

阴道毛滴虫半胱氨酸蛋白酶基因变异分析

贾万忠, 李志, 赵亮, 伦照荣*

【摘要】 目的 研究阴道毛滴虫半胱氨酸蛋白酶(TvCP)基因变异特点, 为分子种系发生和遗传进化等提供参考资料。方法 利用 PCR 扩增 *TvCP* 基因, 用 pBS-T 或 pET28b 载体构建重组质粒, 转化大肠埃希菌, 筛选阳性克隆, 测定目的基因序列。通过以序列相似性为基础的数据库搜索方法检索同源序列, 利用 Clustal X、DNASTar 等软件分析 *TvCP* 基因核苷酸和氨基酸序列多态性、遗传变异和进化规律。结果 克隆的 *TvCP2* 和 *TvCP3* 基因与数据库中相应参考序列在核苷酸和氨基酸水平上的相似性均为 99% 以上, 高度保守。克隆的 *TvCP4* 基因与 *TvCP4* 参考序列在核苷酸和氨基酸水平上的相似性为 97.5% 以上, 属于 *TvCP4* 基因家族的两个不同成员。根据阴道毛滴虫基因组数据库信息, 经分析表明 *TvCP4* 酶前体由 305 个氨基酸残基组成, 共有大小一致的 *TvCP4* 同源分子 9 个和 N-末端缺少部分氨基酸残基的同源分子 2 个, 它们之间的氨基酸序列相似性为 62.3%~96.7%。*TvCP4*、*TvCP1-3*、*TvCP12*、*TvCP25* 和半胱氨酸蛋白酶 65 (CP65) 之间也有同源性, 氨基酸序列相似性约为 61%~88.2%。此外, 阴道毛滴虫还拥有为数众多的其他组织蛋白酶 L 样成员, *TvCP* 构成呈高度多样性。序列联配和聚类分析表明 *TvCP* 酶活性位点、酶前体结构中 ERFNIN 样基序和形成二硫键的半胱氨酸残基数目和位置非常保守, 进化树分析也表明它们可能由共同的祖先分子通过基因复制和不断突变进化而来。结论 *TvCP* 组成一蛋白酶家族, 家族内各成员酶活性位点上残基高度保守, 但其他位点上的序列呈现高度多态性; 推测 *TvCP* 家族各成员可能由一个共同的祖先基因分子进化而来, 在保持酶基本功能的同时又赋予多样的功能。

【关键词】 阴道毛滴虫; 半胱氨酸蛋白酶基因; 变异; 进化树

中图分类号: R382.211 文献标识码: A

Genetic Variation and Clustal Analysis of *Trichomonas vaginalis* Cysteine Proteases

JIA Wan-zhong, LI Zhi, ZHAO Liang, LUN Zhao-rong*

(Center for Parasitic Organisms, School of Life Sciences, Zhongshan University, Guangzhou 510275, China)

【Abstract】 Objective To clone the genes coding for cysteine proteases (CPs, TvCPs) from *Trichomonas vaginalis* and to analyze their genetic variations with the related sequences from NCBI database (GenBank) and *T. vaginalis* Genome Project database from The Institute for Genomic Research (TIGR). **Method** *TvCP* genes were amplified using PCR, and inserted into vector pET28b or pBS-T. The recombinant plasmids were then transformed to *Escherichia coli* BL21 or Top10 strain. The recombinant plasmids were used for sequencing. Homologous *TvCP* genes were blasted based on NCBI GenBank and TIGR *T. vaginalis* Genome Project database. The sequences of cloned *TvCP* genes were aligned and clustered by Clustal X (1.83 version) with retrieved sequences. Comparisons of amino acids among cathepsin L-like *TvCPs*, human L-like cathepsins and papaya papain were performed using DNASTar software, and their phylogenetic tree was constructed based on neighbor-joining method using Clustal X. **Results** Two *TvCP3* clones and one *TvCP2* had a high identity of more than 99% with their responding *TvCPs*. Three clones of *TvCP4* genes, GZ-CP4-clone 1-3, belonged to two members of a family showing a high percentage identity of more than 97.5% with the sequences of *TvCP4* genes from databases (GenBank and TIGR) both at amino acid and nucleotide levels. Nine homologous *TvCP4* pro-enzymes with 304 amino acids and other two members with deletions of N-terminal sequence existed in *T. vaginalis* sharing a similarity of 62.3-96.7% amino acids, which may evolve by means of gene replication and deletion. *TvCP1-4*, *TvCP12*, *TvCP25* and CP65 had an identity of 61-88.2% at amino acid levels. So far, all reported sequences of C1 family from *T. vaginalis* belonged to capanthesin L-like subfamily with the same enzymatic active sites, conserved cysteine residues and similar structural features such as ERFNIN-like motif in pro-enzyme region, suggesting that they might result from gene

