

胶孢炭疽菌 RGS 蛋白生物信息学分析

韩长志

(西南林业大学 林学院;云南省森林灾害预警与控制重点实验室,昆明 650224)

摘 要:胶孢炭疽菌侵染众多植物引起的炭疽病,给各国农林业生产造成了巨大经济损失. RGS 作为植物病原菌 G 蛋白信号转导过程中的重要调节因子,在众多生理生化过程中发挥着重要作用. 本研究基于前期研究所获得的胶孢炭疽菌中所含有的 12 个 RGS,通过保守结构域、理化性质、疏水性、细胞信号肽、跨膜区结构、亚细胞定位以及二级结构等生物信息学分析,明确上述 RGS 在保守结构域、理化性质、疏水性、信号肽、跨膜区域、二级结构等方面均具有较大的一致性特点;此外,通过对上述 RGS 进行遗传关系比较分析,发现不同菌株中的 RGS 彼此之间具有较近的亲缘关系. 该研究为深入开展胶孢炭疽菌 RGS 功能研究打下坚实的理论基础.

关键词:胶孢炭疽菌;G 蛋白信号调控因子;信号肽;二级结构

中图分类号:Q81

文献标志码:A

胶孢炭疽菌[*Colletotrichum gloeosporioides* (Penz.)Penz. et Sacc.]可以侵染多种植物引起炭疽病,目前已经报道的植物主要有桃、板栗、芒果、柿树、枇杷、苹果、核桃、橡胶、砀山梨等众多重要的经济林树种,以及蒜薹、番木瓜、葡萄、草莓、麦冬等^[1],给生产上带来了较大的经济损失. 前人对该菌的研究主要集中于形态特征、生活史以及遗传关系、致病基因等方面^[1],随着炭疽菌属中的禾谷炭疽菌(*Colletotrichum graminicola* (Cesati) Wilson)、希金斯炭疽菌(*Colletotrichum higginsianum* Sacc.)全基因组序列的公布^[2],为同属于该属病菌的胶孢炭疽菌的研究提供了重要的便利条件.

G 蛋白信号调控因子(Regulators of G-protein Signaling, RGS),所具有的主要功能在于通过促进 G 蛋白 α 亚基($G\alpha$)耦连的 GTP 水解,使 $G\alpha$ 和 $G\beta\gamma$ 亚基发生重新聚合,导致 G 蛋白失活而实现快速关闭上述信号途径,即对异三聚体 G 蛋白途径具有负调控作用^[3-4]. 基于此,学术界对真菌 RGS 的研究从其结构特征、定位情况以及功能等方面开展了大量工作^[5]. 然而,对于一些真菌中 RGS 的功能了解却十分有限^[5],现已明确真菌中的 RGS 多与菌丝生长发育过程中的有性生殖调控以及侵染过程中的次生代谢产物、色素合成等过程有关^[6].

前期利用 *S. cerevisiae* S288c 中已经报道的 4 个典型 RGS 氨基酸序列,通过在炭疽菌属蛋白质数据库中进行 Blastp 比对分析,以及通过关键词搜索,获得与酿酒酵母 RGS 同源的 *C. graminicola*,*C. higginsianum* 序列,明确 *C. graminicola*,*C. higginsianum* 中分别所存在 6 个、5 个 RGS,并根据其氨基酸大小,分别命名为 CgRGS1, CgRGS2, CgRGS3, CgRGS4, CgRGS5, CgRGS6, ChRGS1, ChRGS2, ChRGS3, ChRGS4, ChRGS5^[7-8]. 目前,对于 *C. gloeosporioides* RGS 生物信息学的分析尚未见报道,本研究基于上述发现的 RGS 氨基酸序列,在美国国家生物信息中心(NCBI)在线进行同源序列搜索,获得胶 *C. gloeosporioides* Nara gc5(缩写为 CgN),*C. gloeosporioides* Cg-14(缩写为 Cg14)中存在的 RGS 序列,利用生物信息学预测网站对其进行保守结构域分析、理化性质分析、疏水性分析、二级结构预测、信号肽分析以及跨膜区结构分析、亚细胞定位等分析,以期为进一步开展该菌 RGS 的后续研究提供重要的理论指导.

