

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200510030527.X

[51] Int. Cl.

C07K 14/37 (2006.01)

C12N 15/31 (2006.01)

C12N 15/63 (2006.01)

C12P 21/02 (2006.01)

C07K 16/00 (2006.01)

G01N 33/53 (2006.01)

[43] 公开日 2007年4月18日

[11] 公开号 CN 1948335A

[22] 申请日 2005.10.14

[21] 申请号 200510030527.X

[71] 申请人 中国科学院上海生命科学研究院

地址 200031 上海市岳阳路 320 号

[72] 发明人 陈江野 毛旭明

[74] 专利代理机构 上海专利商标事务所有限公司

代理人 徐 迅

权利要求书 1 页 说明书 25 页 附图 5 页

[54] 发明名称

白念珠菌菌丝调控因子基因及其用途

[57] 摘要

本发明提供了一种新的菌丝调控因子 - CaSnf2 蛋白, 编码 CaSnf2 蛋白的多核苷酸和经重组技术产生这种 CaSnf2 蛋白的方法。本发明还公开了编码这种 CaSnf2 蛋白的多核苷酸的用途。CaSnf2 是白念珠菌菌丝生长所必需。免疫共沉淀实验证明 CaSnf2 因子和 CaSwi1 因子能够在白念珠菌体内相互作用, 这种相互作用不管在菌体还是菌丝条件下都能发生, 说明在白念珠菌中存在着 CaSwi/Snf 复合物调控白念珠菌的形态发生。小鼠系统感染实验表明, *casnf2/casnf2* 缺失株没有毒性, 这表明 CaSnf2 是白念珠菌中一个重要的毒性因子。

- 1.一种分离的白念珠菌 CaSnf2 多肽，其特征在于，该多肽选自下组：
 - (a) 具有 SEQ ID NO: 2 氨基酸序列的多肽；
 - (b) 将 SEQ ID NO: 2 氨基酸序列经过一个或多个氨基酸残基的取代、缺失或添加而形成的，并且具有调控白念珠菌菌丝形成的功能的由 (a) 衍生的多肽。
- 2.如权利要求1所述的多肽，其特征在于，该多肽是具有 SEQ ID NO: 2 氨基酸序列的多肽。
- 3.一种分离的多核苷酸，其特征在于，它包含一核苷酸序列，该核苷酸序列选自下组：
 - (a) 编码如权利要求1所述多肽的多核苷酸；
 - (b) 与多核苷酸(a)互补的多核苷酸。
- 4.如权利要求3所述的多核苷酸，其特征在于，该多核苷酸编码具有 SEQ ID NO: 2 所示氨基酸序列的多肽。
- 5.如权利要求3所述的多核苷酸，其特征在于，该多核苷酸的序列选自下组的一种：
 - (a) 具有 SEQ ID NO: 1 中 1-5070 位的序列；
 - (b) 具有 SEQ ID NO: 1 中 1-5073 位的序列。
- 6.一种载体，其特征在于，它含有权利要求3所述的多核苷酸。
- 7.一种遗传工程化的宿主细胞，其特征在于，它含有权利要求6所述的载体。
- 8.一种多肽的制备方法，其特征在于，该方法包含：
 - (a) 在适合表达的条件下，培养权利要求7所述的宿主细胞；
 - (b) 从培养物中分离出白念珠菌 CaSnf2 蛋白多肽。
- 9.一种能与权利要求1所述的白念珠菌 CaSnf2 蛋白特异性结合的抗体。
- 10.一种检测样品中是否存在 CaSnf2 蛋白的方法，其特征在于，包括：

将样品与权利要求9所述的抗体接触，

观察是否形成抗体复合物，形成了抗体复合物就表示样品中存在 CaSnf2 蛋白。

白念珠菌菌丝调控因子基因及其用途

技术领域

本发明属于生物技术领域，具体地说，本发明涉及新的编码白念珠菌菌丝调控因子 CaSnf2 的多核苷酸，以及此多核苷酸编码的多肽。本发明还涉及此多核苷酸和多肽的用途和制备。更具体地，本发明涉及从白念珠菌的基因组中克隆到白念珠菌菌丝生长和毒性相关因子基因 *CaSNF2* 及其用途。

背景技术

白念珠菌也称为白色念珠菌(*Candida albicans*)，它是一种临床上分离到的一种机会性人体致病真菌，在免疫下调的病人中，如器官移植病人，艾滋病毒感染者等，可以引起广泛的浅部和深部系统感染，感染部位包括口腔，女性阴道等，引起鹅口疮，阴道炎等疾病，也可以侵入表皮和内皮细胞进入血液到达内脏器官，如肾脏，脑部等，导致败血症，严重可以导致死亡(Odds, F.C. 1994. J Am Acad Dermatol. 31: S2-S5.)。

白念珠菌在不同的生长条件下，呈现不同的生长形态，包括菌体(yeast form)，假菌丝(pseudohyphae)和菌丝(hyphae)。各种形态之间的相互转化能力直接影响白念珠菌的致病能力(Odds, F.C. 1985. Crit Rev Microbiol. 12: 45-93; Brown, A.J.P. et al. 1999. Trends Microbiol. 7: 334-338.)，菌丝生长缺陷的菌株其系统感染能力下降或消失(Lo, H.J. et al. 1997. Cell. 90: 939-949; Braun, B.R. et al. 2001. EMBO J. 20: 4753-61; Hwang, C.S. et al. 2003. Mol Microbiol. 47: 1029-43.)。

许多培养条件可以引起白念珠菌的形态转换，包括血清，温度，pH 值，氮源利用以及氧压等等。细胞内调节白念珠菌形态转换的分子机制主要有 MAPK 途径(mitogen-activated protein kinase pathway)和 cAMP/PKA 途径(cAMP-dependent protein kinase A pathway)。同时还发现 Cph2 介导的，Efg1 介导的以及 pH 应答的信号途径也都和白念珠菌的形态发生相关(Lane, S. et al. 2001. Mol Cell Biol. 21: 6418-28; Stoldt, V.R., et al. 1997. EMBO J. 16: 1982-1991; El Barkani, A. et al. 2000. Mol Cell Biol. 20: 4635-4647.)。在白念珠菌中还存在抑制效应的信号途径，主要是 Tup1 介导的信号途径，依靠特异性 DNA 结合抑制因子 Rfg1, Nrg1 等来发挥作用(Kadosh, D. et al. 2001. Mol Cell Biol. 21: 2496-2505; Braun, B.R. et al. 2001. EMBO J. 20: 4753-4761.)。

由于白念珠菌会严重威胁人们的身体健康，因此，本领域迫切需要开发与白念珠菌的生长或毒性有关的各种蛋白或因子，以便更好地防治白念珠菌的引起的感染。

发明内容

本发明的目的是提供一种新的与白念珠菌生长有关的菌丝调控因子 CaSnf2 蛋白以

及其片段、类似物和衍生物。

本发明的另一目的是提供编码这些多肽的多核苷酸。

本发明的另一目的是提供生产这些多肽的方法以及该多肽和编码序列的用途。

在本发明的第一方面，提供新颖的分离出的 CaSnf2 多肽，它包括：具有 SEQ ID NO: 2 氨基酸序列的多肽、或其保守性变异多肽、或其活性片段、或其活性衍生物。

较佳地，该多肽选自下组：

(a) 具有 SEQ ID NO: 2 氨基酸序列的多肽；

(b) 将 SEQ ID NO: 2 氨基酸序列经过一个或多个（(如 1-100 个，较佳地 1-50 个，更佳地 1-20 个，最佳地 1-10 个)）氨基酸残基的取代、缺失或添加而形成的，并且具有调控白念珠菌菌丝形成的功能的由 (a) 衍生的多肽。

更佳地，该多肽是具有 SEQ ID NO: 2 氨基酸序列的多肽。

在本发明的第二方面，提供编码分离的这些多肽的多核苷酸，该多核苷酸包含一核苷酸序列，该核苷酸序列与选自下组的一种核苷酸序列有至少 70% 相同性：(a) 编码上述白念珠菌 CaSnf2 多肽的多核苷酸；和 (b) 与多核苷酸 (a) 互补的多核苷酸。较佳地，该多核苷酸编码具有 SEQ ID NO: 2 所示氨基酸序列的多肽。更佳地，该多核苷酸的序列是选自下组的一种：(a) 具有 SEQ ID NO: 1 中 1-5070 位的序列；(b) 具有 SEQ ID NO: 1 中 1-5073 位的序列。

在本发明的第三方面，提供了含有上述多核苷酸的载体，以及被该载体转化或转导的宿主细胞或者被上述多核苷酸直接转化或转导的宿主细胞。

在本发明的第四方面，提供了制备具有白念珠菌 CaSnf2 蛋白活性的多肽的方法，该方法包含：(a) 在适合表达白念珠菌 CaSnf2 蛋白的条件下，培养上述被转化或转导的宿主细胞；(b) 从培养物中分离出具有白念珠菌 CaSnf2 蛋白活性的多肽。

在本发明的第五方面，提供了与上述的白念珠菌 CaSnf2 多肽特异性结合的抗体。

在本发明的第六方面，提供了模拟、促进、拮抗白念珠菌 CaSnf2 多肽活性的化合物，以及抑制白念珠菌 CaSnf2 多肽的表达的化合物。还提供了筛选和/或制备这些化合物的方法。较佳地，该化合物是白念珠菌 CaSnf2 多肽的编码序列或其片段的反义序列。

在本发明的第七方面，提供了检测样品中是否存在 CaSnf2 蛋白的方法，它包括：将样品与 CaSnf2 蛋白的特异性抗体接触，观察是否形成抗体复合物，形成了抗体复合物就表示样品中存在 CaSnf2 蛋白。

在本发明的第八方面，提供了一种检测与白念珠菌 CaSnf2 多肽异常表达相关的疾病或疾病易感性的方法，该方法包括：检测编码所述多肽的核酸序列中是否存在突变。

在本发明的第九方面，提供了本发明多肽和编码序列的用途。本发明多肽可被用于筛选促进白念珠菌 CaSnf2 多肽活性的激动剂，或者筛选抑制白念珠菌 CaSnf2 多肽活性的拮抗剂、或者被用于肽指纹图谱鉴定。本发明的白念珠菌 CaSnf2 蛋白的编码序列或其片段，可被作为引物用于 PCR 扩增反应，或者作为探针用于杂交反应，或者用于制

造基因芯片或微阵列。在一个优选例中，白念珠菌 *CaSnf2* 多肽或基因被用于制备检测白念珠菌的试剂，或者筛选抑制 *CaSnf2* 表达或活性的抑制剂或拮抗剂，或作为筛选抑制 *CaSnf2* 表达或活性的靶点。

在本发明的第十方面，提供了一种药物组合物，它含有安全有效量(如 0.001-99.99wt%)的本发明的白念珠菌 *CaSnf2* 多肽的拮抗剂或抑制剂或抗体以及药学上可接受的载体。这些药物组合物可治疗白念珠菌感染等病症。

本发明的其它方面由于本文的技术的公开，对本领域的技术人员而言是显而易见的。

附图说明

下列附图用于说明本发明的具体实施方案，而不用于限定由权利要求书所界定的本发明范围。

图 1A 显示了白念珠菌 *CaSNF2* 基因开放阅读框的脱氧核糖核苷酸序列。

图 1B 显示了白念珠菌 *CaSnf2* 蛋白质的氨基酸序列。

图 1C 显示了白念珠菌 *CaSnf2* 蛋白因子和酿酒酵母 *ScSnf2* 蛋白因子结构域比较。两者都还有相似的 DEXDc, HELICc 和 BROMO 结构域。

图 2A 显示了在白念珠菌中敲除 *CaSNF2* 基因的策略及其染色体上的酶切图谱。

图 2B 显示了在白念珠菌 *CaSNF2* 基因的敲除过程中各个菌株的 Southern 杂交图谱。所用探针为 *CaSNF2* 编码框 5'端的 1kb 片段。

图 3 显示了白念珠菌 *casnf2/casnf2* 缺失株在菌丝生长条件下形成假菌丝。培养条件为液体 YPD 加 10%血清和 Lee's 培养基在 37°C 诱导以及包埋在 YPS 中室温培养。

图 4 显示了白念珠菌 *casnf2/casnf2* 缺失株在菌体生长条件下都形成假菌丝。培养条件为液体和固体 YPD 培养基，以及固体 SLAD 培养基，温度 30°C。液体 YPD 中生长的各个菌株细胞用染料 Calcofluor White 染色显示几丁质的分布。

图 5 显示了 *CaSnf2* 因子和 *CaSwi1* 因子能够在白念珠菌体内相互作用。培养条件为液体 YPD, 25°C, 液体 YPD 加 10%血清和 Lee's 培养基在 37°C 诱导。

图 6 显示了在小鼠系统感染实验中，*CaSNF2* 基因的敲除导致白念珠菌毒性消失。注射液浓度为 5×10^7 细胞/毫升，每只小鼠尾静脉注射 100 μ l 菌液，共注射 8 只小鼠。

具体实施方式

本发明人通过广泛而深入的研究，首次在白念珠菌中克隆了一种调控白念珠菌菌丝生长的基因 *CaSnf2*。具体地，本发明人利用同源重组原理，在白念珠菌中敲除 *CaSNF2*，构建了 *casnf2/casnf2* 缺失株，该缺失株不能形成菌丝，这说明 *CaSnf2* 是白念珠菌菌丝生长所必需。此外，不管在菌体还是菌丝培养条件下，白念珠菌 *casnf2/casnf2* 缺失株都以假菌丝形态生长。免疫共沉淀实验证明 *CaSnf2* 因子和 *CaSwi1* 因子能够在白念珠菌体内相互作用，这种相互作用不管在菌体还是菌丝条件下都能发生，说明在白念珠菌

中存在着 CaSwi/Snf 复合物调控白念珠菌的形态发生。小鼠系统感染实验表明 *casnf2/casnf2* 缺失突变株没有毒性，这说明 CaSnf2 是白念珠菌中一个重要的毒性因子。在此基础上完成了本发明。

在本发明中，术语“CaSnf2 蛋白”、“CaSnf2 多肽”或“菌丝调控因子 CaSnf2”可互换使用，都指具有白念珠菌菌丝调控因子 CaSnf2 氨基酸序列 (SEQ ID NO:2) 的蛋白或多肽。它们包括含有或不含起始甲硫氨酸的菌丝调控因子 CaSnf2。

如本文所用，“分离的”是指物质从其原始环境中分离出来(如果是天然物质，原始环境即是天然环境)。如活体细胞内的天然状态下的多聚核苷酸和多肽是没有分离纯化的，但同样的多聚核苷酸或多肽如从天然状态中同存在的其他物质中分开，则为分离纯化的。

如本文所用，“分离的 CaSnf2 蛋白或多肽”是指 CaSnf2 多肽基本上不含天然与其相关的其它蛋白、脂类、糖类或其它物质。本领域的技术人员能用标准的蛋白质纯化技术纯化 CaSnf2 蛋白。基本上纯的多肽在非还原聚丙烯酰胺凝胶上能产生单一的主带。