基金项目: 广东省自然科学基金 (No. 036606)

作者单位: 中山大学生命科学学院寄生生物研究中心, 广州 510275

* 通讯作者, E-mail: lsslzr@mail.sysu.edu.cn

duplication and mutations. **Conclusion** TvCPs belong to cathepsin L-like family with genetic diversity, but they have the same active amino acid residues, cysteine residues and similar structural characteristics, suggesting that they may derive from one ancestor.

【Key words】 *Trichomonas vaginalis*; Cysteine proteinase; Variation; Phylogenetic tree

Supported by the Natural Science Fund of Guangdong Province (No. 036606)

* Corresponding author, E-mail: lsslzr@mail.sysu.edu.cn

阴道毛滴虫(*Trichomonas vaginalis*)是一种兼性厌氧鞭毛虫,寄生于人体泌尿道、生殖道,可引起阴道毛滴虫病(trichomoniasis),该病的主要症状包括女性阴道炎、宫颈炎、尿道炎,男性前列腺炎及其他生殖道和尿道症状^[1-3]。流行病学研究结果显示,阴道毛滴虫病与子宫颈癌、人类免疫缺陷病毒(HIV)、乙型肝炎病毒(HBV)的感染以及早产等有密切关系^[1-3]。

半胱氨酸蛋白酶(cysteine proteinase, or cysteine protease, CP)是阴道毛滴虫重要致病因子之一,参与多种病理损伤及免疫逃避过程^[4-11]。阴道毛滴虫富含CP,至少在 23 种以上蛋白酶中大部分为CP,相对分子质量(M_r)约为 23 000~110 000^[8]。阴道毛滴虫CP(TvCP)主要由半胱氨酸蛋白酶A类(papain-like protease, 类木瓜蛋白酶)和半胱氨酸蛋白酶D类(legumin-like, 豆样蛋白酶)组成^[9,12]。迄今为止,已克隆的CA类TvCP基因均为组织蛋白酶L样亚族成员,共有7个(*Tvcp1*~4、*Tvcp12*、*Tvcp25*和*Tvcp65*),但亚族内成员的确切数目尚不十分清楚^[13]。新近发表的阴道毛滴虫基因组序列草图显示,阴道毛滴虫CP数量大、且种类繁多^[9]。为进一步了解阴道毛滴虫半胱氨酸蛋白酶基因的特征,本研究对TvCP2、TvCP3和TvCP4蛋白的基因序列进行测定,同时运用生物信息学方法对其基因的核苷酸及其编码蛋白的氨基酸序列进行分析。

材料与方 法

1 虫株、菌株与质粒

阴道毛滴虫临床分离株GZTV20(人型支原体阴性、对甲硝唑敏感)和GZTV29(人型支原体阳性、对甲硝唑不敏感),均采自广州市第二人民医院临床就诊患者。质粒pET28b(+)和大肠埃希菌(*Escherichia coli*)BL21(DE32)和Top10菌株,由本研究中心保存。质粒pBS-T购自北京天为时代生物公司。

2 主要试剂与酶

Taq DNA 酶、T₄ DNA 连接酶为宝生物工程(大连)有限公司产品,DNA 胶纯化试剂盒为上海生工生物工程有限公司产品,DNA 标志物(DL2000)为北京