收稿日期:2014-07-20

基金项目:云南省优势特色重点学科生物学一级学科建设项目(50097505);云南省高校林下生物资源保护及利用科技创新团队(2014015);云南省教育厅科学研究基金项目(2014Y330).

作者简介:韩长志(1981-),男,河北石家庄人,西南林业大学讲师,博士,研究方向为经济林木病害生物防治与真菌分子生物学, E-mail: hanchangzhi2010@163.com.

1 材料与方法

1.1 材料

依据禾谷炭疽菌 RGS 的名称,对所获得的胶孢炭疽菌 RGS 分别进行命名(表 1).

表 1 胶孢炭疽菌 RGS 基本情况

胶孢炭疽菌	蛋白质登录号	禾谷炭疽菌 RGS 同源	希金斯炭疽菌 RGS 同源	氨基酸长度/aa	文中名称
CgN	ELA28178.1	CgRGS1	ChRGS2	533	CgNRGS1
Cg14	EQB53246.1	CgRGS1	ChRGS2	274	Cg14RGS1
CgN	ELA25443.1	CgRGS2	ChRGS1	367	CgNRGS2
Cg14	EQB54821.1	CgRGS2	ChRGS1	367	Cg14RGS2
CgN	ELA32294.1	CgRGS3	—	597	CgNRGS3
Cg14	EQB56965.1	CgRGS3	—	597	Cg14RGS3
CgN	ELA32529.1	CgRGS4	ChRGS3	736	CgNRGS4
Cg14	EQB54662.1	—	ChRGS3	437	Cg14RGS4
CgN	ELA31285.1	CgRGS5	ChRGS4	875	CgNRGS5
Cg14	EQB45833.1	CgRGS5	ChRGS4	873	Cg14RGS5
CgN	ELA34238.1	CgRGS6	ChRGS5	1224	CgNRGS6
Cg14	EQB57438.1	CgRGS6	ChRGS5	1224	Cg14RGS6

1.2 方法

1.2.1 保守结构域预测 利用 SMART 网站^[9]在线分析 RGS 所具有的保守结构域特征.

1.2.2 蛋白质理化性质及疏水性预测 利用蛋白质数据库^[10]在线进行分析预测 RGS 等电点、分子质量及氨基酸组成等特征.利用 ProtScale 程序^[10]对 RGS 进行疏水性测定.

1.2.3 蛋白质转运肽及信号肽预测 对蛋白质转运肽的预测利用 TargetP 1.1 Server 在线分析实现^[11],信号肽的预测则是利用 SignalP 3.0 Server^[12]在线分析实现.

1.2.4 蛋白质二级结构及跨膜区结构预测 对蛋白质二级结构预测采用 PHD^[13]在线分析实现.同时,对 RGS 的跨膜区结构预测,利用 TMHMM Server v. 2.0 实现^[12].

1.2.5 亚细胞定位分析 对 RGS 进行亚细胞定位分析,利用 ProtComp v9.0 实现^[14],以期获得蛋白质的定位情况.

1.2.6 系统进化树构建 利用 ClustalX^[15]对 CgN 以及 Cg14 中所含 RGS 氨基酸序列进行多重比对分析,随后利用 MEGA 5.2.2 软件^[16]构建系统进化树;采用邻近法构建系统发育树,各分支之间的距离计算采用 p-distance 模型,系统可信度检测采用自举法重复 1 000 次进行.