本发明的多肽可以是重组多肽、天然多肽、合成多肽，优选重组多肽。本发明的多肽可以是天然纯化的产物，或是化学合成的产物，或使用重组技术从原核或真核宿主(例如，细菌、酵母、高等植物、昆虫和哺乳动物细胞)中产生。根据重组生产方案所用的宿主，本发明的多肽可以是糖基化的，或可以是非糖基化的。本发明的多肽还可包括或不包括起始的甲硫氨酸残基。

本发明还包括白念珠菌 CaSnf2 蛋白的片段、衍生物和类似物。如本文所用，术语“片段”、“衍生物”和“类似物”是指基本上保持本发明的天然白念珠菌 CaSnf2 蛋白相同的生物学功能或活性的多肽。本发明的多肽片段、衍生物或类似物可以是(i)有一个或多个保守或非保守性氨基酸残基(优选保守性氨基酸残基)被取代的多肽，而这样的取代的氨基酸残基可以是也可以不是由遗传密码编码的，或(ii)在一个或多个氨基酸残基中具有取代基团的多肽，或(iii)成熟多肽与另一个化合物(比如延长多肽半衰期的化合物，例如聚乙二醇)融合所形成的多肽，或(iv)附加的氨基酸序列融合到此多肽序列而形成的多肽(如前导序列或分泌序列或用来纯化此多肽的序列或蛋白原序列，或与抗原 IgG 片段的形成的融合蛋白)。根据本文的教导，这些片段、衍生物和类似物属于本领域熟练技术人员公知的范围。

在本发明中，术语“白念珠菌 CaSnf2 多肽”指具有白念珠菌 CaSnf2 蛋白活性的 SEQ ID NO: 2 序列的多肽。该术语还包括具有与白念珠菌 CaSnf2 蛋白相同功能的、SEQ ID NO: 2 序列的变异形式。这些变异形式包括(但并不限于)：一个或多个(通常为 1-50 个，较佳地 1-30 个，更佳地 1-20 个，最佳地 1-10 个)氨基酸的缺失、插入和/或取代，以及在 C 末端和/或 N 末端添加一个或数个(通常为 20 个以内，较佳地为 10 个以内，更佳地为 5 个以内)氨基酸。例如，在本领域中，用性能相近或相似的氨基酸进行取代时，通常不会改变蛋白质的功能。又比如，在 C 末端和/或 N 末端添加一个或数个氨基酸通

常也不会改变蛋白质的功能。该术语还包括白念珠菌 CaSnf2 蛋白的活性片段和活性衍生物。该术语还包括与 SEQ ID NO: 2 所示的氨基酸序列具有至少 70%，更佳地至少 80%，更佳地至少 90%，最佳地至少 95% 序列相同性，并且具有调控白念珠菌菌丝形成的功能的多肽。

该多肽的变异形式包括：同源序列、保守性变异体、等位变异体、天然突变体、诱导突变体、在高或低的严紧度条件下能与白念珠菌 CaSnf2 DNA 杂交的 DNA 所编码的蛋白、以及利用抗白念珠菌 CaSnf2 多肽的抗血清获得的多肽或蛋白。本发明还提供了其他多肽，如包含白念珠菌 CaSnf2 多肽或其片段的融合蛋白。除了几乎全长的多肽外，本发明还包括了白念珠菌 CaSnf2 多肽的可溶性片段。通常，该片段具有白念珠菌 CaSnf2 多肽序列的至少约 10 个连续氨基酸，通常至少约 30 个连续氨基酸，更佳地至少约 50 个连续氨基酸，更佳地至少约 80 个连续氨基酸，最佳地至少约 100 个连续氨基酸。

本发明还提供白念珠菌 CaSnf2 蛋白或多肽的类似物。这些类似物与天然白念珠菌 CaSnf2 多肽的差别可以是氨基酸序列上的差异，也可以是不影响序列的修饰形式上的差异，或者兼而有之。这些多肽包括天然或诱导的遗传变异体。诱导变异体可以通过各种技术得到，如通过辐射或暴露于诱变剂而产生随机诱变，还可通过定点诱变法或其他已知分子生物学的技术。类似物还包括具有不同于天然 L-氨基酸的残基(如 D-氨基酸)的类似物，以及具有非天然存在的或合成的氨基酸(如 β 、 γ -氨基酸)的类似物。应理解，本发明的多肽并不限于上述例举的代表性的多肽。

修饰(通常不改变一级结构)形式包括：体内或体外的多肽的化学衍生形式如乙酰化或羧基化。修饰还包括糖基化，如那些在多肽的合成和加工中或进一步加工步骤中进行糖基化修饰而产生的多肽。这种修饰可以通过将多肽暴露于进行糖基化的酶(如哺乳动物的糖基化酶或去糖基化酶)而完成。修饰形式还包括具有磷酸化氨基酸残基(如磷酸酪氨酸，磷酸丝氨酸，磷酸苏氨酸)的序列。还包括被修饰从而提高了其抗蛋白水解性能或优化了溶解性能的多肽。

在本发明中，“白念珠菌 CaSnf2 蛋白保守性变异多肽”指与 SEQ ID NO: 2 的氨基酸序列相比，有至多 10 个，更佳地至多 8 个，更佳地至多 5 个，最佳地至多 3 个氨基酸被性质相似或相近的氨基酸所替换而形成多肽。这些保守性变异多肽最好根据表 1 进行氨基酸替换而产生。

表 1

最初的残基	代表性的取代	优选的取代
Ala (A)	Val; Leu; Ile	Val
Arg (R)	Lys; Gln; Asn	Lys
Asn (N)	Gln; His; Lys; Arg	Gln
Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro; Ala	Ala

His (H)	Asn; Gln; Lys; Arg	Arg
Ile (I)	Leu; Val; Met; Ala; Phe	Leu
Leu (L)	Ile; Val; Met; Ala; Phe	Ile
Lys (K)	Arg; Gln; Asn	Arg
Met (M)	Leu; Phe; Ile	Leu
Phe (F)	Leu; Val; Ile; Ala; Tyr	Leu
Pro (P)	Ala	Ala
Ser (S)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr; Phe	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser	Phe
Val (V)	Ile; Leu; Met; Phe; Ala	Leu

本发明的多核苷酸可以是 DNA 形式或 RNA 形式。DNA 形式包括 cDNA、基因组 DNA 或人工合成的 DNA。DNA 可以是单链的或是双链的。DNA 可以是编码链或非编码链。编码成熟多肽的编码区序列可以与 SEQ ID NO:1 所示的编码区序列相同或者是简并的变异体。如本文所用,“简并的变异体”在本发明中是指编码具有 SEQ ID NO:2 的蛋白质,但与 SEQ ID NO:1 所示的编码区序列有差别的核酸序列。

编码 SEQ ID NO:2 的成熟多肽的多核苷酸包括:只编码成熟多肽的编码序列;成熟多肽的编码序列和各种附加编码序列;成熟多肽的编码序列(和任选的附加编码序列)以及非编码序列。

术语“编码多肽的多核苷酸”可以是包括编码此多肽的多核苷酸,也可以是还包括附加编码和/或非编码序列的多核苷酸。

本发明还涉及上述多核苷酸的变异体,其编码与本发明有相同的氨基酸序列的多肽或多肽的片段、类似物和衍生物。此多核苷酸的变异体可以是天然发生的等位变异体或非天然发生的变异体。这些核苷酸变异体包括取代变异体、缺失变异体和插入变异体。如本领域所知的,等位变异体是一个多核苷酸的替换形式,它可能是一个或多个核苷酸的取代、缺失或插入,但不会从实质上改变其编码的多肽的功能。

本发明还涉及与上述的序列杂交且两个序列之间具有至少 50%,较佳地至少 70%,更佳地至少 80%相同性的多核苷酸。本发明特别涉及在严格条件下与本发明所述多核苷酸可杂交的多核苷酸。在本发明中,“严格条件”是指:(1)在较低离子强度和较高温度下的杂交和洗脱,如 0.2×SSC, 0.1%SDS, 60°C;或(2)杂交时加有变性剂,如 50%(v/v) 甲酰胺, 0.1%小牛血清/0.1% Ficoll, 42°C等;或(3)仅在两条序列之间的相同性至少在 90%以上,更好是 95%以上时才发生杂交。并且,可杂交的多核苷酸编码的多肽与 SEQ ID NO:2 所示的成熟多肽有相同的生物学功能和活性。

本发明还涉及与上述的序列杂交的核酸片段。如本文所用,“核酸片段”的长度至少含 15 个核苷酸,较好是至少 30 个核苷酸,更好是至少 50 个核苷酸,最好是至少 100 个核苷酸以上。核酸片段可用于核酸的扩增技术(如 PCR)以确定和/或分离编码 CaSnf2 蛋白的多聚核苷酸。

本发明中的多肽和多核苷酸优选以分离的形式提供，更佳地被纯化至均质。

本发明的白念珠菌 CaSnf2 核苷酸全长序列或其片段通常可以用 PCR 扩增法、重组法或人工合成的方法获得。对于 PCR 扩增法，可根据本发明所公开的有关核苷酸序列，尤其是开放阅读框序列来设计引物，并用市售的 cDNA 库或按本领域技术人员已知的常规方法所制备的 cDNA 库作为模板，扩增而得有关序列。当序列较长时，常常需要进行两次或多次 PCR 扩增，然后再将各次扩增出的片段按正确次序拼接在一起。

一旦获得了有关的序列，就可以用重组法来大批量地获得有关序列。这通常是将其克隆入载体，再转入细胞，然后通过常规方法从增殖后的宿主细胞中分离得到有关序列。

此外，还可用人工合成的方法来合成有关序列，尤其是片段长度较短时。通常，通过先合成多个小片段，然后再进行连接可获得序列很长的片段。

目前，已经可以完全通过化学合成来得到编码本发明的 CaSnf2 蛋白(或其片段，或其衍生物)的 DNA 序列。然后可将该 DNA 序列引入本领域中已知的各种现有的 DNA 分子(或如载体)和细胞中。此外，还可通过化学合成将突变引入本发明的 CaSnf2 蛋白序列中。

应用 PCR 技术扩增 DNA/RNA 的方法(Saiki, et al. Science 1985;230: 1350-1354)被优选用于获得本发明的基因。特别是很难从文库中得到全长的 cDNA 时，可优选使用 RACE 法(RACE-cDNA 末端快速扩增法)，用于 PCR 的引物可根据本文所公开的本发明的序列信息适当地选择，并可用常规方法合成。可用常规方法如通过凝胶电泳分离和纯化扩增的 DNA/RNA 片段。

本发明也涉及包含本发明的多核苷酸的载体，以及用本发明的载体或 CaSnf2 蛋白编码序列经基因工程产生的宿主细胞，以及经重组技术产生本发明所述多肽的方法。

通过常规的重组 DNA 技术(Science, 1984; 224: 1431)，可利用本发明的多聚核苷酸序列可用来表达或生产重组的 CaSnf2 多肽。一般来说有以下步骤：

(1). 用本发明的编码白念珠菌 CaSnf2 多肽的多核苷酸(或变体)，或用含有该多核苷酸的重组表达载体转化或转导合适的宿主细胞；

(2). 在合适的培养基中培养的宿主细胞；

(3). 从培养基或细胞中分离、纯化蛋白质。

本发明中，白念珠菌 CaSnf2 多核苷酸序列可插入到重组表达载体中。术语“重组表达载体”指本领域熟知的细菌质粒、噬菌体、酵母质粒、植物细胞病毒、哺乳动物细胞病毒如腺病毒、逆转录病毒或其他载体。在本发明中适用的载体包括但不限于：在细菌中表达的基于 T7 的表达载体(Rosenberg, et al. Gene, 1987, 56: 125)；在哺乳动物细胞中表达的 pMSXND 表达载体(Lee and Nathans, J Bio Chem. 263: 3521, 1988)和在昆虫细胞中表达的来源于杆状病毒的载体。总之，只要能在宿主体内复制和稳定，任何质粒和载体都可以用。表达载体的一个重要特征是通常含有复制起点、启动子、标记基因和翻译控制元件。

本领域的技术人员熟知的方法能用于构建含白念珠菌 CaSnf2 编码 DNA 序列和合适的转录/翻译控制信号的表达载体。这些方法包括体外重组 DNA 技术、DNA 合成技术、体内重组技术等(Sambrook, et al. *Molecular Cloning, a Laboratory Manual*, Cold Spring Harbor Laboratory, New York, 1989)。所述的 DNA 序列可有效连接到表达载体中的适当启动子上,以指导 mRNA 合成。这些启动子的代表性例子有:大肠杆菌的 lac 或 trp 启动子;λ 噬菌体 PL 启动子;真核启动子包括 CMV 立即早期启动子、HSV 胸苷激酶启动子、早期和晚期 SV40 启动子、反转录病毒的 LTRs 和其他一些已知的可控制基因在原核或真核细胞或其病毒中表达的启动子。表达载体还包括翻译起始用的核糖体结合位点和转录终止子。

此外,表达载体优选地包含一个或多个选择性标记基因,以提供用于选择转化的宿主细胞的表型性状,如真核细胞培养用的二氢叶酸还原酶、新霉素抗性以及绿色荧光蛋白(GFP),或用于大肠杆菌的四环素或氨苄青霉素抗性。

包含上述的适当 DNA 序列以及适当启动子或者控制序列的载体,可以用于转化适当的宿主细胞,以使其能够表达蛋白质。

宿主细胞可以是原核细胞,如细菌细胞;或是低等真核细胞,如酵母细胞;或是高等真核细胞,如哺乳动物细胞。代表性例子有:大肠杆菌,链霉菌属;鼠伤寒沙门氏菌的细菌细胞;真菌细胞如酵母;植物细胞;果蝇 S2 或 Sf9 的昆虫细胞;CHO、COS、293 细胞、或 Bowes 黑素瘤细胞的动物细胞等。

本发明的多核苷酸在高等真核细胞中表达时,如果在载体中插入增强子序列时将会使转录得到增强。增强子是 DNA 的顺式作用因子,通常大约有 10 到 300 个碱基对,作用于启动子以增强基因的转录。可举的例子包括在复制起始点晚期一侧的 100 到 270 个碱基对的 SV40 增强子、在复制起始点晚期一侧的多瘤增强子以及腺病毒增强子等。

本领域一般技术人员都清楚如何选择适当的载体、启动子、增强子和宿主细胞。

用重组 DNA 转化宿主细胞可用本领域技术人员熟知的常规技术进行。当宿主为原核生物如大肠杆菌时,能吸收 DNA 的感受态细胞可在指数生长期后收获,用 CaCl₂ 法处理,所用的步骤在本领域众所周知。另一种方法是使用 MgCl₂。如果需要,转化也可用电穿孔的方法进行。当宿主是真核生物,可选用如下的 DNA 转染方法:磷酸钙共沉淀法,常规机械方法如显微注射、电穿孔、脂质体包装等。

获得的转化子可以用常规方法培养,表达本发明的基因所编码的多肽。根据所用的宿主细胞,培养中所用的培养基可选自各种常规培养基。在适于宿主细胞生长的条件下进行培养。当宿主细胞生长到适当的细胞密度后,用合适的方法(如温度转换或化学诱导)诱导选择的启动子,将细胞再培养一段时间。