天为时代生物公司产品。

3 TvCP 基因克隆

阴道毛滴虫基因组DNA提取、目的基因PCR扩增、质粒DNA提取及重组质粒DNA鉴定等,按参考文献[14]进行。根据已发表的TvCP基因序列^[13,15],用Oligo软件设计引物,用引物TvCP2-F(5'-ATG TTT GCT TTC TTG CTT TCT-3')和TvCP2-R(5'-TTA GAG AGC CTT TGG AAT GAT AGC-3')扩增完整开放性阅读框(ORF);用引物TvCP3-F(5'-TCG AAT TCG TTC TCT GCT TTC TTT GCT AC-3')和TvCP3-R(5'-AAC CTC GAG CTC TAC CTG TGG GAT GAT AA-3')扩增ORF,用引物TvCP4-F(5'-TCG AAT TCT ATT GCT AAC GCT GAC TGT GAC TGG-3')和TvCP4-R(5'-AA CTC GAG CTT GTC TTG AGG GAT GCA GGC CA-3')扩增成熟肽部分(下划线部分分别为EcoR I和Xho I酶切位点,引物由北京赛百胜生物公司合成)。扩增条件:94℃ 5 min;94℃ 30 s,55℃ 30 s,72℃ 90 s,35个循环;72℃ 10 min。PCR扩增产物在10 g/L琼脂糖凝胶中电泳,切下目的条带,用DNA胶纯化试剂盒回收纯化产物,将目的基因和质粒载体pET28b(+)DNA或pBS-T载体连接重组,经PCR鉴定为阳性重组质粒,阳性克隆由上海联合基因公司测序。用DNASTAR软件分析和比对相关序列,确定重组质粒序列的正确性。

4 TvCP 基因序列比较分析

利用DNASTAR软件,将测序得到的TvCP2、TvCP3和TvCP4基因序列及其编码氨基酸序列与美国基因组研究所(TIGR)*T. vaginalis*数据库(<http://www.tigr.org/tdb/e2kl/tvlg/>)和美国国立生物技术信息中心(NCBI)数据库(GenBank)(<http://www.ncbi.nlm.nih.gov/>)的相应参考序列(TvCP2-3的登录号分别为X77219、X77220,TvCP4的登录号分别为X77221、92602.m00073和96423.m00256)进行相似性和差异性比较。用ClustalX软件(1.83)对TvCP4基因进行多重序列比对和联配(alignment),并根据邻接法(neighbor-joining method)绘制TvCP4分子系统进化树,用自展法

(bootstrap test) 检验进化树^[16-18]。

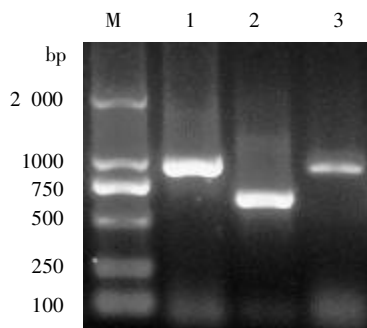
5 TvCP 中 L 样组织蛋白酶多重序列的比对和分子进化树构建

利用 Blast 检索工具从 GenBank 和 TIGR 中搜寻阴道毛滴虫 TvCP2-4 蛋白酶的同源序列, 用 ClustalX 软件进行多重序列比对和联配 (alignment), 根据邻接法 (Neighbor-joining method) 绘制 TvCP 分子系统进化树^[16-18]。

结 果

1 TvCP 基因克隆与酶结构特点分析

1.1 TvCP 基因克隆及序列测定 阴道毛滴虫基因组 DNA 扩增 TvCP 基因的电泳分析结果如图 1 所示, 目的片段大小与预期一致。



M: DNA 标志物 (DL2000), 1: *TvCP3*, 2: *TvCP4*, 3: *TvCP2*。
M: DNA marker (DL2000), 1: *TvCP3*, 2: *TvCP4*, 3: *TvCP2*。