2 结果与分析

2.1 胶孢炭疽菌 12 个 RGS 均具有典型的保守域

利用 SMART 在线分析,结果显示,CgN 以及 Cg14 中所含的 12 条 RGS 序列均具有典型的 RGS 保守域结构,同时,还具有诸如 DEP,PAC,PXA,PX 等保守结构域(图 1).这与 *Aspergillus nidulans*,*Gibberella zeae*,*Cryptococcus neoformans*,*Magnaporthe oryzae* 中所含有 RGS 的结构相似^[6].另外,对上述 12 条 RGS 序列分别进行保守结构域对比分析,结果显示,除 CgNRGS4 和 Cg14RGS4 在所含 RGS 保守结构域及其位置方面存在一定差异外,其他 RGS 在保守结构域及其位置方面均存在着较大的一致性(图 1).

通过 TMHMM Server v. 2.0 分析,CgNRGS1,Cg14RGS1,CgNRGS4,Cg14RGS4,CgNRGS5,Cg14RGS5 均不具有典型的跨膜区结构,而 CgNRGS2,Cg14RGS2,CgNRGS3,Cg14RGS3,CgNRGS6,Cg14RGS6 均具有数量不同的跨膜结构,上述结果与 SMART 在线分析结果一致(图 1),具体而言,CgN-

RGS2, Cg14RGS2 位于 C 端具有 3 个跨膜区; CgNRGS3, Cg14RGS3 在位于 N 端具有 7 个跨膜区; CgNRGS6, Cg14RGS6 也在位于 N 端含有两个跨膜区(图 2).