在上面的方法中的重组多肽可在细胞内、或在细胞膜上表达、或分泌到细胞外。如果需要,可利用其物理的、化学的和其它特性通过各种分离方法分离和纯化重组的蛋白。这些方法是本领域技术人员所熟知的。这些方法的例子包括但并不限于:常规的复性处理、用蛋白沉淀剂处理(盐析方法)、离心、渗透破菌、超处理、超离心、分子筛层

析(凝胶过滤)、吸附层析、离子交换层析、高效液相层析(HPLC)和其它各种液相层析技术及这些方法的结合。

重组的白念珠菌 CaSnf2 蛋白或多肽有多方面的用途。这些用途包括(但不限于):用于筛选促进或对抗 CaSnf2 蛋白功能的抗体、多肽或其它配体。用表达的重组白念珠菌 CaSnf2 蛋白筛选多肽库可用于寻找有治疗价值的能抑制白念珠菌 CaSnf2 蛋白功能的多肽分子。

另一方面,本发明还包括对白念珠菌 CaSnf2 DNA 或是其片段编码的多肽具有特异性的多克隆抗体和单克隆抗体,尤其是单克隆抗体。“特异性”是指抗体能结合于白念珠菌 CaSnf2 基因产物或片段。较佳地,指那些能与白念珠菌 CaSnf2 基因产物或片段结合但不识别和结合于其它非相关抗原分子的抗体。本发明中抗体包括那些能够结合并抑制白念珠菌 CaSnf2 蛋白的分子,也包括那些并不影响白念珠菌 CaSnf2 蛋白功能的抗体。本发明还包括那些能与修饰或未经修饰形式的白念珠菌 CaSnf2 基因产物结合的抗体。

本发明不仅包括完整的单克隆或多克隆抗体,而且还包括具有免疫活性的抗体片段,如 Fab'或(Fab)₂片段;抗体重链;抗体轻链;遗传工程改造的单链 Fv 分子(Ladner 等人,美国专利 No. 4,946,778);或嵌合抗体,如具有鼠抗体结合特异性但仍保留来自人的抗体部分的抗体。

本发明的抗体可以通过本领域内技术人员已知的各种技术进行制备。例如,纯化的白念珠菌 CaSnf2 基因产物或者其具有抗原性的片段,可被施用于动物以诱导多克隆抗体的产生。与之相似的,表达白念珠菌 CaSnf2 蛋白或其具有抗原性的片段的细胞可用于免疫动物来生产抗体。本发明的抗体也可以是单克隆抗体。此类单克隆抗体可以利用杂交瘤技术来制备(见 Kohler 等人, Nature 256:495, 1975; Kohler 等人, Eur. J. Immunol. 6:511, 1976; Kohler 等人, Eur. J. Immunol. 6:292, 1976; Hammerling 等人, In Monoclonal Antibodies and T Cell Hybridomas, Elsevier, N.Y., 1981)。本发明的抗体包括能阻断白念珠菌 CaSnf2 蛋白功能的抗体以及不影响白念珠菌 CaSnf2 蛋白功能的抗体。本发明的各类抗体可以利用白念珠菌 CaSnf2 基因产物的片段或功能区,通过常规免疫技术获得。这些片段或功能区可以利用重组方法制备或利用多肽合成仪合成。与白念珠菌 CaSnf2 基因产物的未修饰形式结合的抗体可以用原核细胞(例如 *E. Coli*)中生产的基因产物来免疫动物而产生;与翻译后修饰形式结合的抗体(如糖基化或磷酸化的蛋白或多肽),可用真核细胞(例如酵母或昆虫细胞)中产生的基因产物来免疫动物而获得。

抗白念珠菌 CaSnf2 蛋白的抗体可用于免疫组织化学技术中,检测活检标本中的白念珠菌 CaSnf2 蛋白。

本发明中的抗体可用于治疗或预防与白念珠菌 CaSnf2 蛋白相关的疾病。给予适当剂量的抗体可以刺激或阻断白念珠菌 CaSnf2 蛋白的产生或活性。

抗体也可用于设计成针对体内某一特殊部位的免疫毒素。如白念珠菌 CaSnf2 蛋白高亲和性的单克隆抗体可与细菌或植物毒素(如白喉毒素,蓖麻蛋白,红豆碱等)共价结

合。一种通常的方法是用巯基交联剂如 SPDP，攻击抗体的氨基，通过二硫键的交换，将毒素结合于抗体上，这种杂交抗体可用于杀灭表达 CaSnf2 蛋白的白念珠菌。

多克隆抗体的生产可用白念珠菌 CaSnf2 蛋白或多肽免疫动物，如家兔，小鼠，大鼠等。多种佐剂可用于增强免疫反应，包括但不限于弗氏佐剂等。

利用本发明的 CaSnf2 蛋白，通过各种常规筛选方法，可筛选出与 CaSnf2 蛋白发生相互作用的物质，如抗体、抑制剂、激动剂或拮抗剂等。

本发明的 CaSnf2 蛋白的抗体、抑制剂、或拮抗剂等，当在治疗上进行施用(给药)时，可提供不同的效果。通常，可将这些物质配制于无毒的、惰性的和药学上可接受的水性载体介质中，其中 pH 通常约为 5-8，较佳地 pH 约为 6-8，尽管 pH 值可随被配制物质的性质以及待治疗的病症而有所变化。配制好的药物组合物可以通过常规途径进行给药，其中包括(但并不限于)：口腔内、阴道内、肌内、腹膜内、静脉内、皮下、皮内、或局部给药。

例如，本发明的 CaSnf2 蛋白的抗体、抑制剂、或拮抗剂可直接用于疾病治疗，例如，用于抑制白念珠菌的正常生长，进而减轻白念珠菌造成的感染。在使用本发明 CaSnf2 蛋白时，还可同时使用其他治疗剂，如其他抗真菌剂氟康唑等。

本发明还提供了一种药物组合物，它含有安全有效量的抗 CaSnf2 多肽的抗体或其拮抗剂以及药学上可接受的载体或赋形剂。这类载体包括(但并不限于)：盐水、缓冲液、葡萄糖、水、甘油、乙醇、及其组合。药物制剂应与给药方式相匹配。本发明的药物组合物可以被制成针剂形式，例如用生理盐水或含有葡萄糖和其他辅剂的水溶液通过常规方法进行制备。诸如片剂和胶囊之类的药物组合物，可通过常规方法进行制备。药物组合物如针剂、溶液、片剂和胶囊宜在无菌条件下制造。活性成分的给药量是治疗有效量，例如每天约 1 微克/千克体重-约 5 毫克/千克体重。此外，本发明的多肽还可与其他治疗剂一起使用。

使用药物组合物时，是将安全有效量的 CaSnf2 蛋白的拮抗剂或抗体施用于哺乳动物，其中该安全有效量通常至少约 10 微克/千克体重，而且在大多数情况下不超过约 8 毫克/千克体重，较佳地该剂量是约 10 微克/千克体重-约 1 毫克/千克体重。当然，具体剂量还应考虑给药途径、病人健康状况等因素，这些都是熟练医师技能范围内的。

能与白念珠菌 CaSnf2 蛋白结合的多肽分子可通过筛选由各种可能组合的氨基酸结合于固相物组成的随机多肽库而获得。筛选时，宜对白念珠菌 CaSnf2 蛋白分子进行标记。

本发明还涉及定量和定位检测白念珠菌 CaSnf2 蛋白水平的诊断试验方法。这些试验是本领域所熟知的，且包括 FISH 测定和放射免疫测定。试验中所检测的白念珠菌 CaSnf2 蛋白水平，可以用作判断白念珠菌是否会造成严重的感染。

一种检测检测样品中是否存在 CaSnf2 蛋白的方法是利用 CaSnf2 蛋白的特异性抗体进行检测，它包括：将样品与 CaSnf2 蛋白特异性抗体接触；观察是否形成抗体复合物，形成了抗体复合物就表示样品中存在 CaSnf2 蛋白。

CaSnf2 蛋白的多聚核苷酸可用于 CaSnf2 蛋白相关疾病的诊断。在诊断方面，CaSnf2 蛋白的多聚核苷酸可用于检测 CaSnf2 蛋白的表达与否。如 CaSnf2 DNA 序列可用于对活检标本的杂交以判断 CaSnf2 蛋白的表达异常。杂交技术包括 Southern 印迹法，Northern 印迹法、原位杂交等。这些技术方法都是公开的成熟技术，相关的试剂盒都可从商业途径得到。本发明的多核苷酸的一部分或全部可作为探针固定在微阵列 (microarray) 或 DNA 芯片 (又称为“基因芯片”) 上，用于分析组织中基因的差异表达分析和基因诊断。用 CaSnf2 蛋白特异的引物进行 RNA-聚合酶链反应 (RT-PCR) 体外扩增也可检测 CaSnf2 蛋白的转录产物。

检测 CaSnf2 基因的突变也可用于诊断 CaSnf2 蛋白相关的疾病。CaSnf2 蛋白突变的形式包括与正常野生型 CaSnf2 DNA 序列相比的点突变、易位、缺失、重组和其它任何异常等。可用已有的技术如 Southern 印迹法、DNA 序列分析、PCR 和原位杂交检测突变。另外，突变有可能影响蛋白的表达，因此用 Northern 印迹法、Western 印迹法可间接判断基因有无突变。

在本发明的一个实例中，提供了一种分离的多核苷酸，它编码具有 SEQ ID NO: 2 所示氨基酸序列的多肽。本发明的多核苷酸是从白念珠菌 cDNA 文库中分离出的。其序列如 SEQ ID NO:1 所示，它包含的多核苷酸序列全长为 5073 个碱基，其开放读框位于 1-5070 位，编码全长为 1690 个氨基酸的白念珠菌 CaSnf2 蛋白 (SEQ ID NO: 2)。CaSnf2 蛋白为治疗白念珠菌感染提供新的治疗途径，具有潜在的应用前景。

下面结合具体实施例，进一步阐述本发明。应理解，这些实施例仅用于说明本发明而并不用于限制本发明的范围。下列实施例中未注明具体条件的实验方法，通常按照常规条件如 Sambrook 等人，分子克隆：实验室手册 (New York: Cold Spring Harbor Laboratory Press, 1989) 中所述的条件，或按照制造厂商所建议的条件。

实施例

材料

限制性内切酶、T4 DNA 连接酶、探针标记试剂盒均购自美国 Invitrogen 公司；消解酶 (Zymolyase 100T) 购自日本 Seikagaku 公司；酸洗玻璃珠 (425~600 μ m)，Calcoflour White 荧光染料，FLAG M2 抗体，蛋白酶抑制剂均购自 Sigma 公司，proteinG-bead 悬液购自 Roche 公司，GFP 抗体购自 Santa Cruz 公司，ECL 试剂购自 Pierce 公司，用于 PCR 扩增的 KOD plus 购自日本的 Toyobo 公司。

通用方法：

(1) 白念珠菌基因组 DNA 抽提

白念珠菌培养过夜用 ddH₂O 洗 1 次，悬浮在 500 μ l 溶液 A 中 (1M 山梨醇, 100mM

EDTA pH8.0), 加入 5 μ l 20mg/ml 的消解酶(Zymolase), 37 $^{\circ}$ C 放置 1 小时后, 高速离心去上清, 细胞用 500 μ l TE 缓冲液(20mM Tris-HCl pH7.5, 1mM EDTA)洗一次, 悬浮在 350 μ l TE 缓冲液中, 并加入 90 μ l 溶液 B(250mM EDTA pH8.0, 400mM Tris-HCl pH8.0, 2% SDS), 65 $^{\circ}$ C 放置 30 分钟, 加入 80 μ l 5M KAc, 冰上放置 1 小时, 高速离心 5 分钟, 吸取上清并加入 1ml 的无水乙醇, -20 $^{\circ}$ C 放置 20 分钟, 高速离心 5 分钟, 沉淀用 70%乙醇洗一次, 离心后烘干。

(2) Southern 分析

白念珠菌基因组 DNA 抽提后, 基因组 DNA 用 HindIII 完全酶切, 电泳完成后的琼脂糖胶分别用变性液和中和液浸泡 45 min, 尼龙膜用双蒸水或 10xSSC 浸透, 搭好转膜平台, 胶与膜间用 Parafilm 封闭, 加一个 500g 砝码。以 10xSSC 为转膜液转移过夜, 第二天漂洗晾干交联。交联后的膜装入杂接管, 加入 10ml 预杂交液(6xSSC, 5xdenhardt's Reagent, 0.5%SDS, 100 μ g/ml 鱼精 DNA 于 42 $^{\circ}$ C 预杂交 12h, 探针在 100 $^{\circ}$ C 变性 5min, 加入 150 μ l(约 1 / 3 所标记的探例)42 $^{\circ}$ C 杂交 10-16h。0.1xSSC, 0.1%SDS 洗两到三次, 每次 40min, 取出膜, 晾干, 室温或-70 $^{\circ}$ C 压片。探针标记用 Invitrogene 公司的随机引物标记试剂盒, 照其说明操作。将含 25ng DNA 的溶液于 100 $^{\circ}$ C 加热变性 5-10min, 置于冰浴中, 再依次加入随机引物缓冲液混合物(Random Primers Buffer Mixture), 2 μ l dCTP, 2 μ l dGTP, 2 μ l dTTP, 3 μ l [α - 32 P]-dATP (10 μ Ci/ μ l), 混匀后加入 1 μ l Klenow 酶, 于 25 $^{\circ}$ C 保温 1h, 以 Sephadex G-25 柱以 3000rpm 离心 4min, 收集离心液, 即为纯化后的探针溶液。探针于 100 $^{\circ}$ C 变性 5min 冰上冷却后加入杂交体系中。

(3) 白念珠菌的转化

准备 PEG / LiAc 溶液(pH 7.5)10ml; 在 1.5ml Eppendorf 管中依次加入质粒 5 μ g, 10 μ l 10mg/ml 鱼精 DNA, 混匀; 加入 0.1ml 感受态细胞和 0.6ml PEG / LiAc, vortex 混匀; 30 $^{\circ}$ C 200rpm 培养 30min; 加入 70 μ l DMSO, 轻轻混匀; 42 $^{\circ}$ C 热冲击 15min, 期间不时轻轻摇匀; 冰浴 2min; 高速离心 15s, 吸掉上清, 用 0.2ml TE 重悬细胞; 涂布于营养筛选 SD 平板上。

(4) 白念珠菌 Calcoflour White 荧光染料的染色

白念珠菌各个菌株在 YPD 里培养过夜, 转接到新鲜的 YPD 培养基中培养 5 小时, 收集细胞, 用 ddH₂O 洗 3 次, 悬浮在 70%的乙醇中, 室温振摇 1 小时, 再用 ddH₂O 洗 3 次, 悬浮在 1 μ g/ml 的 Calcoflour White 荧光染料(购自 Sigma 公司)的溶液中, 室温振摇 15 分钟, 用 ddH₂O 洗 3 次, 在荧光显微镜下观察。