图 1 *TvCP* 基因 PCR 扩增产物的电泳图谱

Fig. 1 Electrophoresis of PCR products of *TvCP* genes encoding *T. vaginalis* cysteine proteinase

含有 *TvCP* 基因的阳性克隆的测序结果表明, *TvCP2*、*TvCP3* 及 *TvCP4* 基因分别为 954、957 及 672 bp。*TvCP2* 基因与 GenBank 中 *TvCP2* 基因参考序列 (登录号为 X77219) 相比, ORF 第 402 位为 C 而非 G (即 $tcc \rightarrow tcg$, 同义密码子), 676 位为 C 而非 A ($cag \rightarrow aag$, 非同义密码子, 赖氨酸 K \rightarrow 谷氨酰胺 Q), 789 位为 C 而非 T (即 $ggc \rightarrow ggt$, 同义密码子)。*TvCP3* 基因序列与 *TvCP3* 基因参考序列 (登录号为 X77220) 相比, 1 个克隆的 ORF 第 66 位碱基为 T 而非 C ($ttt \rightarrow ttc$, 同义密码子), 另 1 个克隆的 ORF 第 66 位也为 T、第 774 位碱基为 G 而非 A (即 $tcg \rightarrow tea$, 同义密码子) 和第 382 位为 T 而非 C (即 $tcc \rightarrow ccc$, 非同义突变, 丝氨酸 S \rightarrow 脯氨酸 P)。

从 GenBank 和 TIGR *T. vaginalis* 数据库中检索的 *TvCP4* 基因序列, 初步比对和分析结果, 共得到 3 个明显不同的基因序列: ① CNCD147-CP4 (AY-679763,

X77221) 和 96423.m00256 (XM_001325970), ② 92602.m00073 (XM_001310116), ③ 90246.m00109 (XM_001316379) 和 CNCD147-CPT^[9,19]。本试验对 *TvCP4* 基因 3 个克隆进行了测序, *TvCP4*-clone1 与 *TvCP4* 参考序列 (CNCD147-CP4, X77221, AY679763) ORF 相比第 765 位 T 替代 C ($tgt \rightarrow tgc$, 同义突变)、第 826 位 G 替代 A ($gct \rightarrow act$, 非同义突变); *TvCP4*-clone2 第 289 位为 G 而非 A ($gaa \rightarrow aaa$, 非同义突变)、第 391 位 A 取代 C ($aac \rightarrow cac$, 非同义突变)、第 826 位 T 替代 A ($tca \rightarrow act$ 非同义突变); *TvCP4*-clone3 第 474 位 T 代替 C ($tat \rightarrow tac$, 同义突变)、第 477 位 C 代替 T ($gac \rightarrow gat$, 同义突变) 和第 826 位 T 代替 A ($tca \rightarrow act$, 非同义突变)。

氨基酸序列变异位点分析结果见表 1。由克隆的 *TvCP4* 基因片段核苷酸序列推导的氨基酸序列, 与数据库 *TvCP4* 基因参考序列构建的分子系统进化树见图 2。GZ-CP4-clone1 与 96423.m00256 氨基酸序列完全相同, 与 CNDC-CP4 只有 1 个氨基酸残基的差异 (1/198); GZ-CP4-clone2 分别与 CNDC-CP4 和 92602.m00073 相差 3 个和 4 个氨基酸残基 (3/198-4/198); GZ-CP4-clone3 与 CNDC-CP4 和 96423.m00256 均相差 5 个氨基酸残基 (5/198)。

表 1 *TvCP4* 蛋白氨基酸序列变异分析 (位点以全酶序列为准)
Table 1 Mutations of amino acid residues of *TvCP4* protein

半胱氨酸 蛋白酶 4 <i>TvCP4</i>	突变氨基酸残基位置 Sites of mutations of amino acid residue									
	19	81	97	10	12	13	16	16	19	27
92602.m00073	S	V	E	G	A	N	I	N	K	S
96423.m00256	G	A	K	D	S	T	V	K	K	A
CNDC-CP4	G	A	K	D	S	T	V	K	K	T
<i>TvCP4</i> -clone1	*	*	K	D	S	T	V	K	K	A
<i>TvCP4</i> -clone2	*	*	K	D	S	T	I	N	K	S
<i>TvCP4</i> -clone3	*	*	E	D	A	N	V	K	E	S