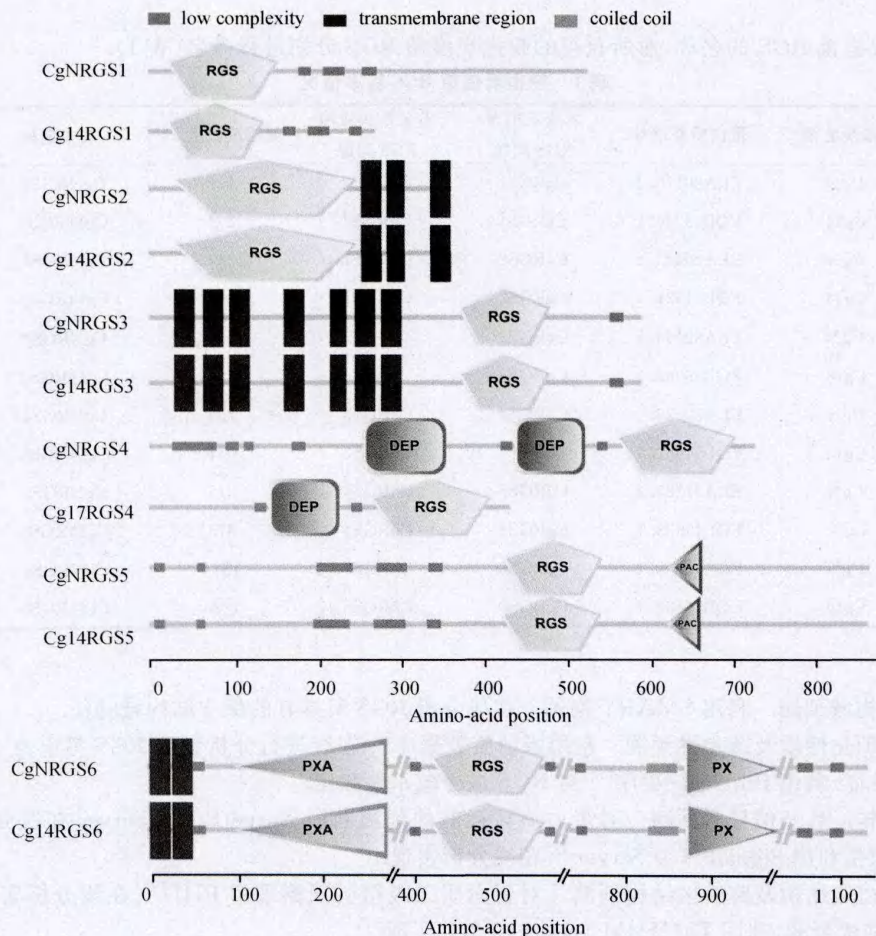


图1 胶孢炭疽菌 RGS 保守结构域分析

2.2 蛋白质理化性质及疏水性预测

对胶孢炭疽菌 RGS 理化性质进行分析,结果显示,尽管 CgN 以及 Cg14 中所含 RGS 彼此之间在相对分子质量、理论等电点、负电荷氨基酸残基数、正电荷氨基酸残基数、分子式、原子数量以及不稳定性系数、脂肪族氨基酸指数、总平均亲水性等方面均存在着差异,但是,就某一个 RGS 而言,两者又存在着较大的相似性(表 2)。比如, CgNRGS2 与 Cg14RGS2 相比,两者在理论等电点、负电荷氨基酸残基数、正电荷氨基酸残基数相同,原子数量、不稳定系数、脂肪族氨基酸指数方面也仅有细微差异。总体而言,胶孢炭疽菌 RGS 不稳定性系数均大于 40,属于不稳定蛋白;总平均亲水性(GRAVY)均小于 0,总体均为亲水性蛋白(表 2)。

对 CgNRGS1 和 Cg14RGS1, CgNRGS2 和 Cg14RGS2, CgNRGS3 和 Cg14RGS3, CgNRGS4 和 Cg14RGS4, CgNRGS5 和 Cg14RGS5, CgNRGS6 和 Cg14RGS6 分别进行对比分析,结果显示, CgNRGS1 和 Cg14RGS1 均含有较高比例的 Ser 和最低比例的 Cys, CgNRGS2 和 Cg14RGS2 均含有较高比例的 Leu 和最低比例 Cys, Trp, CgNRGS3 和 Cg14RGS3, CgNRGS6 和 Cg14RGS6 均含有较高比例的 Leu 和最低比例的 Cys, CgNRGS4 和 Cg14RGS4, CgNRGS5 和 Cg14RGS5 均含有较高比例的 Ser 和最低比例的 Trp(数据未显示)。

利用 ProtScale 程序,对胶孢炭疽菌 12 个 RGS 进行疏水性分析,结果显示,就亲水性最强氨基酸残基而言, CgNRGS2 和 Cg14RGS2, CgNRGS3 和 Cg14RGS3, CgNRGS6 和 Cg14RGS6 所含氨基酸相同,分别为

S, R/S/E, E; 就疏水性最强氨基酸残基而言, 除 CgNRGS4 和 Cg14RGS4 外, CgNRGS1 和 Cg14RGS1, CgNRGS2 和 Cg14RGS2, CgNRGS3 和 Cg14RGS3, CgNRGS5 和 Cg14RGS5, CgNRGS6 和 Cg14RGS6 均具有相同的氨基酸, 分别为 T, A, L, E, V. 对上述 12 个 RGS 的疏水性、亲水性数值进行统计分析, 结果显示, CgNRGS1, Cg14RGS1, CgNRGS2, Cg14RGS2, CgNRGS3, Cg14RGS3, CgNRGS4, Cg14RGS4, CgNRGS5, Cg14RGS5, CgNRGS6, Cg14RGS6 的亲水性氨基酸残基总和分别为 -356.930, -177.499, -155.662, -160.535, -265.742, -269.067, -456.330, -204.788, -647.540, -633.549, -663.479, -661.166, 而疏水性氨基酸残基数值总和则分别为 -28.878, 27.19, 127.033, 126.508, 206.190, 205.332, 45.736, 23.768, 44.799, 46.804, 196.082, 192.181, 上述结果表明, 除 CgNRGS4 和 Cg14RGS4 外, CgN 以及 Cg14 中对应的 RGS, 在“亲水性最强氨基酸残基位置、数值”、“疏水性最强氨基酸残基位置、数值”、“疏水性氨基酸残基数值总和”以及“亲水性氨基酸残基数值总和”等方面均具有较高的一致性。