(5) 白念珠菌总蛋白抽提, 免疫沉淀和 Western Blot 检测

白念珠菌菌株在 YPD 中培养至 OD=1, 收集细胞, 用 400 μ l 酵母细胞裂解液(20 mM

Tris-HCl pH 7.5, 1 mM EDTA, 150 mM NaCl, 0.5% NP-40)洗涤 3 次后重悬于 400 μ l 酵母细胞裂解液, 加入 0.5mm 珠(bead)0.5g 和蛋白酶抑制剂 (protease inhibitor cocktail, 购自 Sigma 公司) 于 0 $^{\circ}$ C, 在 Fastprep120 上以 5m/s 振荡三次, 每次 30 秒。12000rpm 离心 15min, 吸取上清, 再离心 15min, 收集上清为细胞裂解物。取 500 μ l 细胞裂解物, 加 anti-FLAG M2 单克隆抗体(购自 Sigma 公司)(1 μ g/反应), 在 4 $^{\circ}$ C 振荡 1h。12000rpm 离心 2min, 在上清中加入 30 μ l 用酵母细胞裂解液平衡过的 50% protein G-agarose (购自 Roche 公司)悬液, 4 $^{\circ}$ C 振荡 2 h。免疫沉淀物用 IP 缓冲液(20 mM Tris-HCl pH 7.5, 2 mM EGTA, 150 mM NaCl, 1% NP-40, 1mM DTT)洗三次。然后重悬在 1 \times 蛋白质电泳上样缓冲液中, 进行 10%的 SDS-PAGE 电泳。

用湿法将蛋白条带转移到硝酸纤维素膜上, 用含 5%脱脂奶粉的 TBS-T(10 mM Tris-HCl(pH 7.5), 150 mM NaCl, 0.1% Tween-20)溶液封闭膜 2-3 小时, anti-FLAG 和 anti-GFP(Santa Cruz)(抗体稀释度为 1:1000)4 $^{\circ}$ C 结合过夜, 用 TBS-T 室温洗膜四次, 每次 15 分钟。用二抗于室温结合 2 小时, 再用 TBS-T 洗膜四次, 每次 15 分钟。用吸水纸吸尽多余的液体, 用 ECL 试剂显色。

将测定的 cDNA 序列与已有的公共 DNA 序列数据库进行比较, 结果发现有一个 cDNA 克隆的 DNA 序列为新的全长 cDNA。

CaSnf2 cDNA 为 5073bp(图 1A 和 SEQ ID NO:1), 含有完整的开放性读框 5073bp, 编码含 1690 氨基酸残基的多肽(图 1B 和 SEQ ID NO: 2)。

实施例 1

白念珠菌 *CaSNF2* 基因的获得和结构分析

利用同源序列搜索的方法从白念珠菌 (*Candida albicans*) 基因组序列中 (<http://genolist.pasteur.fr/CandidaDB/>)。

根据搜索结果, 合成以下引物:

上游引物: ATGAATCGTCAACCTACAAGAGAG (SEQ ID NO: 3)

下游引物: TCAATCAAATTTGCTGGTGTAGACTC (SEQ ID NO: 4)

以野生型白念珠菌基因组 DNA 为模板, 通过常规的 PCR 反应获得长度约 5.1Kb 的扩增产物。

通过合成一系列引物对扩增产物所含的 DNA 序列进行双向测定。计算机分析表明, 该扩增产物所含的全长 DNA 是一个新的 DNA 序列(如 SEQ ID NO: 1 和图 1A 所示), 编码一个新的 1690 个氨基酸的蛋白质(如 SEQ ID NO: 2 和图 1B 所示)。此蛋白质被命名为 CaSnf2 蛋白, 其编码基因命名为 *CaSNF2* 基因。

利用 BLAST 结构域搜索和同源性比较分析, 发现该基因产物和酿酒酵母

(*Saccharomyces cerevisiae*)的 *ScSNF2* 基因产物有一定的同源性, 一致性(identity)达49%, 因此把这个基因命名为 *CaSNF2*, 其对应的编码产物为 *CaSnf2*。

通过结构域分析, 发现 *CaSnf2* 和 *ScSnf2* 都有几个相似的结构域 (图 1C)。 *CaSnf2* 和 *ScSnf2* 都含有 DEXDc 和 HELICc 结构域, 这两个结构域及其中间的片段是结合 ATP 并且水解 ATP 所必需的, 而且这个区域对整个 Swi/Snf 复合物进行染色体重塑也是必需的(Mohrmann, L. et al. 2005. *Biochim Biophys Acta* 1681: 59-73.)。同时 *CaSnf2* 和 *ScSnf2* 都含有 BROMO 结构域, 这个结构域能够与赖氨酸残基乙酰化后的蛋白相互作用(Ladurner, A.G. 2003. *Mol Cell*. 11: 365-76.)。因此序列比较分析表明白念珠菌 *CaSNF2* 基因是酿酒酵母 *ScSNF2* 基因的同源基因, 白念珠菌 *CaSnf2* 因子是酿酒酵母 *ScSnf2* 因子的同源蛋白。

白念珠菌 *CaSNF2* 基因的核苷酸序列已送 GenBank 登录, 在本申请之前尚未公开。

实施例 2

白念珠菌中 *CaSNF2* 基因的敲除

为了研究白念珠菌 *CaSNF2* 基因在白念珠菌形态发生和毒性表现中的功能, 本实施例首先在白念珠菌中敲除 *CaSNF2* 基因。具体方法如下:

敲除采用图 2A 所示的策略。

利用 5'引物:CGGGATCCATGAATCGTCAACCTACAAGAGAG (SEQ ID NO: 5)

和 3'引物:GAAGATCTGTTGTTGAAGGGCATATTGTTG (SEQ ID NO: 6)

从野生型白念珠菌株基因组 DNA 中用 PCR 的方法扩增大约 0.9kb 的片段, 连接到质粒 pCUB6(Praveen Singh 等人, *Infect Immun*. 2001 December; 69(12): 7898-7903)的 BglIII 位点,

利用 5'引物:CAGGATCCGAACAGAAGAGTCTACACCAG (SEQ ID NO: 7)

和 3'引物:ACATGCATGCGTTCCACAAGTGTTCTATACC (SEQ ID NO: 8)

从野生型白念珠菌株基因组 DNA 中用 PCR 的方法扩增大约 1.0kb 的片段, 继续连接到质粒 pCUB6 的 BamH-SphI 位点, 从而体外构建了 *CaSNF2* 基因敲除质粒 pCaSNF2-KO, 在此质粒中 *CaSNF2* 开放阅读框(open reading frame, ORF)中约 4.0 kb DNA 片段被 *HisG-URA3-HisG* 替代.质粒 pCaSNF2-KO 用 PstI 酶切并转化常规的白念珠菌 *ura*⁻营养缺陷型菌株, 在缺少尿嘧啶的合成培养基上可以筛选到转入质粒的转化子.通过 0.9kb 和 1.0kb 两个同源片段和基因组上的 *CaSNF2* 同源片段重组, 可以把染色体上的 *CaSNF2* 基因中的 4.0 kb DNA 同源片段用 *HisG-URA3-HisG* 给替换掉, 从而破坏染色体上的 *CaSNF2* 基因, 正确插入的转化子通过 Southern 杂交分析确定.筛选标志 *URA3* 和一个拷贝 *HisG* 序列可以在含 5-氟乳清酸(5-fluoro-orotic acid, 5-FOA)平板上通过两个同向 *HisG* 同源序列在同一条染色体上的重组而环出, 丢失了 *URA3* 筛选标志的重组子可以在 5-FOA 平板上生长, 通过负向筛选得到(Boeke et al. 1984. *Mol Gen Genet*. 197: 345-346.), 从而可以进行下一轮的转化进而敲除另一条染色体上的 *CaSNF2* 基因。

所得重组子的基因型用 Southern 杂交技术检测确定(图 2B)。HindIII 酶切各个菌株基因组 DNA, 用 1.0kb 片段作探针杂交。根据白念珠菌基因组序列 (<http://genolist.pasteur.fr/CandidaDB/>), 发现野生型菌株显示一条杂交条带, 单拷贝缺失株显示两条杂交条带, 在 5-FOA 平板上将一份拷贝的 *HisG* 和 *URA3* 环出后, 又出现了一条较小的杂交条带, 通过第二轮的转化和环出, 可以将染色体上的第二拷贝的 *CaSNF2* 基因敲除。图 2B 显示了 *CaSNF2* 基因敲除中各个菌株的 Southern 杂交分析图谱。

实施例 3

CaSNF2 基因的敲除对白念珠菌菌丝形成的影响

通过同源重组的方法在白念珠菌中敲除了 *CaSNF2* 基因, Southern 杂交分析确证这个敲除是成功的。*CaSNF2* 基因的敲除, 引起了白念珠菌一系列细胞和菌落形态的变化。

白念珠菌菌丝形成能力对其侵入和感染是必需的。*CaSNF2* 基因的敲除大大降低了白念珠菌菌丝形成能力。在含 YPD 加 10%胎牛血清培养基上, 经过 37°C, 3.5 小时培养后, 可以看到野生型菌株形成典型的菌丝(hyphae), 延长的管状细胞和平行生长的细胞壁, 细胞之间没有明显的缢缩, 而 *casnf2/casnf2* 缺失株则形成假菌丝, 没有形成这种典型的菌丝, 细胞之间是链状相连。一部分细胞要伸长一些(图 3), 个别细胞伸长程度为原来的 2-3 倍, 极少量细胞呈单个生长细胞, 而且也形成较长的菌丝, 不同细胞间形态差异比较大。在 Lee's 液体培养中, 细胞形态和在 YPD 加 10%血清培养中基本一致, 说明 *CaSNF2* 基因的缺失确实在很大程度上阻断了白念珠菌菌丝的形成。

在微氧的菌丝诱导条件下, 当野生型菌株被包埋在固体 YPS 培养基中, 经过 3 天后可形成明显的丝状生长。而 *casnf2/casnf2* 缺失株没有形成菌丝, 菌落仍然平滑呈球形, 说明 *CaSNF2* 的缺失阻断了在微氧条件下的菌丝生长。

实施例 4

白念珠菌 *casnf2/casnf2* 缺失株在菌体生长条件下的表型

CaSNF2 基因的敲除导致白念珠菌在菌体生长条件下形态发生变化(图 4)。*casnf2/casnf2* 缺失株在 YPD 或者 SD 培养基, 30°C 培养条件下, 几个到几十个细胞会形成有分枝的链状细胞, 细胞基本呈椭圆型或者略微延长, 呈单极分裂模式, 细胞和细胞之间相互粘连, 而且通过 Calcofluor White 染料染色, 在细胞之间可以看到明显的缢缩。这种细胞形态和白念珠菌中假菌丝形态很相似, 而且这种细胞形态也不是细胞分裂不完全所致, 因为 DAPI 染色表明每个细胞中都含有一个完整的细胞核。

野生型菌株在同样培养条件下, 呈现典型的分散的球形细胞生长, 也有 2-4 个细胞连在一起, 那是正在分裂的细胞。在固体培养基上, 菌落形态由野生型菌株光滑的表面, 规则的边缘转变为 *casnf2/casnf2* 缺失株皱褶的表面和不规则的边缘, 而且菌落和培养基间以及菌和菌之间的粘附明显下降。而在酿酒酵母 *ScSNF2* 的缺失并没有引起细胞和

菌落形态的变化，也说明了 *CaSNF2* 和 *ScSNF2* 在体内可能发挥不同的作用。

同样在固体 SLAD 培养基上，野生型菌落基本呈均向生长的球形，而两个缺失株都呈现典型的假菌丝生长形态。在白念珠菌 *CBK1*，*FKH2* 缺失株和酿酒酵母 *FKH1*，*FKH2* 双缺失株中也可形成类似的细胞和菌落形态(McNemar, M.D. et al. 2002. J Bacteriology. 184: 2058–206; Hollenhorst, P.C. et al. 2000. Genetics. 154: 1533–1548; Bensen, E.S. et al. 2002. Eukaryotic Cell. 1: 787–798.)。

实施例 5

白念珠菌蛋白因子 *CaSnf2* 和 *CaSwi1* 能在体内相互作用

在酿酒酵母里，*ScSwi1* 和 *ScSnf2* 都是 *ScSwi/Snf* 复合物的组分，两者在体内非常紧密的结合在一起。而且任何一个因子的缺失都引起相同的突变表型(Peterson, C.L. et al. 1992. Cell. 68: 573-583; Peterson, C.L. et al. 1994. Proc Natl Acad Sci U S A. 91: 2905-8.)。为了检测在白念珠菌中 *CaSwi1* 和 *CaSnf2* 能否也相互作用，本发明人构建了 GFP-*CaSwi1* 和 FLAG-*CaSnf2* 融合表达菌株。方法如下：

利用 5'引物:GGGGATCCGGACACCTACACCAAACA (SEQ ID NO: 9)

和 3'引物: CCGCTCGAGCCATTCACACCCTGCCATA (SEQ ID NO: 10)

从野生型白念珠菌株基因组 DNA 中用 PCR 的方法扩增大约 0.9kb 的片段，连接到质粒 p584(参见美国专利 6,911,468、5,520,253、或 5,477,351)的 *KpnI*-*XhoI* 位点，

利用 5'引物:TCCCCGCGGTACCTTGAGATTTGGCTGTAACA (SEQ ID NO: 11)

和 3'引物: TCCCCGCGCCATGTCTGATTGGTTGAATG (SEQ ID NO: 12)

从野生型白念珠菌株基因组 DNA 中用 PCR 的方法扩增大约 0.9kb 的片段，继续连接到质粒 p584 的 *SacII* 位点，从而得到质粒 p*CaSWI1*-GFP，用 *KpnI* 酶切质粒 p*CaSWI1*-GFP 并转化白念珠菌，通过同源重组的原理，使得 *CaACT1p*-GFP-*URA3*-GFP 插入到染色体上 *CaSWI1* 基因的起始密码子前面。通过在 5-FOA 平板上的负向选择，环出一个拷贝的 *GFP* 和 *URA3*，从而得到菌株能在 *CaACT1* 启动子调控下表达融合蛋白 GFP-*CaSwi1*，GFP 融合在 *CaSwi1* 的 N 端。

利用 5'引物:GAGGATCCTCTGACGACGATGATGACAATG (SEQ ID NO: 13)和

3'引物: ACATGCATGCCTTGTCATCGTCATCCTTGTAATCGATGTCATGATCTTATAATCACCGTCATGGTCTTTGTAGTCATCAAATTTGCTGGTGTAGACTC (SEQ ID NO: 14)

从野生型白念珠菌株基因组 DNA 中用 PCR 的方法扩增大约 0.9kb 的片段，连接到常规的质粒 pFLAG-Act1(Umeyama, T., Nagai, Y., Niimi, M., 和 Uehara, Y. (2002). Construction of FLAG tagging vectors for *Candida albicans*. Yeast 19, 611–618.)的 *BamHI*-*SphI* 位点，得到质粒 p4FLAG-*CaSNF2*。用 *PstI* 酶切 p4FLAG-*CaSNF2* 并转化融合表达 GFP-*CaSwi1* 的菌株，通过位点特异性的同源重组，质粒 p4FLAG-*CaSNF2* 被重组到染色体的 *CaSNF2* 位点，从而得到另一菌株，该菌株中 4 个 FLAG 融合在 *CaSnf2* 的 C 端，