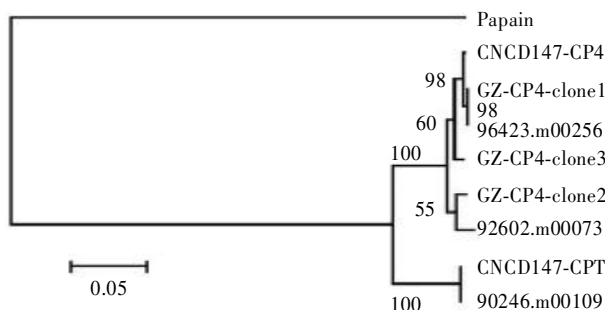
注: 92602.m00073 和 96423.m00256 来自 TIGR 数据库, CNDC-CP4 来自 GenBank, * 表示未测定。

Note: 92602.m00073 and 96423.m00256 from TIGR database, CNDC-CP4 from GenBank, * Not available.

2 TvCP 的数据库 Blast 检索、TvCP 聚类及变异特点分析

2.1 TvCP 的数据库 Blast 检索 阴道毛滴虫 (G3 株) 全基因组测序草图表明, C1 族半胱氨酸蛋白酶基因共 47 个, 其中 L 样组织蛋白酶 41 个, 与克隆的基因 *TvCP2*、*TvCP3* 和 *TvCP4* 以及已报道克隆的 *TvCP* 其他基因-*TvCP1* (X77218)、*TvCP12* (AY371180、AY-463697)、*TvCP25* (AY371181) 和 *TvCP65* (AY463696) 有较高相似性, 具有 L 样组织蛋白酶典型结构的基因有 26 个。

2.2 *TvCP4* 多态性、变异特点分析 在阴道毛滴虫



注：节点处数字为自展值（重复 100 次），标尺为遗传距离（%）；CNCD147-CPT 和 CNCD147-CP4 分别为 CNCD147 分离株的 CPT 和 CP4，90246.m00-109、92602.m00073 及 96423.m00256 分别为 TIGR 数据库 TvCP 基因，GZ-CP4-clone1、clone2 及 clone3 分别为阴道毛滴虫广州株 CP4 基因测序克隆 1、2、3。

Note: Numbers at nodes indicate the number of repeated times (100 repeats), the scales represent the genetic distance unit (percent divergence or percentage), papain used as an outgroup.

图 2 根据 TvCP4 分子第 95~293 位之间氨基酸残基应用 ClustalX 程序的邻接法构建的系统进化树

Fig. 2 Phylogenetic tree based on 199 amino acid residues (No.95-293, R-C) from TvCP4 constructed by ClustalX with the neighbor-joining method

(G3 株) 中，TvCP4 全酶氨基酸序列相同（305 个氨基酸残基），共有这样的 CP 分子 9 个，另外 TVAG_151460 和 TVAG_484350 在酶前体区 N 端分别缺少长度 20 个和 14 个氨基酸残基的一小段外，其氨基酸序列与 TvCP4 高度相似，它们共同组成一基因亚家族。TvCP4 成员间氨基酸序列相似性高达 62.3%~96.7%，为同源蛋白酶分子。氨基酸序列差异性和相似性比较见表 2。

表 2 与 TvCP4 大小相同的各成员间氨基酸序列变异性和相似性比较
Table 2 Divergence (%) and similarity (%) of amino acid residues of TvCPs with the same number of amino acid residues