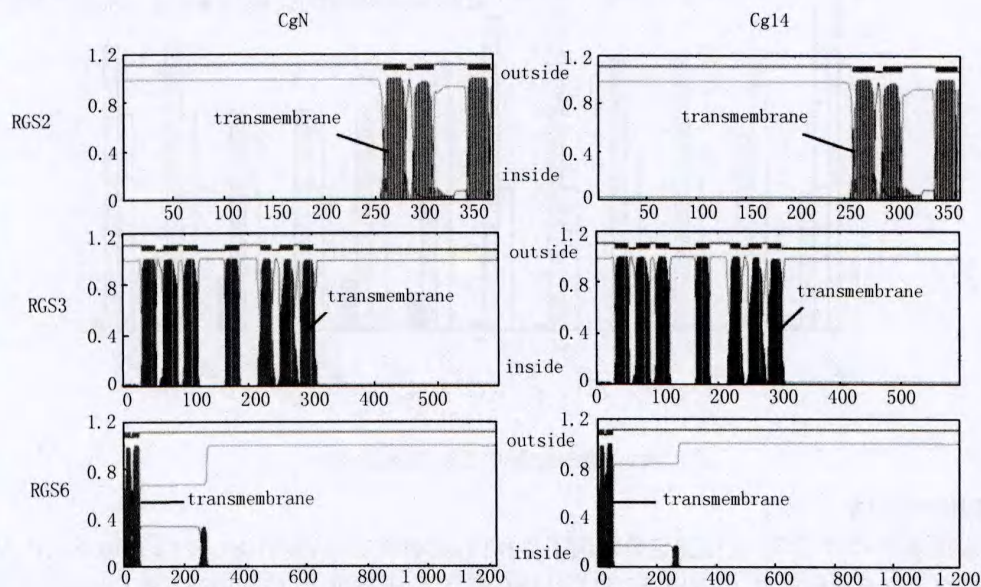


图 2 胶孢炭疽菌 RGS 跨膜结构域分析

表 2 胶孢炭疽菌 RGS 蛋白质基本理化性质

名称	相对分子质量/kDa	理论等电点	负电荷氨基酸残基数	正电荷氨基酸残基数	分子式	原子数	半衰期/h	不稳定性系数	脂肪族氨基酸指数	总平均亲水性
CgNRGS1	58.495	5.44	71	53	C ₂₅₃₂ H ₃₉₇₈ N ₇₃₄ O ₈₂₂ S ₂₀	8 086	30	59.2	67.02	-0.639
Cg14RGS1	29.742	5.29	34	26	C ₁₂₈₃ H ₂₀₁₅ N ₃₇₁ O ₄₂₂ S ₁₁	410 2	30	51.56	64.89	-0.58
CgNRGS2	41.306 4	6.11	43	40	C ₁₈₅₈ H ₂₉₀₅ N ₄₉₉ O ₅₃₇ S ₁₅	581 4	30	55.9	89.86	-0.072
Cg14RGS2	41.322 3	6.11	43	40	C ₁₈₅₈ H ₂₉₀₅ N ₄₉₉ O ₅₄₀ S ₁₄	581 6	30	56.61	89.32	-0.093
CgNRGS3	67.618 7	9.09	55	64	C ₃₀₈₂ H ₄₇₃₉ N ₈₁₉ O ₈₅₇ S ₂₀	951 7	30	40.74	85.63	-0.098
Cg14RGS3	67.619 6	9.01	55	63	C ₃₀₈₄ H ₄₇₃₀ N ₈₁₈ O ₈₅₇ S ₂₀	950 9	30	40.75	84.82	-0.105
CgNRGS4	80.994 3	9.09	76	87	C ₃₄₉₅ H ₅₅₈₀ N ₁₀₂₈ O ₁₁₄₁ S ₂₃	112 67	30	48.42	66.98	-0.59
Cg14RGS4	48.880 2	8.33	55	58	C ₂₁₂₁ H ₃₃₉₉ N ₆₁₃ O ₆₇₂ S ₂₀	682 5	30	39.87	75.88	-0.482
CgNRGS5	97.748 1	8.1	114	116	C ₄₂₂₇ H ₆₆₉₇ N ₁₂₈₃ O ₁₃₃₅ S ₂₈	135 70	30	50.51	68.47	-0.681
Cg14RGS5	97.608	7.43	115	115	C ₄₂₂₉ H ₆₆₉₇ N ₁₂₇₇ O ₁₃₃₂ S ₂₇	135 62	30	51.6	69.64	-0.664
CgNRGS6	137.220 9	6.22	170	162	C ₆₀₂₄ H ₉₆₉₁ N ₁₇₅₃ O ₁₈₃₉ S ₃₅	193 42	30	51.5	90.74	-0.366
Cg14RGS6	137.238 8	6.12	170	161	C ₆₀₁₉ H ₉₆₈₅ N ₁₇₅₁ O ₁₈₄₆ S ₃₅	193 36	30	51.62	90.58	-0.367