并在 *CaSNF2* 自身启动子下表达,同时也表达 GFP-CaSwi1 融合蛋白。质粒 pFLAG-Act1 用 *StuI* 酶切并转化融合表达 GFP-CaSwi1d 菌株,得到菌株仅仅表达 GFP-CaSwi1 融合蛋白。

通过免疫共沉淀实验发现,在两个融合蛋白共表达的菌株中,GFP-CaSwi1 能够被抗 FLAG 的抗体共沉淀并且通过 Western blot 能够检测到。作为对照,在仅仅表达 FLAG 空载体的菌株中,GFP-CaSwi1 不能被免疫共沉淀下来(图 5)。由此可见,在白念珠菌中,CaSwi1 和 CaSnf2 也能相互作用,而且它们的相互作用不管在 YPD 培养的菌体条件下,还是血清诱导和 Lee's 培养基中都能发生(图 5),说明了 CaSwi1 和 CaSnf2 也是紧密的结合在一起,而非瞬时作用,同时也说明在白念珠菌中也存在着保守的 CaSwi/Snf 复合物。

实施例 6:

CaSNF2 基因的敲除导致白念珠菌毒性的下降

本实施例通过小鼠系统感染实验检测 *CaSNF2* 基因的敲除对白念珠菌毒性的影响。方法如下:以体重在 16-18g ICR 雄性小鼠为实验对象,通过尾静脉注射 100 μ l 白念珠菌各菌株,并培养观测 25 天。各个菌株注射浓度为 5×10^7 细胞 / 毫升。

以上结果表明, *CaSNF2* 基因的敲除导致白念珠菌菌丝生长的缺陷,也就是使得白念珠菌即使在强烈的菌丝诱导条件下也无法形成菌丝。通过小鼠尾静脉注射野生型菌株和缺失株进行小鼠系统感染实验来检测 *CaSNF2* 基因的被坏是否导致了菌株毒性下降。发现野生型菌株有较强的毒性,小鼠在注射了 5×10^6 野生型细胞后于 14 天后全部死亡;同样条件下,小鼠注射了 5×10^6 *casnf2* / *casnf2* 缺失株细胞后存活 25 天以上,证明 *casnf2* / *casnf2* 缺失株没有毒性(图 6),说明了 CaSnf2 在白念珠菌中是一个必需的毒性因子。

实施例 7

CaSnf2 蛋白重组表达和纯化

在该实施例中,以实施例 1 中的 PCR 扩增产物为模板,用序列如下的 5'和 3'端的 PCR 寡核苷酸引物进行扩增,获得白念珠菌 CaSnf2 DNA 作为插入片段。

PCR 反应中使用的 5'端寡核苷酸引物序列为:

5'-CGGGATCCATGAATCGTCAACCTACAAGAGAG- 3' (SEQ ID NO: 15)

该引物含有 BamHI 限制性内切酶的酶切位点,在该酶切位点之后是由起始密码子开始的部分编码序列;

3'端引物序列为:

5'-CGGGATCCTCAATCAAATTTGCTGGTGTAGACTC-3' (SEQ ID NO: 16)

该引物含有 BamHI 限制性内切酶的酶切位点、翻译终止子和白念珠菌 CaSnf2 的部分编码序列。

白念珠菌 CaSnf2 蛋白 cDNA PCR 产物纯化后经 BamHI 酶切再与质粒 pGEX-2T (购于美国 Invitrogen 公司)按常规方法重组形成载体 pGEX-2T-CaSnf2 并转化至感受态大肠杆菌 DH5 α , 挑取阳性克隆用 EcoRI 酶切鉴定插入方向, 酶切产物在 0.8%琼脂糖凝胶电泳分析。鉴定后纯化并测序(ABI 公司的 377 型测序仪, BigDye Terminator 试剂盒, PE 公司)。经测序证实, 已插入了完整的 CaSnf2 编码序列。

挑表达 CaSnf2 的阳性 DH5 α 克隆接种于 100ml 2 \times YTA 培养基中, 37 $^{\circ}$ C 300rpm 振荡培养 12-15hr, 1:10 稀释于预热的 2 \times YTA 培养基继续振荡培养 1.5hr, 加 100mM IPTG 至 0.1mM 后 30 $^{\circ}$ C 诱导 2-6hr, 5,000g 4 $^{\circ}$ C 离心 10min 去上清, 置冰上用 50ml 1 \times PBS (0.14M NaCl, 2.7 mM KCl, 10.1mM Na₂HPO₄, 1.8mM KH₂PO₄, pH7.3) 重悬, 超声(B. Braun Labsonic U)破碎后再加入 20% Triton X-100 至 1%轻摇 30min, 然后 12,000g 4 $^{\circ}$ C 离心 10min, 上清用 0.8 μ m 滤膜过滤后, 过 1ml 50%谷胱甘肽 Sepharose 4B 层析柱, 1 \times PBS 充分洗涤后, 加入 500ul 谷胱甘肽洗脱缓冲液 (10 mM 谷胱甘肽, 50 mM Tris-HCl, pH 8.0)室温静置 30 分钟后收集洗脱液, 重复洗脱 2-3 次, 得到白念珠菌 CaSnf2 蛋白。

实施例 8

抗 CaSnf2 蛋白抗体的产生

将实施例 7 中获得的重组白念珠菌 CaSnf2 蛋白用来免疫动物以产生抗体, 具体方法如下。重组分子用层析法进行分离后备用。也可用 SDS-PAGE 凝胶电泳法进行分离, 将电泳条带从凝胶中切下, 并用等体积的完全 Freund's 佐剂乳化。用 50-100 μ g/0.2ml 乳化过的蛋白, 对小鼠进行腹膜内注射。14 天后, 用非完全 Freund's 佐剂乳化的同样抗原, 对小鼠以 50-100 μ g/0.2ml 的剂量进行腹膜内注射以加强免疫。每隔 14 天进行一次加强免疫, 至少进行三次。获得的抗血清的特异反应活性用它在体外沉淀白念珠菌 CaSnf2 蛋白基因翻译产物的能力加以评估。结果发现, 抗体可特异性地与本发明的 CaSnf2 蛋白发生结合。

讨论

酿酒酵母(*Saccharomyces cerevisiae*)SNF2 基因(*ScSNF2*)最先在调控单倍体细胞 *HO* 基因的正常表达克隆得到的(Stern, M. et al. 1984. J Mol Biol. 178: 853-868.)。ScSNF2 编码的蛋白 ScSnf2 定位在细胞核中, 具有 DEXDc, HELICc 和 BROMO 结构域。DEXDc, HELICc 结构域以及两结构域中间的部分对 ScSnf2 的 ATPase 活性是必需的(Mohrmann, L. et al. 2005. Biochim Biophys Acta. 1681: 59-73.), 而 BROMO 结构域能够和赖氨酸残基被乙酰化的蛋白相互作用(Ladurner, A.G. 2003. Mol Cell. 11: 365-76; Aileen, K. et al. 1987. Genet. 116: 523-530.)。ScSNF2 基因的缺失, 在单倍体细胞中可阻断 a 型和 α 型细胞生殖型之间的转换, 在双倍体细胞间引起减数分裂缺陷, 导致生长缺陷, 对非发酵糖类(如半乳糖, 蔗糖, 棉子糖, 甘油等)利用的缺陷以及 a, α 型细胞特异性基因(如 *STE6*, *MCM1*, *MAT α 1* 等)以及其它和代谢相关基因(如 *SUC2*, *INO1*, *ADH2*, *PHO85* 等)表

达的下调(Aileen, K. et al. 1987. Genet. 116: 523-530; Peterson, C.L. et al. 1992. Cell. 68: 573-583; Sudarsanam, P. et al. 2000. PANS. 97: 3364-3369.)。遗传学和生物化学证据表明在酿酒酵母细胞中, ScSnf2 是由 11 个成分组成的 Swi/Snf 复合物中的成员之一, 在 ATP 存在的条件下, 该复合物可以和 RNA 聚合酶 II 全酶(RNA polymerase holoenzyme), 转录激活因子(transcriptional activator), 染色质上的组蛋白, 非组蛋白成分以及其它抑制因子相互作用, 改变染色质的结构, 去除抑制效应, 把 RNA 聚合酶 II 全酶募集到靶基因的启动子上, 从而激活转录(Burns, L.G. et al. 1997. Mol Cell Biol. 17:4811-4819; Kruger, et al. 1995. Genes Dev. 9: 2770-2779; Igor, M. et al. 1997. EMBO J. 16: 6263-6271; Christopher, J. et al. 1996. Cell. 84: 235-244; Neely, K.E. et al. 2002. Mol Cell Biol. 22: 1615-1625; Yudkovsky, N. et al. 1999. Genes Dev. 13: 2369-2374.)。同时该复合物也有可能使染色质结构更加紧密而抑制靶基因的转录(Joseph, A. et al. 2002. Genes Dev. 16: 2231-2236.)。

白念珠菌(*Candida albicans*)是一种人体机会性致病真菌, 能引起广泛的深部和浅部感染, 但是其真正的致病因子和致病机理没有完全研究清楚。利用生物信息学比较分析, 本发明人在白念珠菌的基因组序列中发现一个新基因, 该基因编码的产物和酿酒酵母(*Saccharomyces cerevisiae*)中的 *SNF2* 基因编码的蛋白有一定的同源性, 一致性达 49%, 以及结构域的相似性, 因此命名为 *CaSNF2*。利用同源重组原理, 在白念珠菌中敲除 *CaSNF2*, 构建了 *casnf2/casnf2* 缺失株, 该缺失株不能形成菌丝, 说明 *CaSnf2* 是白念珠菌菌丝生长所必需。不管在菌体还是菌丝培养条件下, 白念珠菌 *casnf2/casnf2* 突变株都以假菌丝形态生长。免疫共沉淀实验证明 *CaSnf2* 因子和 *CaSwi1* 因子能够在白念珠菌体内相互作用, 这种相互作用不管在菌体还是菌丝条件下都能发生, 说明在白念珠菌中存在着 *CaSwi/Snf* 复合物调控白念珠菌的形态发生。小鼠系统感染实验表明 *casnf2/casnf2* 缺失株没有毒性, 说明 *CaSnf2* 是白念珠菌中一个重要的毒性因子。

在本发明提及的所有文献都在本申请中引用作为参考, 就如同每一篇文献被单独引用作为参考那样。此外应理解, 在阅读了本发明的上述讲授内容之后, 本领域技术人员可以对本发明作各种改动或修改, 这些等价形式同样落于本申请所附权利要求书所限定的范围。

序列表

- <110> 中国科学院上海生命科学研究院
- <120> 白念珠菌菌丝调控因子基因及其用途
- <130> 057914
- <160> 16
- <170> PatentIn version 3.2
- <210> 1
- <211> 5073
- <212> DNA
- <213> 白念珠菌(*Candida albicans*)
- <220>
- <221> CDS
- <222> (1)..(5070)

```

<400> 1
atgaatcgtc aacctacaag agaggatatt caaagagcga ttcaacgctg gcatcaaatg      60
aaacaacaat atggagacca agttcaactt aatcctgaat ttgttaaatt aaccaagttc      120
ttgaatactt tgaaaatgca acaacaacgt tttcagcaac aataccaaca acaacaacaa      180
caacaacaac agcagcagca gcagcagcag cagcaacaac aacaacaaca gcagcaacaa      240
caaagtctaa accattcaca acaatcgcea ttgctacaaa atgcacaggg ccaaactcca      300
caacaacctc caactcctca acagttttct aatttcaatc agaacgggta taatgggtcaa      360
cagttttctt ctcaagtaca ctccactgct attgggtgggt ctttatcgac ttcaggacat      420
ggaaccccat tagttacaaa tgccaatttg atgacaggaa agaaaaaac aaggacacca      480
aatgcccatt ttgtaacca gacggctgct ggaacaccat tacaacaaca acaacaacaa      540
caacaacaac aacaacagcc ttccctcat ggaacaatt cgaatcctat gctaaccag      600
acggctcagc aaccaccaca atctcaacaa cgtcaacaaa accaaccacc aagtcccaa      660
tcagcattca ctaatcaaca attccaatta ttgaaatctc agcttcaagc attcaagtat      720
tttgtgagag ctcttagtca gggtaaggt caataaccac aaaacttgat agcatatgtt      780
tcgaaatcat catctgctat ggccaatgat atgtacttac cagcagtga cggcctcaa      840
actaatggta tggatcgtac aatgcaaatg ccacaatcaa taccttcaca accacaacaa      900
tatgcccttc aacacaacaa cgatttgctg aacctgaatc caaagtcaac tggaggcacc      960
cctgaaatcc cagaaaagaa aaaaggaaag cgtggacctt aaccaagaa tccgaaaaaa      1020
cctacaagaa aacagttgag agaagaagaa cagagacttg cattggaaaa acaaaagaaa      1080
gaacttgaac aaaatagact caagagtagt gctcctcaag cattcccgcc tcaagcaggt      1140
ttacaaggac aagctccttt ccaccacaa ccaccacagc agtcacaaca acatgtacct      1200
caaccacctc cagcateaac ttcatctagt ccaccgggtg ggttgccaca gctgcaaccg      1260
caacagcaac agcaacagcc atctctcca attactaaac ctgccacacc tcaaccttta      1320
ttccctgate catctctccc agtgaatata aagagtgtag taccggacaa agcaaacac      1380
aagaaggtaa taataccggt aactaaacca aatattgaag tagacacctt cgagttattt      1440
gacattatca gtgatgaggt gaaagatata ccgtttaata ctttatatgc tccacagagt      1500
agatttcaga tcccttctgt tttgcctgat ggtataaata tggaagatat ttatgtgaac      1560
agagaaggat atatgcaaat tacaatagaa caagagaagg agagattaag aaaacaatt      1620
gacagtttga atgaaaaaga cactgaaaag aaattggaac ttgaaacaca attaagtcaa      1680
ttggaattga ttccttatca gaaagattta cgtggtaaag ttcttataca atcttggttt      1740
gggaaatcat tacttcttaa ttcacatcca aactttttag caagattcag ttcattatct      1800
atggacagtg taatcatgac aaacagattt taccgactcc aattggaatc catgatgaga      1860
gaacaaaata agaacaatgg caaaactatt gaagaatca taaatttcag tgatcgaagt      1920
agcatcaaa cctgcaagaa atcagaccgg ttgtcaaggt ttatgactaa aattaataat      1980
ttccataatc aaactgcaa ggaagaacag aaaaagtgg aaaaatggc taacaacagt      2040
ttgcaagcat tgaactgaa tgaatgtgaa gcttatttga aattgttggg tcatcaaaag      2100
gatacaagaa ttaccattt attagaacaa acaaatcaat ttttgactc tttggctctt      2160
gcagtgcaaa gtcaacaaaa agaggctcag gacaatttag catattcagg tctgcccata      2220
gaaccagcat cagttgaacc ccttgatgat gagaagagag aaaaattga ttattataat      2280
gttgctcata gaattaaaga agaagtcacc aagcaacctt caatattggt tgggggtact      2340
ttgaaggagt atcaattgaa aggtttacaa tggatggtt cattgtttaa taatcatttg      2400
aacggtatct tggcagatga gatgggtttg ggtaaaacaa ttcaaacatc ttcattactc      2460
acatatcttg ttggaagtga aaaaattcct ggtccatttt tagtaattgt tcccttatca      2520
acagtaacca attggaattt agaattttaa aaatgggctc cctcaattaa aaaaattacc      2580
tataaagta ctccaaatca acgtaaagt atgcaacacg atatcagaac cgggaatttc      2640
caatttagtat tgacgacatt tgaatattgt attaaagata aaggattatt gggtagaattc      2700
aaatgggtcc atatgattat tgaatgaagt catcgtatga agaattgctaa ttcgaaatta      2760
cttgagacat tgacacaata ttaccatagt gattatcgtt tgattttgac tggactcca      2820
ttgcaaaata acttaccaga attatgggcc ttgttaaatt ttgttttacc caaaatttc      2880
aactctgtga aatcatttga tgaatggttc aatacaccat ttgccaatc tgggtgcaa      2940
gataagatag aattgacaga agaagaacaa ttgttgggtg ttagaagatt gcataaagtt      3000
ttaagaccgt tccctttaaag aagattaaag aaagatgttg aaaaagattt accaacaacg      3060
gtgaaaaaag ttgtcaaatg taaatcatcg gcattgcaat ctaaattata tcaacaagt      3120
ttgaggtata atatgttga tgctggagat cctgccaatg gatcagtgcc cgttactata      3180
aaaaacgcca acaatcaaat aatgcaattg aaaaaattt gtaatcacc ttttgtttat      3240
gaagaagtgg agaattttag taatcctaatt attgaaacca acgatcagat ttggagagta      3300
    
```

