



棉花 NBS 类抗病基因类似物的生物信息学分析

Bio-informatics Analysis of the NBS Resistance

Gene Analogs in Cotton

孙涛^{1,2}, 卢美光², 简桂良², 林凤¹

(1. 沈阳农业大学生物科学技术学院, 沈阳 110161;

2. 中国农业科学院植物保护研究所, 植物病虫害生物学国家重点实验室, 北京 100094)

在长期的进化过程中,植物与病原物相互作用,协同进化,形成了复杂的植物抗病机制,现在普遍认为,在分子水平上植物的抗病性是由抗病基因(Resistance gene, R gene)介导的防御反应。目前已经从 11 种植物中克隆出 40 多个抗病基因,序列分析表明它们的产物在氨基酸水平上有相同特征的保守结构域。这为基于同源序列的抗病候选基因的克隆(Homology-based candidate gene method)提供了理论上的可能性。根据已克隆基因的保守区设计引物,利用 PCR 技术扩增植物基因组 DNA 或 cDNA,得到的扩增产物经测序和序列对比即可作为抗病基因类似物(Resistance gene analog, RGA)。与已克隆抗病基因同源性较高的 RGAs 可作为探针,经分子标记定位后进一步筛选抗病近等位基因系文库,可获得抗病基因候选序列,经转化鉴定即可获得目的基因。

1996 年 Leister 等首次利用此方法克隆抗病基因,随后,该方法在很多植物种都得到利用。近年来,国内外也发表了一系列棉花的抗病基因类似物。收集这些 RGAs 序列,并利用生物信息学的手段对这些序列进行分析,一方面是对过去工作的总结,另一方面也为这些 RGAs 序列的利用及棉花抗病基因的最终克隆提供可靠的数据和信息支持。

1 数据来源及方法

使用软件及程序:NCBI(National Center for Biotechnology Information, USA)提供的 ENTREZ 搜索程序,BLAST 程序;DNAMAN 等软件。

数据来源:本文涉及的 RGAs 核苷酸及其编

码的氨基酸序列,大部分来源于 NCBI (<http://www.ncbi.nlm.nih.gov>),一部分序列是方宣钧实验室 2003 年发表。利用 NCBI 的 Entrez 程序在数据库中搜索棉花抗病基因类似物,获取其登录号及核苷酸序列,以及这些 RGAs 编码的氨基酸序列。对提供的氨基酸序列利用 BLASTp 进行分析,寻找其保守区域。对 NBS 类型的 RGAs 编码的氨基酸序列利用 DNAMAN 进行多重对比分析,并利用 NCBI 中的 BLASTp 程序与已克隆的植物抗病基因进行同源性比对,分析棉花 RGAs 的结构特点。

2 棉花抗病基因类似物搜索结果

Entrez 的搜索结果,共发现棉花抗病基因类似序列 106 条,其中陆地棉 IM216 的 RGAs 4 条(Genbank 登录号 AY705379-AY705376);陆地棉 Auburn 634 RNR 的 RGAs 共 17 条(AY581719-AY581703);陆地棉 M-249 RGAs 33 条(AF469105-AF469073);海岛棉 pima 90 的 RGAs 51 条(AY331210-AY331191, AY244704-AY244674);陆地棉徐州 142 的 RGA 1 条(AY040533)。另有方宣钧实验室发表的海岛棉 7124 的 RGAs 11 条(Gbrga 2, 8, 39, 70, 72, 508, 522, 535, 540, 550, 568)。

3 棉花 RGAs 编码的氨基酸序列

从 117 条 RGAs 中获得氨基酸序列 74 条,经保守区分析,除 AY040533 编码的蛋白外其余均含有 NBS 类型蛋白特有的结构域(基序)。对含有这些基序的氨基酸序列(共 44 条)进行同源性分析和多重对比分析,发现这 44 条氨基酸的同源性都在 20% 以上,序列可分为三类,有两类属于