		相似性 Identity percentage (%)										
TvCP 编号 No.		1	2	3	4	5	6	7	8	9	10	11
差异性 Divergence (%)	1	-	96.7	92.1	84.2	74.8	74.8	71.1	75.1	75.9	78.4	66.9
	2	3.4	-	90.8	84.9	73.8	74.1	70.8	73.8	74.9	78.0	66.6
	3	8.3	9.8	-	84.6	74.4	75.7	71.8	75.1	75.6	78.7	64.9
	4	17.8	16.9	17.3	-	75.8	71.6	68.1	71.6	73.7	75.4	66.7
	5	30.8	32.3	31.3	29.8	-	69.8	66.2	69.2	72.2	72.1	62.6
	6	30.8	31.8	29.4	36.3	38.5	-	86.9	78.0	74.9	75.1	64.3
	7	36.4	36.9	35.4	42.1	44.7	14.5	-	75.1	72.2	72.8	63.9
	8	30.3	32.4	30.3	35.7	39.6	26.0	30.3	-	77.3	77.4	64.9
	9	29.1	30.6	29.6	32.0	34.8	30.6	34.8	27.1	-	76.6	64.3
	10	25.6	26.0	25.1	30.3	34.8	30.3	33.8	27.0	28.1	-	65.6
	11	43.6	44.1	47.1	43.3	51.4	48.3	48.9	47.1	48.3	45.9	-

注：1~11 为阴道毛滴虫半胱氨酸蛋白酶 (TvCP) 基因，来自 TIGR 数据库；1: TVAG-467970, 2: TVAG-355480, 3: TVAG-298080, 4: TVAG-151460, 5: TVAG-437820, 6: TVAG-228450, 7: TVAG-043620, 8: TVAG-096740, 9: TVAG-484350, 10: TVAG-465470, 11: TVAG-164120; 除 TVAG_151460 和 TVAG-484350 外，其余 TvCP 分子均由 305 个氨基酸残基组成。

Note: No.1-11 represents TvCP molecules from TIGR database. TvCPs consist of 305 amino acid residues except for TVAG_151460 and TVAG-484350.

2.3 其他 TvCP 多态性、变异特点分析 其他 TvCP 与 TvCP4 大小接近，序列相似，均具有 L 样组织蛋白酶特征性的 ERFNIF 样基序和其他特征，但个别酶分子如 TVAG_228710 和 TVAG_146680 的酶催化活性位点 C 残基已发生变异，均突变为 G 残基。

2.4 TvCP 与人组织蛋白酶 L 比较和分子系统进化树构建 TvCP 与木瓜蛋白酶 (papain, AAB02650)、人组织蛋白酶 (cathepsinL 登录号分别为 M20496、NM_148979 和 NM_004390, cathepsin K 登录号为 AAH16058, cathepsin S 登录号为 NM_004079) 氨基酸序列大小相近，结构相似，酶活性位点和组成二硫键的半胱氨酸残基相当保守^[20-24]，它们之间氨基酸序列差异性和相似性比较结果见表 3。用邻接法根据 TvCP 活性位点 Q 和形成第 3 个二硫键最后一个 C 之间氨基酸序列构建的分子系统进化树(无根)见图 3。从中看出，人 L 样组织蛋白酶及木瓜蛋白酶聚在一起，大多数 TvCP 分子也聚在一起。而部分 TvCP 分子如 TVAG_228710、TVAG_146680、TVAG_325040 和 TVAG_267850 与人 L 样组织蛋白酶及木瓜蛋白酶聚在一起，反映了物种的形成和进化上亲缘关系。

讨 论

阴道毛滴虫 TvCP2 和 TvCP3 氨基酸序列在不同虫株间高度保守，TvCP4 基因由至少 9 个成员组成，大小相同(305 个氨基酸残基)，氨基酸和核苷酸相似性约 63% 以上。Mallinson 等^[25]认为，TvCP4 基因很可能是多拷贝的，因为该基因组成小家族并且各成员间高度相似，但不同探针在核酸杂交时可能会出现不同的结果。本实验未能获得与 90246.m00109 (TVAG_355480)CP4 基因相同的序列，是由于所用上游引物不能与其互补配对，却能与 TVAG_467970 (96423.m00256) 和 TVAG_355480 (92602.m00073) 两个成员的基因序列互补配对所造成。结果显示 GZ-CP4-clone2 和 clone3 序列间有 6 个氨基酸残基差异，这可能与使用的虫株未经克隆或者序列测定差异所致。但有趣的是 TvCP4 家族中 96423.m00256 和 92602.m00073 分子的氨基酸序列在不同虫株之间仅存在微小差异，如 96423.m00256 (G3 株)、CNDC-CP4 (CNDC-117 株) 和 TvCP4-clone1 (GZ 株) 之间相差 1~3 个氨基酸残基，92602.m00073 和 TvCP4-clone2 (GZ 株) 之间相差 4 个氨基酸残基。