gctggtaaat ttgaattatt agacaaagtc ttacctaaat tcaaagctac tggtcataaa 3360
 gttttgattt tcttccaaat gactcaaatt atgaatatca tggaaagactt ttacgattc 3420
 agaggatag aatataatgag attggatggt ggaaccaaag ctgatgatag aactgattta 3480
 ttgaaaagtt ttaatgcccc agattctgat tatttttgtt tctttttatc cactcgtgct 3540
 ggtgggttag gtcttaattt acaaaactgcc gatacagtc ttttttttga tactgattgg 3600
 aatccctatc aagatttaca agctcaagat agagcccatc gtattgggtca aaagaatgaa 3660
 gttagaatat taagattgat tacggaaaac tcggtggaag aaatgatttt ggaagagct 3720
 cataaaaagt tggagattga tggtaaaagt attcaagccg gtaaatcga taacaaatcc 3780
 actgctgaag aacaagaagc tatgttaaga gcattaatag aaaaagaaga tgaacgtaga 3840
 cagaaggtg gtaccgacga agaagaagaa gatttggatg atgatgaatt gaatcaaatt 3900
 attgccagaa atgaaaacga gttgggtggt tttaggaaaa tggatgaaga aagatacctt 3960
 gctaccaaga atgctccata cccatccaga ttgtataccg aggaagagtt gcctgaaatt 4020
 tataagatag atccagaaga acttttcaag aaagaagacg ttgcactgga agagtatggt 4080
 cgtgggtgta gagaaggaa aatattacaa tacgatgata atttaactga agagcaatgg 4140
 ttgaagaaaa ttgaaggat ggtatctgac gacgatgatg acaatgacga tgacgatggt 4200
 aatgttgata tgagtgttc tgaatggaa gccaaagcga agaaccgaa aggaagaaga 4260
 ggtcggaaac ccaaggttgc acgtgtttaa gatgaagaaa gtcaaaactga atcggatgtt 4320
 atttctgtta agcgtcaatt tcctgaagat gcagatgact ttataccacc aaagagacag 4380
 aaatctgcta caccaggtgg cactactact tcgggtagag gcagaggtag gggccgtggt 4440
 agagtcgtg gacgtggaag aggcagagcc agaggatctt tgttgtctcg ttatacact 4500
 tctgtagacc ccttaactgc agacgagaga tctactttac aaaatcaaat tgaacacata 4560
 ttggggttga tcattaatta taagaatgaa catgatagag tattgagtga attgtttttg 4620
 gtgaagccac ccaagagatt ttaccctgat tactatgttt tgataaagca tccaattgca 4680
 cttgacgtga ttaaaaagag aacagcgtca aaatcttaca gcaagattag ggaatttttg 4740
 gaggatattc atttgatgtt caccaacgcc aagatatata atgaagaagg ttcaattgtt 4800
 tatcaagatg cagcattctt ggaaagatta tcaatggaca aatcaaga attatcagcc 4860
 aatctttcag aagcagaagt caataagatt ttggatttcg ctgagtttga cgaatgttt 4920
 agtttaaac cattagtcc ttcaactgct atcaaacatc caattgaagc taaattagag 4980
 aaaatagata aaggtgaagc aattgacagt ccattactca atgcttctac cactgctgga 5040
 acagaagagt ctacaccagc aaattttgat tga 5073

<210> 2
 <211> 1690
 <212> PRT
 <213> 白念珠菌(Candida albicans)

<400> 2
 Met Asn Arg Gln Pro Thr Arg Glu Asp Ile Gln Arg Ala Ile Gln Arg
 1 5 10 15
 Trp His Gln Met Lys Gln Gln Tyr Gly Asp Gln Val Gln Leu Asn Pro
 20 25 30
 Glu Phe Val Lys Leu Thr Lys Phe Leu Asn Thr Leu Lys Met Gln Gln
 35 40 45
 Gln Arg Phe Gln Gln Gln Tyr Gln Gln Gln Gln Gln Gln Gln Gln
 50 55 60
 Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
 65 70 75 80
 Gln Ser Leu Asn His Ser Gln Gln Ser Pro Leu Leu Gln Asn Ala Gln
 85 90 95
 Gly Gln Thr Pro Gln Gln Pro Pro Thr Pro Gln Gln Phe Ser Asn Phe
 100 105 110
 Asn Gln Asn Gly Tyr Asn Gly Gln Gln Phe Ser Ser Gln Val His Ser
 115 120 125
 Pro Ala Ile Gly Gly Ser Leu Ser Thr Ser Gly His Gly Thr Pro Leu
 130 135 140
 Val Thr Asn Ala Asn Leu Met Thr Gly Lys Lys Asn Thr Arg Thr Pro
 145 150 155 160
 Asn Ala Gln Phe Gly Asn Gln Thr Ala Ala Gly Thr Pro Leu Gln Gln
 165 170 175
 Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Phe Pro His Gly Asn
 180 185 190
 Asn Ser Asn Pro Met Leu Asn Gln Thr Ala Gln Gln Pro Pro Gln Ser
 195 200 205
 Gln Gln Arg Gln Gln Asn Gln Pro Pro Ser Pro Gln Ser Ala Phe Thr
 210 215 220
 Asn Gln Gln Phe Gln Leu Lys Ser Gln Leu Ala Phe Lys Tyr
 225 230 235 240
 Phe Val Arg Ala Pro Ser Gln Gly Gln Gly Gln Ile Pro Gln Asn Leu
 245 250 255
 Ile Ala Tyr Val Ser Asn Pro Ser Ser Ala Met Ala Asn Asp Met Tyr
 260 265 270
 Leu Pro Ala Val Asn Arg Pro Gln Thr Asn Gly Met Asp Arg Thr Met
 275 280 285
 Gln Met Pro Gln Ser Ile Pro Ser Gln Pro Gln Gln Tyr Ala Leu Gln
 290 295 300
 Gln Gln Asn Asp Leu Leu Asn Leu Asn Pro Lys Ser Thr Gly Gly Thr
 305 310 315 320

Pro Glu Ile Pro Glu Lys Lys Lys Gly Lys Arg Gly Pro Lys Pro Lys
 325 330 335
 Asn Pro Lys Lys Pro Thr Lys Lys Gln Leu Arg Glu Glu Glu Gln Arg
 340 345 350
 Leu Ala Leu Glu Lys Gln Arg Gln Glu Leu Glu Gln Asn Arg Leu Lys
 355 360 365
 Ser Ser Ala Pro Gln Ala Phe Pro Pro Gln Ala Gly Leu Gln Gly Gln
 370 375 380
 Ala Pro Phe Pro Pro Gln Pro Pro Gln Gln Ser Gln Gln His Val Pro
 385 390 395 400
 Gln Pro Pro Pro Ala Ser Thr Ser Ser Ser Pro Pro Gly Gly Leu Pro
 405 410 415
 Gln Leu Gln Pro Gln Gln Gln Gln Gln Pro Ser Arg Pro Ile Thr
 420 425 430
 Lys Pro Ala Thr Pro Gln Pro Leu Phe Pro Asp Pro Ser Pro Pro Val
 435 440 445
 Asn Ile Lys Ser Val Val Pro Asp Lys Ala Asn Asn Lys Lys Val Ile
 450 455 460
 Ile Pro Val Thr Lys Pro Asn Ile Glu Val Asp Thr Phe Glu Leu Phe
 465 470 475 480
 Asp Ile Ile Ser Asp Glu Val Lys Asp Ile Pro Phe Asn Thr Leu Tyr
 485 490 495
 Ala Pro Gln Ser Arg Phe Gln Ile Pro Ser Phe Leu Pro Asp Gly Ile
 500 505 510
 Asn Met Glu Asp Ile Tyr Val Asn Arg Glu Gly Tyr Met Gln Ile Thr
 515 520 525
 Ile Glu Gln Glu Lys Glu Arg Leu Arg Lys Gln Ile Asp Ser Leu Asn
 530 535 540
 Glu Lys Asp Thr Glu Lys Lys Leu Glu Leu Glu Thr Gln Leu Ser Gln
 545 550 555 560
 Leu Glu Leu Ile Pro Tyr Gln Lys Asp Leu Arg Gly Lys Val Leu Ile
 565 570 575
 Gln Ser Trp Phe Gly Lys Ser Leu Leu Pro Asn Ser His Pro Asn Phe
 580 585 590
 Leu Ala Arg Phe Ser Ser Leu Ser Met Asp Ser Val His Met Thr Thr
 595 600 605
 Asp Leu Tyr Arg Leu Gln Leu Glu Ser Met Met Arg Glu Gln Asn Lys
 610 615 620
 Lys His Gly Lys Thr Ile Glu Glu Ile Ile Asn Phe Ser Asp Arg Ser
 625 630 635 640
 Ser Ile Lys Ala Val Lys Lys Ser Asp Arg Leu Ser Arg Phe Met Thr
 645 650 655
 Lys Ile Asn Asn Phe His Asn Gln Thr Ala Lys Glu Glu Gln Lys Lys
 660 665 670
 Leu Glu Lys Met Ala Lys Gln Arg Leu Gln Ala Leu Lys Leu Asn Asp
 675 680 685
 Glu Glu Ala Tyr Leu Lys Leu Leu Asp His Thr Lys Asp Thr Arg Ile
 690 695 700
 Thr His Leu Leu Glu Gln Thr Asn Gln Phe Leu Asp Ser Leu Ala Leu
 705 710 715 720
 Ala Val Gln Ser Gln Gln Lys Glu Ala Gln Asp Asn Leu Ala Tyr Ser
 725 730 735
 Gly Arg Ala Ile Glu Pro Ala Ser Val Glu Pro Leu Asp Asp Glu Lys
 740 745 750
 Arg Glu Lys Ile Asp Tyr Tyr Asn Val Ala His Arg Ile Lys Glu Glu
 755 760 765
 Val Thr Lys Gln Pro Ser Ile Leu Val Gly Gly Thr Leu Lys Glu Tyr
 770 775 780
 Gln Leu Lys Gly Leu Gln Trp Met Val Ser Leu Phe Asn Asn His Leu
 785 790 795 800
 Asn Gly Ile Leu Ala Asp Glu Met Gly Leu Gly Lys Thr Ile Gln Thr
 805 810 815
 Ile Ser Leu Leu Thr Tyr Leu Val Glu Val Lys Lys Ile Pro Gly Pro
 820 825 830
 Phe Leu Val Ile Val Pro Leu Ser Thr Val Thr Asn Trp Asn Leu Glu
 835 840 845
 Phe Glu Lys Trp Ala Pro Ser Ile Lys Lys Ile Thr Tyr Lys Gly Thr
 850 855 860
 Pro Asn Gln Arg Lys Val Met Gln His Asp Ile Arg Thr Gly Asn Phe
 865 870 875 880
 Gln Leu Val Leu Thr Thr Phe Glu Tyr Val Ile Lys Asp Lys Gly Leu
 885 890 895
 Leu Gly Arg Ile Lys Trp Val His Met Ile Ile Asp Glu Gly His Arg
 900 905 910
 Met Lys Asn Ala Asn Ser Lys Leu Ser Glu Thr Leu Thr Gln Asn Tyr
 915 920 925
 His Ser Asp Tyr Arg Leu Ile Leu Thr Gly Thr Pro Leu Gln Asn Asn