TIR-NBS, 第三类属于 non-TIR-NBS。海岛棉与陆地棉在这两个类型中均有分布, 说明棉花与其他的双子叶植物一样, 都存在 TIR-NBS 和 non-TIR-NBS 两种类型的抗病基因。从多重对比的结果可以看出, 这些 RGAs 编码的产物在五个区域具有保守结构, 即 NBS 序列中的 P-loop, Kinase-2, Kinase-3, 跨膜区“GLPL”以及在 Kinase-3 和“GLPL”之间的类 LZ(Leucine Zipper) 的保守区。在棉花 RGAs 中 P-loop 的共有序列可总结为 G(W/D/M)GGVG(K/E)TT(L/I), 大多数序列的 P-loop 结构为:GGGVGKTT, 同源性较强。Kinase-2 共有序列可总结为(V/L)(L/I/V)(L/M/I)D(D/G/W/N)(V/A/L/I)(D/W)。广泛存在的是 LLVLDDV。Kinase-3 的共有序列为(G/V/N)(S/Y/C)(k/R/T)aaa(T/I/A)(T/S)R, 最多的是 GSRaaaTTR, 中间

三个氨基酸多为 V,I,L 中的一个。跨膜区也表现出较强的保守性, 氨基酸序列大都为 GL-PLAL。

4 已克隆 NBS 类 R 基因与棉花 RGAs 所编码蛋白质的氨基酸同源对比

利用 Entrez 搜索已克隆的植物抗病基因所编码的蛋白质序列, 选取 NBS 类共 13 个 R 蛋白, 与棉花 RGAs 所编码的氨基酸序列进行同源对比(表 1), 同源性最强的是 AY244694 所编码的氨基酸序列与烟草抗 TMV 的 N 基因编码的氨基酸序列, 以及 Gbrga535 与番茄抗枯萎病的 I2C-1 基因所编码的氨基酸序列, 同源性达 50% 以上。

表 1 棉花 RGAs 产物与 NBS 类 R 蛋白氨基酸序列的同源性分析

Table 1 The homologous analysis of the amino acids in cotton RGAs to cloned NBS R gene

R 基因	来源	蛋白质登录号	病原菌	同源性最好的 RGA	同源性
N	烟草	A54810	tobacco mosaic virus	AY244694	53.2%
L6	亚麻	AAA91021	<i>Melampsora lini</i>	AY244674	37.3%
M	亚麻	T18548	<i>Melampsora lini</i>	AY244693,694,695	43.8%
I2C-1	番茄	AAB63274	<i>Fusarium oxysporum</i>	Gbrga535	53.3%
Mi-1.2	番茄	AA667238	<i>Meloidogyne incognita</i>	Gbrga508	40.0%
Xa1	水稻	T00020	<i>Xanthomonas oryzae pv.</i>	Gbrga535	38.8%
Pib	水稻	BAA93618	<i>Magna porthe grisea</i>	AF469105	30.3%
Prf	番茄	AAC49408	<i>Pseudomonas. pv. tomato</i>	AY244700	40.9%
RPM1	拟南芥	CAA61131	<i>Pseudomonas. pv. Maculicola</i>	AF469092, AF469101	30.0%
RPP1	拟南芥	AAC72977	<i>Peronospora parasitica</i>	AF469105	37.5%
RPS2	拟南芥	Q42484	<i>Pseudomonas syringae pv.</i>	AY244704	45.3%
P2	亚麻	AAK28806	<i>Melampsora lini</i>	AY244694	39.1%
RX	马铃薯	CAB50786	potato virus X	AY244700	36.4%

5 讨论

目前克隆的棉花 RGAs 序列同抗病基因均具有一定的同源性, 说明用此方法进行棉花抗病基因的克隆是可行的。但是, 迄今为止, 棉花 RGAs 的克隆还处在其利用的初级阶段, 即棉花的 RGAs 测序、定位以及利用其进行棉花抗病基因的系统发育研究。如要真正发挥其在抗病基因克隆方面的功能, 还需要在几个方面作出努力。在实验材料上, 要培育针对某一病害的高抗品系及近等位基因系。在引物设计方面, 要加强对已克

隆 R 基因以及棉花 RGAs 的研究, 设计比较精确的引物, 避免非特异性扩增。在试验技术方面, 目前扩增产物的检测大多采用 1% 的琼脂糖电泳, 但是根据前人的经验以及笔者自己的试验过程来看, 1% 的琼脂糖不能很好的显示扩增产物的多态性, 而采用 5% 的 PAGE 凝胶电泳可使多态性大大增强。同时, 棉花抗病基因的成功克隆, 还需要在抗病基因的分子标记, 棉花遗传图谱构建, ESTs(Expressed Sequence Tags)序列的克隆等方面取得进展。 ●