT (1999) Genetic and biochemical interactions involving tricarboxylic acid cycle (TCA) function using a collection of mutants defective in all TCA cycle genes. *Genetics* 152: 153-166.
- 350