930	935	940
Leu Pro Glu Leu Trp Ala	Leu Leu Asn Phe Val	Leu Pro Lys Ile Phe
945	950	955
Asn Ser Val Lys Ser Phe	Asp Glu Trp Phe Asn Thr	Pro Phe Ala Asn
	965	970
Thr Gly Gly Gln Asp Lys	Ile Glu Leu Thr Glu Glu	Thr Leu Leu
	980	985
Val Ile Arg Arg Leu His	Lys Val Leu Arg Pro Phe	Leu Leu Arg Arg
	995	1000
Leu Lys Lys Asp Val Glu	Lys Asp Leu Pro Asn Lys	Val Glu Lys
1010	1015	1020
Val Val Lys Cys Lys Ser	Ser Ala Leu Gln Ser Lys	Leu Tyr Gln
1025	1030	1035
Gln Met Leu Arg Tyr Asn	Met Leu Tyr Ala Gly Asp	Pro Ala Asn
1040	1045	1050
Gly Ser Val Pro Val Thr	Ile Lys Asn Ala Asn Asn	Gln Ile Met
1055	1060	1065
Gln Leu Lys Lys Ile Cys	Asn His Pro Phe Val Tyr	Glu Glu Val
1070	1075	1080
Glu Asn Leu Ile Asn Pro	Asn Ile Glu Thr Asn Asp	Gln Ile Trp
1085	1090	1095
Arg Val Ala Gly Lys Phe	Glu Leu Leu Asp Lys Val	Leu Pro Lys
1100	1105	1110
Phe Lys Ala Thr Gly His	Lys Val Leu Ile Phe Phe	Gln Met Thr
1115	1120	1125
Gln Ile Met Asn Ile Met	Glu Asp Phe Leu Arg Phe	Arg Gly Met
1130	1135	1140
Lys Tyr Met Arg Leu Asp	Gly Gly Thr Lys Ala Asp	Asp Arg Thr
1145	1150	1155
Asp Leu Leu Lys Ser Phe	Asn Ala Pro Asp Ser Asp	Tyr Phe Cys
1160	1165	1170
Phe Leu Leu Ser Thr Arg	Ala Gly Gly Leu Gly Leu	Asn Leu Gln
1175	1180	1185
Thr Ala Asp Thr Val Ile	Ile Phe Asp Thr Asp Trp	Asn Pro His
1190	1195	1200
Gln Asp Leu Gln Ala Gln	Asp Arg Ala His Arg Ile	Gly Gln Lys
1205	1210	1215
Asn Glu Val Arg Ile Leu	Arg Leu Ile Thr Glu Asn	Ser Val Glu
1220	1225	1230
Glu Met Ile Leu Glu Arg	Ala His Lys Lys Leu Glu	Ile Asp Gly
1235	1240	1245
Lys Val Ile Gln Ala Gly	Lys Phe Asp Asn Lys Ser	Thr Ala Glu
1250	1255	1260
Glu Gln Glu Ala Met Leu	Arg Ala Leu Ile Glu Lys	Glu Asp Glu
1265	1270	1275
Arg Arg Gln Lys Gly Gly	Thr Asp Glu Glu Glu Glu	Asp Leu Asp
1280	1285	1290
Asp Asp Glu Leu Asn Gln	Ile Ile Ala Arg Asn Glu	Asn Glu Leu
1295	1300	1305
Val Val Phe Arg Lys Met	Asp Glu Glu Arg Tyr Leu	Ala Thr Lys
1310	1315	1320
Asn Ala Pro Tyr Pro Ser	Arg Leu Tyr Thr Glu Glu	Glu Leu Pro
1325	1330	1335
Glu Ile Tyr Lys Ile Asp	Pro Glu Glu Leu Phe Lys	Lys Glu Asp
1340	1345	1350
Val Ala Leu Glu Glu Tyr	Gly Arg Gly Ala Arg Glu	Arg Lys Ile
1355	1360	1365
Leu Gln Tyr Asp Asp Asn	Leu Thr Glu Glu Gln Trp	Leu Lys Lys
1370	1375	1380
Ile Glu Gly Met Val Ser	Asp Asp Asp Asp Asn	Asp Asp Asp
1385	1390	1395
Asp Gly Asn Val Asp Met	Ser Asp Ser Glu Met Glu	Ala Lys Pro
1400	1405	1410
Lys Lys Pro Lys Gly Arg	Arg Gly Arg Lys Pro Lys	Val Ala Arg
1415	1420	1425
Val Glu Asp Glu Glu Ser	Gln Thr Glu Ser Asp Val	Ile Ser Val
1430	1435	1440
Lys Arg Gln Phe Pro Glu	Asp Ala Asp Asp Phe Ile	Pro Pro Lys
1445	1450	1455
Arg Gln Lys Ser Ala Thr	Pro Gly Gly Thr Thr Thr	Ser Gly Arg
1460	1465	1470
Gly Arg Gly Arg Gly Arg	Gly Arg Gly Arg Gly Arg	Gly Arg Gly
1475	1480	1485
Arg Gly Arg Gly Ser Leu	Leu Ser Arg Tyr Thr Pro	Ser Val Asp
1490	1495	1500
Pro Leu Thr Ala Asp Glu	Arg Ser Thr Leu Gln Asn	Gln Ile Glu
1505	1510	1515

Asn Ile Leu Gly Leu Ile Ile Asn Tyr Lys Asn Glu His Asp Arg
 1520 1525 1530
 Val Leu Ser Glu Leu Phe Leu Val Lys Pro Pro Lys Arg Phe Tyr
 1535 1540 1545
 Pro Asp Tyr Tyr Val Leu Ile Lys His Pro Ile Ala Leu Asp Val
 1550 1555 1560
 Ile Lys Lys Arg Thr Ala Ser Lys Ser Tyr Ser Lys Ile Arg Glu
 1565 1570 1575
 Phe Leu Glu Asp Ile His Leu Met Phe Thr Asn Ala Lys Ile Tyr
 1580 1585 1590
 Asn Glu Glu Gly Ser Ile Val Tyr Gln Asp Ala Ala Phe Leu Glu
 1595 1600 1605
 Arg Leu Ser Met Asp Lys Phe Lys Glu Leu Ser Ala Asn Leu Ser
 1610 1615 1620
 Glu Asp Glu Ile Asn Lys Ile Leu Asp Phe Ala Glu Phe Asp Glu
 1625 1630 1635
 Met Phe Ser Leu Lys Pro Leu Val Pro Ser Thr Ala Ile Lys His
 1640 1645 1650
 Pro Ile Glu Ala Lys Leu Glu Lys Ile Asp Lys Gly Glu Ala Ile
 1655 1660 1665
 Asp Ser Pro Leu Leu Asn Ala Ser Thr Thr Ala Gly Thr Glu Glu
 1670 1675 1680
 Ser Thr Pro Ala Asn Phe Asp
 1685 1690

<210> 3
 <211> 24
 <212> DNA
 <213> 寡核苷酸

<400> 3
 atgaatcgtc aacctacaag agag 24

<210> 4
 <211> 27
 <212> DNA
 <213> 寡核苷酸

<400> 4
 tcaatcaaaa ttgctggtg tagactc 27

<210> 5
 <211> 32
 <212> DNA
 <213> 寡核苷酸

<400> 5
 cgggatccat gaatcgtaa cctacaagag ag 32

<210> 6
 <211> 30
 <212> DNA
 <213> 寡核苷酸

<400> 6
 gaagatctgt tgttgaagg catattgttg 30

<210> 7
 <211> 29
 <212> DNA
 <213> 寡核苷酸

<400> 7
 caggatccga acagaagagt ctacaccag 29

<210> 8
 <211> 31
 <212> DNA
 <213> 寡核苷酸

<400> 8
 acatgcatgc gttccacaag tgttctatac c 31

<210> 9
 <211> 27
 <212> DNA
 <213> 寡核苷酸

<400> 9 gggatccgg acacctacac caaaaca	27
<210> 10 <211> 28 <212> DNA <213> 寡核苷酸	
<400> 10 ccgctcgagc cattcacacc ctgccata	28
<210> 11 <211> 32 <212> DNA <213> 寡核苷酸	
<400> 11 tccccgcggt accttgagat ttggctgtaa ca	32
<210> 12 <211> 30 <212> DNA <213> 寡核苷酸	
<400> 12 tccccgcggc catgtctgat tgggtgaatg	30
<210> 13 <211> 30 <212> DNA <213> 寡核苷酸	
<400> 13 gaggatcctc tgacgacgat gatgacaatg	30
<210> 14 <211> 100 <212> DNA <213> 寡核苷酸	
<400> 14 acatgcatgc cttgtcatcg tcatccttgt aatcgatgtc atgatcttta taatcacctg	60
catggtcttt gtagtcatca aaatttgcgt gtgtagactc	100
<210> 15 <211> 32 <212> DNA <213> 寡核苷酸	
<400> 15 cgggatccat gaatcgtcaa cctacaagag ag	32
<210> 16 <211> 35 <212> DNA <213> 寡核苷酸	
<400> 16 cgggatcctc aatcaaaatt tgctggtgta gactc	35

atgaatcgtc	aacctacaag	agaggatatt	caaagagcga	ttcaacgctg	gcatcaaatg	60
aaacaacaat	atggagacca	agttcaactt	aatcctgaat	ttgttaaatt	aaccaagttc	120
ttgaataactt	tgaaaatgca	acaacaacgt	ttcagcaac	aataccaaca	acaacaaca	180
caacaacaac	agcagcagca	gcagcagcag	cagcaacaac	aacaacaaca	gcagcaacaa	240
caaagtctaa	accattcaca	acaatcgcca	ttgctacaaa	atgcacaggg	ccaaactcca	300
caacaacctc	caactcctca	acagttttct	aatttcaatc	agaacggtta	taatggtcaa	360
cagttttctt	ctcaagtaca	ctcacctgct	attgggtgggt	ctttatcgac	ttcaggacat	420
ggaaccccat	tagttacaaa	tgccaatttg	atgacaggaa	agaaaaatac	aaggacacca	480
aatggccaat	ttggttaacca	gacggctgct	ggaacaccat	tacaacaaca	acaacaacaa	540
caacaacaac	aacaacagcc	ttccctcat	ggaacaatt	cgaatcctat	gctaaaccag	600
acggctcagc	aaccaccaca	atctcaacaa	cgtaacaaa	accaaccacc	aagtccccaa	660
tcagcattca	ctaatacaaca	attccaatta	ttgaaatctc	agcttcaagc	attcaagtat	720
tttgtgagag	ctcctagtca	gggtcaaggt	caaataccac	aaaacttgat	agcatatggt	780
tcgaatccat	catctgctat	ggccaatgat	atgtacttac	cagcagtgaa	ccggcctcaa	840
actaatggta	tggatcgctac	aatgcaaatg	ccacaatcaa	tacctcaca	accacaacaa	900
tatgcccttc	aacaacaaaa	cgatttgctg	aaactgaaac	caaagtcaac	tggaggcacc	960
cctgaaatcc	cagaaaaaga	aaaaggaaag	cgtagaccta	aaccaagaa	tccgaaaaaa	1020
cctacaaga	aacagttgag	agaagaagaa	cagagacttg	cattggaaaa	acaaagacaa	1080
gaacttgaac	aaaatagact	caagagtagt	gctcctcaag	cattcccgcc	tcaagcaggt	1140
ttacaaggac	aagctccttt	cccaccacaa	ccaccacagc	agtcacaaca	acatgtacct	1200
caaccacctc	cgcatcaaac	ttcatctagt	ccaccgggtg	ggttgccaca	gctgcaaccg	1260
caacagcaac	agcaacagcc	atctcgcca	attactaaac	ctgccacacc	tcaaccttta	1320
ttccctgac	catctctcc	agtgaatata	aagagtgtag	taccgcacaa	agcaaacac	1380
aagaaggtaa	taataccggt	aactaaacca	aatattgaag	tagacacctt	cgagttatit	1440
gacattatca	gtgatgaggt	gaaagatata	ccgtttaata	ctttatatgc	tccacagagt	1500
agatttcaga	tcccttcgtt	tttgectgat	ggtataaata	tggagatat	ttatgtgaac	1560
agagaaggt	atatgcaaat	tacaatagaa	caagagaagg	agagattaag	aaaacaatit	1620
gacagtttga	atgaaaaaga	cactgaaaag	aaattggaac	ttgaaacaca	attaagtcaa	1680
ttggaattga	ttccttatca	gaaagattta	cgtaggtaaag	ttcttataca	atcttggttt	1740
gggaaatcat	tacttcttaa	ttcacatcca	aactttttag	caagattcag	ttcattatct	1800
atggacagtg	ttcatatgac	aacagattta	taccgactcc	aattggaatc	catgatgaga	1860
gaacaaaata	agaaacatgg	caaaactatt	gaagaaatca	taaatttcag	tgatcgaagt	1920
agcatcaaa	ccgtcaagaa	atcagaccgg	ttgtcaaggt	ttatgactaa	aattaataat	1980
ttccataatc	aaactgccaa	ggaagaacag	aaaaagttgg	aaaaaatggc	taaacacagt	2040
ttgcaagcat	tgaactgaa	tgatgaagaa	gcttatttga	aattgttggg	tcatacaaa	2100
gatacaagaa	ttaccattt	attagaacaa	acaaatcaat	ttttggactc	tttggtcttt	2160
gcagtgcaaa	gtcaacaaaa	agaggctcag	gacaatttag	catattcagg	tcgtgccata	2220
gaaccagcat	cagttgaacc	ccttgatgat	gagaagagag	aaaaaattga	ttattataat	2280
gttgctcata	gaattaaaga	agaagtcacc	aagcaacctt	caatattggt	tgggggtact	2340
ttgaaggagt	atcaattgaa	aggtttaca	tggatgggtt	cattgtttaa	taatcatttg	2400
aacggtatct	tggcagatga	gatgggtttg	ggtaaaacaa	ttcaaaactat	ttcattactc	2460
acatatcttg	tggaaagtga	aaaaattcct	ggtccatttt	tagtaattgt	tcccttatca	2520
acagtaacca	attggaatth	agaatttga	aaatgggctc	cctcaattaa	aaaaattacc	2580
tataaaggta	ctccaaatca	acgtaaagtg	atgcaacacg	atcagaac	cgggaaattc	2640
caattagtat	tgacgacatt	tgaatatggt	atnaagata	aaggattatt	gggtagaatc	2700
aaatgggtcc	atatgattat	tgatgaaggt	catcgatga	agaatgctaa	ttcgaatta	2760
tctgagacat	tgacacaaaa	ttaccatagt	gattatcggt	tgattttgac	tggtactcca	2820
ttgcaaaaata	acttaccaga	attatgggcc	ttgttaaatt	ttgttttacc	caaaattttc	2880
aactctgtga	aatcatttga	tgaatgggtc	aatacaccat	ttgccaatac	tgggtggtcaa	2940
gataagatag	aattgacaga	agaagaacaa	ttgttggtga	ttagaagatt	gcataaagtt	3000
ttaagaccgt	tccttttaag	aagtttaaag	aaagatgttg	aaaaagattt	accaaacaag	3060
gtggaaaaag	ttgtcaaatg	taaatcatcg	gcattgcaat	ctaaattata	tcaacaatg	3120
ttgaggtata	atatgttgta	tgctggagat	cctgccaatg	gatcagtgcc	cgttactata	3180
aaaaacgcc	acaatcaaat	aatgcaattg	aaaaaaattt	gtaatcacc	ttttgtttat	3240
gaagaagttg	agaatttgat	taatccta	attgaaacca	acgatcagat	ttggagagta	3300
gctggtaaat	ttgaattatt	agacaaagtc	ttacctaaat	tcaaaactac	tggctataaa	3360
gttttgattt	ttctccaaat	gactcaaat	atgaatatca	tggaaagact	tttacgattc	3420
agaggtatga	aatatatgag	attggatggt	ggaaccaaag	ctgatgatag	aactgattta	3480
ttgaaaagtt	ttaatgcccc	agattctgat	tatttttggt	ttcttttacc	cactcgtgct	3540
ggtgggttag	gtcttaattt	acaaactgcc	gatacagtca	ttatttttga	tactgattgg	3600
aatcctcatc	aagatttaca	agctcaagat	agagcccatc	gtattggtca	aaagaatgaa	3660
gttagaatat	taagattgat	tacggaaaac	tcggtggaag	aaatgatttt	ggaaagagct	3720
cataaaaagt	tggagattga	tggtaaagtg	attcaagccg	gtaaattcga	taacaaatcc	3780