2.3 转运肽及信号肽特征

通过分析,胶孢炭疽菌中含有的 12 个 RGS 定位具有一致性特点,即 CgNRGS1 和 Cg14RGS1 均定位于线粒体,CgNRGS2 和 Cg14RGS2,CgNRGS4 和 Cg14RGS4,CgNRGS5 和 Cg14RGS5 均定于分泌途径上,CgNRGS3 和 Cg14RGS3,CgNRGS6 和 Cg14RGS6 均定位于分泌途径上。

经过 SignalP 3.0 分析,无论是经 NN 计算还是经 HMM 分析,胶孢炭疽菌中 12 个 RGS 均未发现有明显的信号肽序列(数据未显示)。

2.4 二级结构预测

二级结构预测显示,除 CgNRGS4 和 Cg14RGS4 外,CgN 以及 Cg14 中对应的 RGS,在所含的 α 螺旋、 β 转角以及无规卷曲方面所占的比例较为一致(图 3)。

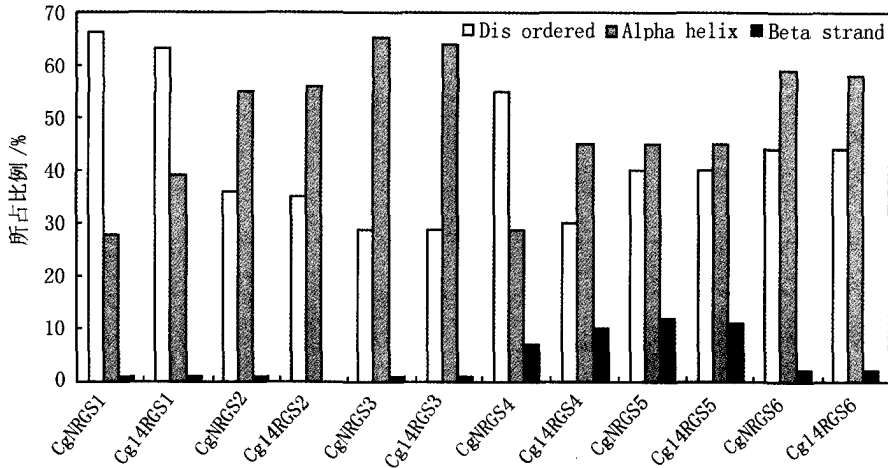


图 3 胶孢炭疽菌 RGS 二级结构特征分析

2.5 亚细胞定位分析

胶孢炭疽菌中 12 个 RGS 亚细胞定位情况不尽相同,具体而言,CgNRGS1 和 Cg14RGS1、CgNRGS5 和 Cg14RGS5 均定位于细胞核内,CgNRGS2 和 Cg14RGS2 定位于质膜上,CgNRGS3 和 Cg14RGS3 定位于内质网上,CgNRGS4 和 Cg14RGS4 定位于细胞质内,CgNRGS6 和 Cg14RGS6 定位于线粒体上(数据未显示)。