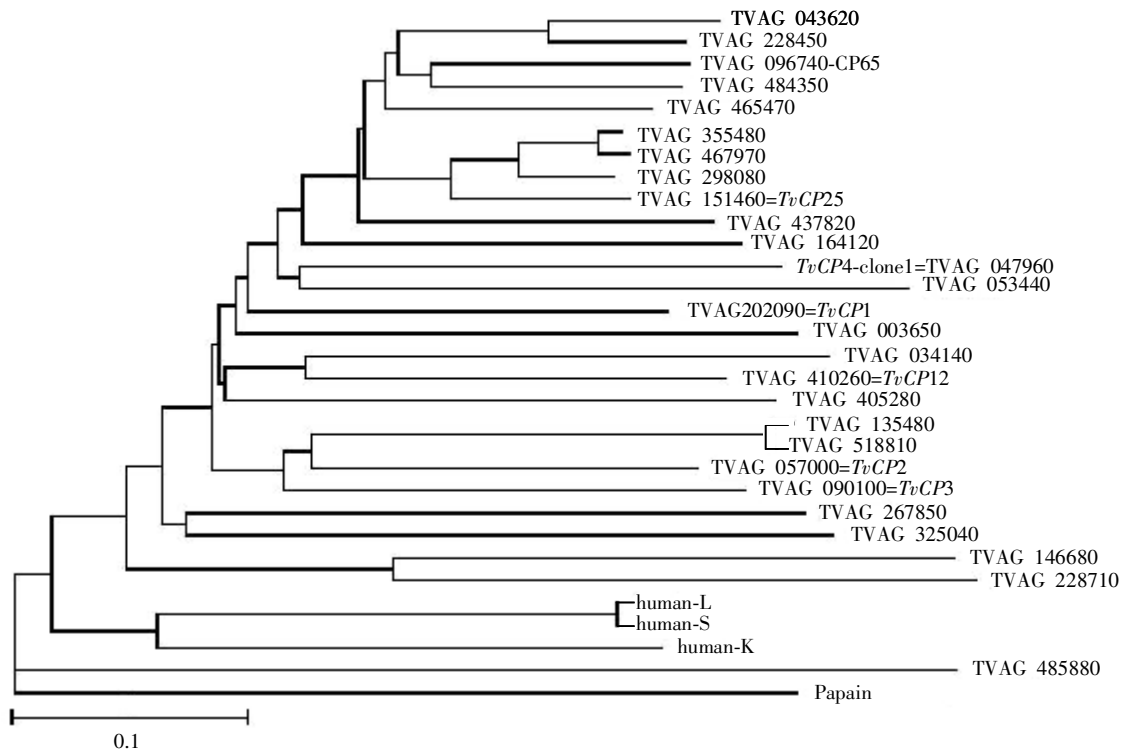
以往研究结果和本研究结果均显示，TvCP1~4、TvCP12、TvCP25 和 CP65 大小相近 (约为 300 个氨基酸残基)，氨基酸序列高度相似，如 TvCP4 与 TvCP25 氨基酸序列相似性达 88.2%，TvCP4 与 TvCP1、

表 3 TvCP 与人组织蛋白酶、木瓜蛋白酶氨基酸序列变异度 (%，对角线上方) 和相似性 (%，对角线下方) 比较
 Table 3 Divergence (%，above diagonal) and identity (%，below diagonal) of amino acids among TvCPs, human cathepsins and papaya papain