图 1A

actgctgaag	aacaagaagc	tatgttaaga	gcattaatag	aaaaagaaga	tgaacgtaga	3840
cagaaaggtg	gtaccgacga	agaagaagaa	gatttggatg	atgatgaatt	gaatcaaatt	3900
attgccagaa	atgaaaacga	gttgggtggt	tttaggaaaa	tgatgaaga	aagatacctt	3960
gctaccaaga	atgctccata	cccattccaga	ttgtataccg	aggaagagtt	gcctgaaatt	4020
tataagatag	atccagaaga	acttttcaag	aaagaagacg	ttgcactgga	agagtatggt	4080
cgtggtgcta	gagaaaggaa	aatattacaa	tacgatgata	atttaactga	agagcaatgg	4140
ttgaagaaaa	ttgaaggtat	ggtatctgac	gacgatgatg	acaatgacga	tgacgatggc	4200
aatgttgata	tgagtgattc	tgaaatggaa	gccaaagccga	agaaaccgaa	aggaagaaga	4260
ggtcggaaac	ccaaggttgc	acgtgttgaa	gatgaagaaa	gtcaaaactga	atcggatggt	4320
atttctgtta	agcgtcaatt	tctgaagat	gcagatgact	ttataccacc	aaagagacag	4380
aaatctgcta	caccaggtgg	cactactact	tcgggtagag	gcagaggtag	gggccgtggt	4440
agaggtcgtg	gacgtggaag	aggcagaggc	agaggatctt	tgttgtctcg	ttatacacct	4500
tctgtagacc	ccttaactgc	agacgagaga	tctactttac	aaaatcaaat	tgaaaacata	4560
ttggggttga	tcattaatta	taagaatgaa	catgatagag	tattgagtga	attgtttttg	4620
gtgaagccac	ccaagagatt	ttaccctgat	tactatgttt	tgataaagca	tccaattgca	4680
cttgacgtga	ttaaaagag	aacagcgtca	aaatcttaca	gcaagattag	ggaatttttg	4740
gaggatattc	atthgatggt	caccaacgcc	aagatatata	atgaagaagg	ttcaattggt	4800
tatcaagatg	cagcattctt	ggaaagatta	tcaatggaca	aattcaaaga	attatcagcc	4860
aatctttcag	aagacgaaat	caataagatt	ttggatttgc	ctgagtttga	cgaaatgttt	4920
agtttaaac	cattagtctc	ttcaactgct	atcaaacatc	caattgaagc	taaattagag	4980
aaaatagata	aagggtgaagc	aattgacagt	ccattactca	atgcttctac	cactgctgga	5040
acagaagagt	ctacaccagc	aaatthgat	tga			5073

图 1A(续)

MNRQPTREDI	QRAIQRWHQM	KQYGDQVQL	NPEFVKLTKF	LNTLKMQQQR	FQQYQQQQQ	60
QQQQQQQQQQ	QQQQQQQQQQ	QSLNHSQQSP	LLQNAQQGTP	QQPPTPQQFS	NFNQNGYNGQ	120
QFSSQVHSPA	IGGSLSTSGH	GTPLVTNANL	MTGKKNTRTP	NAQFGNQATA	GTPLQQQQQQ	180
QQQQQQPPFH	GNNSNPMLNQ	TAQQPPQSQQ	RQQNQPPSPQ	SAFTNQFQL	LKSQQLQAFKY	240
FVRAPSQQGQ	QIPQNLIAVY	SNPSSAMAND	MYLPAVNRPQ	TNGMDRTMQM	PQSIPSPQQ	300
YALQQQNDLL	NLNPKSTGGT	PEIPEKKKKG	RGPKPKPKPK	PTKKQLREEE	QRLALEKQRQ	360
ELEQNRLKSS	APQAFPPQAG	LQGQAPPPPQ	PPQSQQHVP	QPPASTSSS	PPGGLPQLQP	420
QQQQQQPSRP	ITKPA TPQPL	FPDPSPPVNI	KSVVPDKANN	KKVIIPVTKP	NIEVDTFELF	480
DIISDEVKDI	PFNTLYAPQS	RFQIPSFPLD	GINMEDIYVN	REGYMQITIE	QEKERLRKQI	540
DSLNEKDTEK	KLELETQLSQ	LELIPYQKDL	RGKVLIQSWF	GKSLLPNSHP	NFLARFSSLS	600
MDSVHMTDDL	YRLQLESMMR	EQNKKHGKTI	EELINFSDRS	SIKAVKSDR	LSRFMTKINN	660
FHNQTAKEEQ	KKLEKMAKQR	LQALKLNDEE	AYLKLDDHTK	DTRITHLLEQ	TNQFLDSLAL	720
AVQSQQKEAQ	DNLAYSGRAI	EPASVEPLDD	EKREKIDYYN	VAHRIKEEVT	KQPSILVGGT	780
LKEYQLKGLQ	WMVSLFNNHL	NGILADEMGL	GKTIQTISLL	TYLVEVKKIP	GPFLVIVPLS	840
TVTNWNLEFE	KWAPSIKKIT	YKGTNPQRKV	MQHDIRTGNF	QLVLTTFEYV	IKDKLLGRI	900
KWVHMIIDEG	HRMKNANSKL	SETLTQNYHS	DYRLILGTGP	LQNNLPELWA	LLNFVLPKIF	960
NSVKSFDWEF	NTPFANTGGQ	DKIELTEEET	LLVIRRLHKV	LRPFLRLRLK	KDVEKDLPNK	1020
VEKVVCKKSS	ALQSKLYQQM	LRYNMLYAGD	PANGSVPVTI	KNANNQIMQL	KKICNHPFVY	1080
EEVENLINPN	IETNDQIWRV	AGKFELLDKV	LPKFKATGHK	VLIFFQMTQI	MNIMEDFLRF	1140
RGMKYMRLDG	GTKADDRTDL	LKSFNAPDS	YFCFLSTR	GGLGLNLQTA	DTVIFD TDW	1200
NPHQDLQAQD	RAHRIGQKNE	VRILRLITEN	SVEEMILERA	HKKLEIDGKV	IQAGKFDNKS	1260
TAEQEAMLR	ALIEKEDERR	QKGGTDEEEE	DLDDDELNQI	IARNENELVV	FRKMDDEERYL	1320
ATKNAPYPSR	LYTEELPEI	YKIDPEELFK	KEDVALEEY	RGARERKILQ	YDDNLTEEQW	1380
LKKIEGMVSD	DDDDNDDDDG	NVDMDSSEME	AKPKPKGR	GRKPKVARVE	DEESQTESDV	1440
ISVKRQFPED	ADDFIPPKRQ	KSATPGGTTT	SGRGRGRGRG	RGRGRGRGRG	RGSLLSRYTP	1500
SVDPLTADER	STLQNQIENI	LGLIINYKNE	HDRVLSLFL	VKPPKRFYPD	YVLIKHPA	1560
LDVIKRTAS	KSYSKIREFL	EDIHLMFTNA	KIYNEEGSIV	YQDAAFLERL	SMDKFKELSA	1620
NLSEDEINKI	LDFAEFDEMF	SLKPLVPSTA	IKHPIEAKLE	KIDKGEAIDS	PLLNASTTAG	1680
TEESTPANFD						1690

图 1B

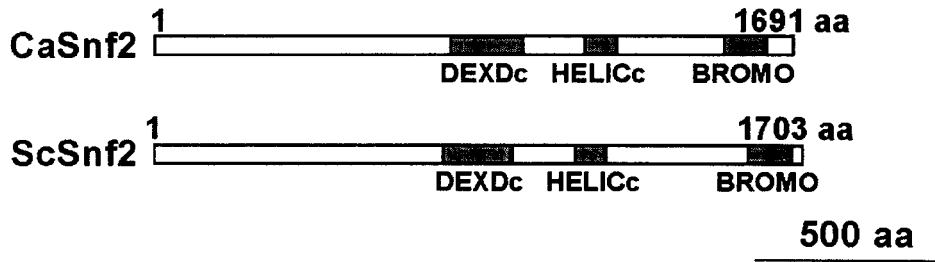


图 1C

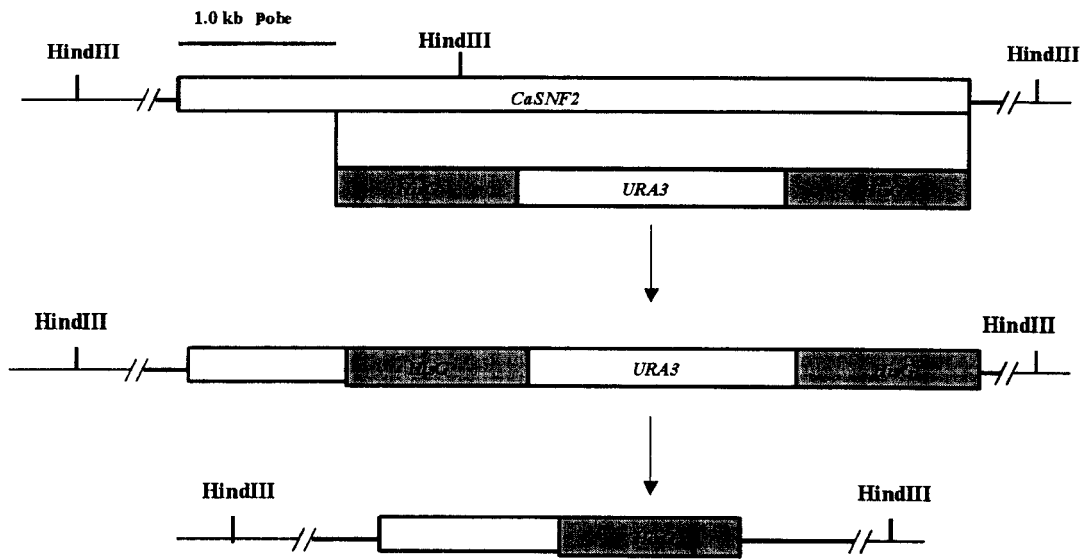


图 2A

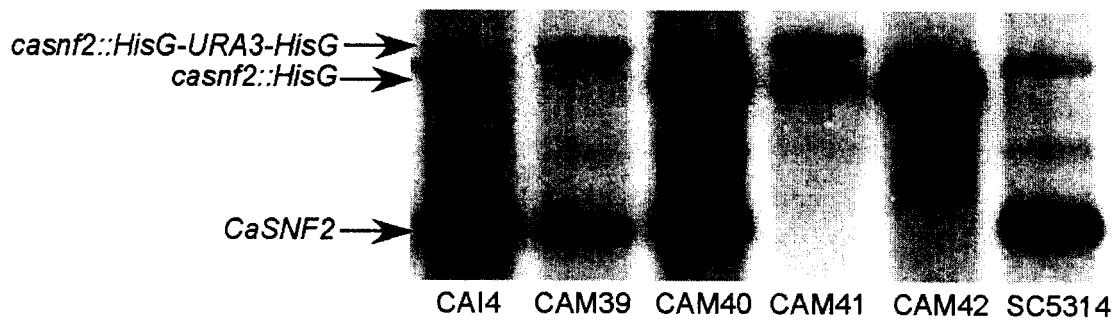


图 2B

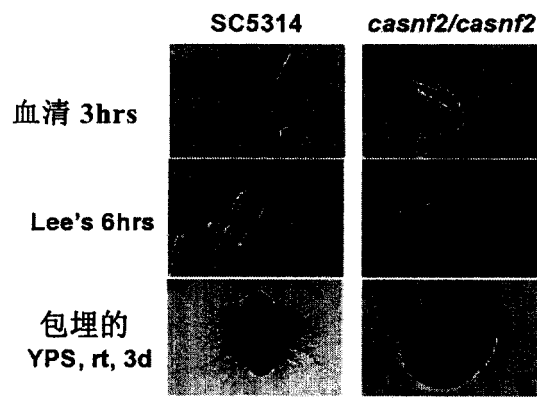


图 3

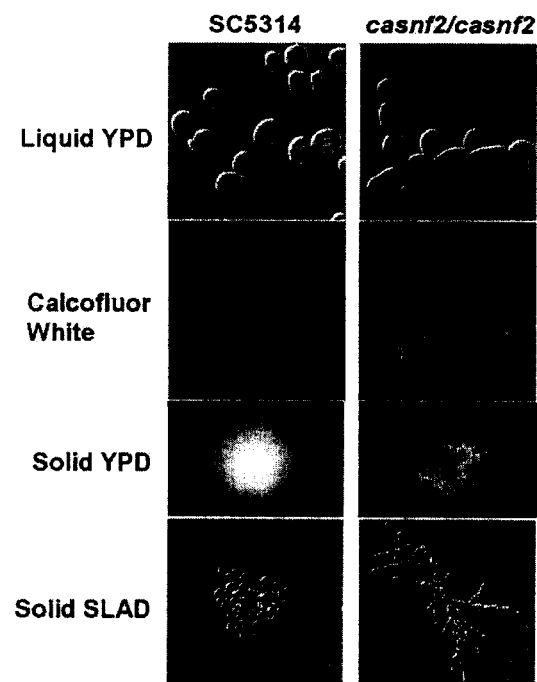


图 4

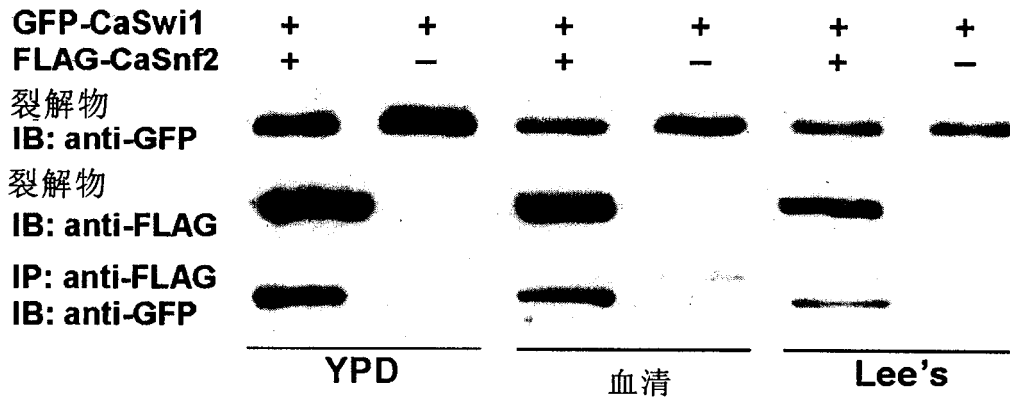


图 5

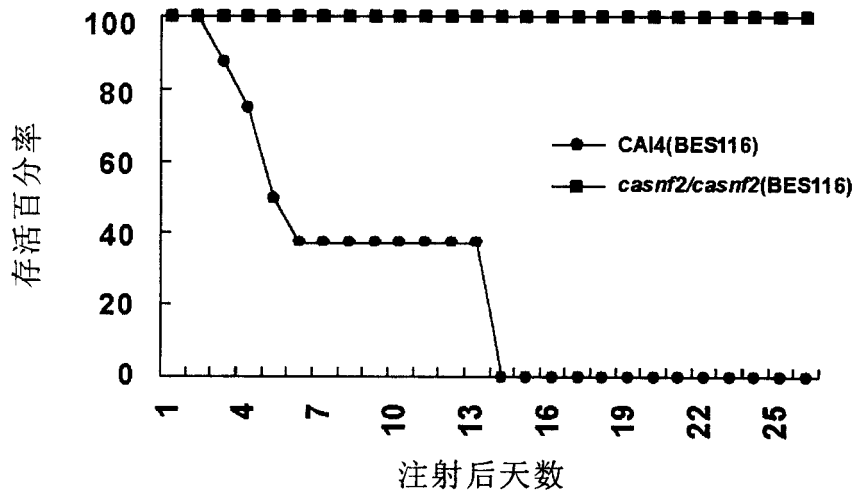


图 6