2.6 遗传关系

通过对 CgN 以及 Cg14 中所含的 12 个 RGS 序列进行聚类分析,结果显示,CgNRGS1 和 Cg14RGS1,CgNRGS2 和 Cg14RGS2,CgNRGS3 和 Cg14RGS3,CgNRGS4 和 Cg14RGS4,CgNRGS5 和 Cg14RGS5,CgNRGS6 和 Cg14RGS6 彼此之间亲缘关系较近,这与通过保守结构域、理化性质、疏水性、二级结构分析所得出的结果相一致;除去菌株方面的差异外,就每一个 RGS 聚类结果而言,RGS1 与 RGS3 亲缘关系较近,RGS2 与 RGS5 亲缘关系较近(图 4)。

3 结论与讨论

C. gloeosporioides 侵染多种经济林植物、观赏植物而引起炭疽病,给农林业生产上造成巨大的经济损失。目前,该菌对多菌灵、甲基托布津、苯菌灵等苯并咪唑类药物产生抗药性情况已经引起广泛关注,对其开展新的药剂筛选以及新的作用机制或靶标的药剂开发具有重要的理论意义。深入解析该病菌致病因子,有助于为开发防治该病害的药剂提供重要的理论意义。目前,已经明确有 MAPK,cAMP 以及 G 蛋白信号等三种保守途径,参与植物病原菌的上述过程,而 RGS 作为 G 蛋白信号转导过程中重要的调节因子,对其开展深入研究具有重要的意义。

本研究通过利用前期所获得的 *C. gloeosporioides* 12 个典型的 RGS 氨基酸序列,利用 SMART, Sig-

nalP, TMHMM 等生物信息学分析网站, 对其进行保守结构域、理化性质、疏水性、信号肽、跨膜区域、二级结构等分析, 结果表明 CgN 以及 Cg14 中所含有的 RGS 除在亚细胞定位外, 其他方面均具有较大的一致性; 此外, 通过对上述 12 个 RGS 进行遗传关系比较分析, 发现 CgNRGS1 和 Cg14RGS1, CgNRGS2 和 Cg14RGS2, CgNRGS3 和 Cg14RGS3, CgNRGS4 和 Cg14RGS4, CgNRGS5 和 Cg14RGS5, CgNRGS6 和 Cg14RGS6 彼此之间亲缘关系较近。

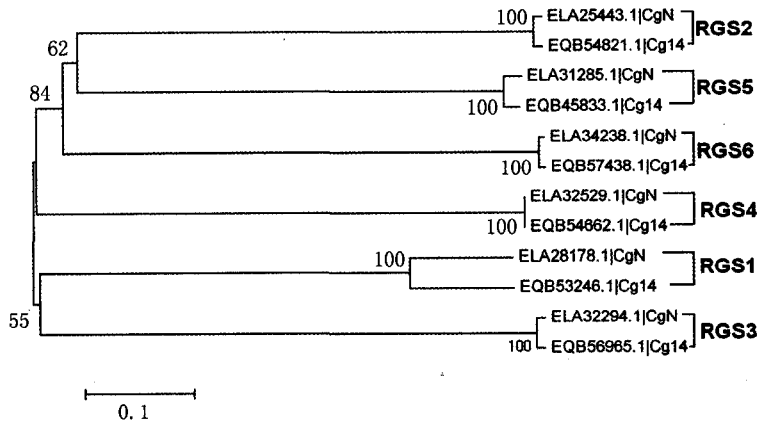


图 4 胶孢炭疽菌 RGS 遗传关系分析

目前, 与 *C. gloeosporioides* 同属于炭疽菌属中的 *C. graminicola* 与 *C. higginsianum* 均已经完成了基因组测序工作, 这对于开展炭疽属真菌致病基因的研究起到了重要推动作用。同时, 随着云南省大面积发展核桃经济林种植, 使得一些病虫害发生较为严重, 特别是由 *C. gloeosporioides* 引起的炭疽病, 给当地生产上造成巨大的经济损失, 深入开展该菌 RGS 等致病基因的研究具有非常重要的理论指导和生产实践意义。

参 考 文 献

- [1] 韩长志. 胶孢炭疽病菌的研究进展[J]. 华北农学报, 2012(S1): 386-389.
- [2] O'Connell R J, Thon M R, Hacquard S, et al. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses[J]. Nature Genetic, 2012, 44(9): 1060-1065.
- [3] Abramow-Newerly M, Roy A A, Nunn C, et al. RGS proteins have a signalling complex: interactions between RGS proteins and GPCRs, effectors, and auxiliary proteins[J]. Cell Signal, 2006, 18(5): 579-591.
- [4] McCudden C R, Hains M D, Kimple R J, et al. G-protein signaling: back to the future[J]. Cell Molecular Life Science, 2005, 62(5): 551-577.
- [5] 赵 勇, 王云川, 蒋德伟, 等. 真菌 G 蛋白信号调控蛋白的功能研究进展[J]. 微生物学通报, 2014, 41(4): 712-718.
- [6] Wang Y, Geng Z, Jiang D, et al. Characterizations and functions of regulator of G protein signaling (RGS) in fungi[J]. Applied Microbiology Biotechnology, 2013, 97(18): 7977-7987.
- [7] 韩长志. 希金斯炭疽菌 RGS 蛋白生物信息学分析[J]. 生物技术, 2014(1): 36-41.
- [8] 韩长志. 禾谷炭疽菌 RGS 蛋白生物信息学分析[J]. 微生物学通报, 2014, 41(8): 1582-1594.
- [9] Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource[J]. Nucleic acids research, 2012, 40(D1): D302-D305.
- [10] Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server[W]. The proteomics protocols handbook; Springer, 2005: 571-607.
- [11] Emanuelsson O, Brunak S, von Heijne G, et al. Locating proteins in the cell using TargetP, SignalP and related tools[J]. Nature protocols, 2007, 2(4): 953-971.
- [12] Bendtsen J D, Nielsen H, von Heijne G, et al. Improved prediction of signal peptides: SignalP 3.0[J]. Journal of molecular biology, 2004, 340(4): 783-795.
- [13] Kelley L A, Sternberg M J. Protein structure prediction on the Web; a case study using the Phyre server[J]. Nat Protoc, 2009, 4(3): 363-371.
- [14] Emanuelsson O, Nielsen H, Brunak S, et al. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence[J]. Journal of molecular biology, 2000, 300(4): 1005-1016.

- [15] Thompson J D, Gibson T J, Higgins D G. Multiple sequence alignment using ClustalW and ClustalX[J]. *Current Protocols in Bioinformatics*, 2002, 2: 2-3.
- [16] Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. *Molecular Biology Evolution*, 2011, 28(10): 2731-2739.

Bioinformatics Analysis on Regulators of G-protein Signaling in *Colletotrichum Gloeosporioides*

HAN Changzhi

(College of Forestry; The Key Laboratory of Forest Disaster Warning and Control of Yunnan Province, Southwest Forestry University, Kunming 650224, China)

Abstract: *Colletotrichum gloeosporioides* can infect many plants, which caused huge economic losses in agriculture and forestry production in many countries. As the important regulatory factor of plant pathogens G protein signal transduction RGS plays an important role in many physiological and biochemical processes. Based on the 12 typical RGS sequences have been reported in *C. gloeosporioides* bioinformatics analysis have been analyzed including conserved domain, the signal peptide, transmembrane domain structure and the secondary structure. In addition, analysis of genetic relationships through comparative 12 RGSs in *C. gloeosporioides* each RGS in *C. gloeosporioides* Cg-14 and *C. gloeosporioides* Nara gc5 has a high sequence homology and close genetic relationship. The study can provide strong theoretical foundation to research the function of RGS in *C. gloeosporioides*.

Keywords: *Colletotrichum gloeosporioides*; RGS; signal peptide; secondary structure