TvCP No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	-	73.2	66.2	61.8	58.8	62.4	64.0	68.9	68.9	69.3	64.9	65.4	62.3	62.3	65.4	61.4
2	43.0	-	70.6	65.4	58.8	65.9	69.7	73.7	73.2	71.9	70.2	68.0	65.4	67.1	69.3	62.7
3	56.1	45.8	-	66.2	60.5	63.9	64.5	69.3	69.3	69.3	65.8	68.9	64.5	64.7	67.5	60.1
4	68.4	58.8	56.8	-	85.5	64.9	66.2	68.9	72.4	71.9	68.0	65.4	61.8	62.7	69.3	62.3
5	76.3	76.3	71.0	57.2	-	61.5	59.6	61.8	62.7	61.8	61.0	58.3	57.9	59.6	61.0	57.0
6	67.9	58.6	63.9	61.9	70.0	-	79.5	77.6	78.0	77.1	80.0	75.1	70.7	71.2	71.2	61.5
7	62.9	48.8	61.9	57.9	75.0	30.4	-	80.3	81.1	79.4	80.7	76.8	70.2	71.9	72.8	64.9
8	50.8	40.2	50.1	51.5	68.5	34.1	28.9	-	93.9	88.2	83.8	79.4	71.1	73.7	79.8	65.8
9	50.8	41.1	50.1	43.6	66.0	33.2	27.3	7.9	-	89.5	83.8	79.4	70.6	71.9	82.0	66.7
10	49.2	43.3	49.4	43.9	67.8	34.3	29.9	15.4	13.5	-	89.5	80.7	69.7	73.2	79.4	67.1
11	60.6	47.8	58.5	53.6	71.0	29.6	28.1	22.9	22.9	24.5	-	78.5	74.1	72.8	76.8	66.7
12	59.5	52.9	51.1	60.1	79.1	38.9	35.4	30.4	30.4	27.5	32.1	-	69.7	71.1	70.6	64.0
13	67.8	59.5	61.9	69.8	80.6	48.2	49.1	47.1	48.1	49.4	40.6	50.1	-	66.2	68.9	63.6
14	67.8	55.0	61.9	67.2	75.0	47.1	45.2	41.5	45.2	41.8	43.3	47.1	58.5	-	68.0	62.7
15	58.3	48.8	53.2	49.4	89.8	46.1	42.4	28.9	25.1	29.1	34.5	47.1	51.1	53.2	-	67.1
16	69.1	65.3	73.0	67.2	82.0	70.8	60.8	58.5	56.3	54.6	56.3	63.1	64.3	66.7	55.9	-
17	70.9	69.7	82.0	82.8	93.8	96.5	86.5	79.1	73.7	78.4	79.1	83.5	85.0	77.7	73.7	86.5
18	88.2	91.5	98.2	81.3	104	82.3	83.5	77.7	74.9	74.9	77.7	74.9	98.2	89.7	77.7	89.7
19	113	113	114	102	125	122	121	123	116	121	116	128	123	119	124	122

注：表内数据为活性位点 Q 和形成二硫键的最后一个 C 残基之间氨基酸序列比较的结果；1：TVAG_518810，2：TvCP2，3：TvCP3，4：TvCP12，5：TVAG_034140，6：CP65，7：TVAG_228450，8：TvCP4，9：TVAG_298080，10：TvCP25，11：TVAG_465470，12：TVAG_437820，13：TVAG_047960，14：TVAG_164120，15：TvCP1，16：TVAG_003650，17：TVAG_267850，18：人组织蛋白酶 L，19：木瓜蛋白酶。

Note: Data are based on amino acid sequences between active site Q and the last C residue that forms the third -S-S- bond. No.1-17 stand for TvCPs from TIGR database. No. 18 is human cathepsin L and No. 19 is papaya papain.



注：标尺为遗传距离 (%)；Papain：木瓜蛋白酶，作为系统进化树的外组；human-L、K、S：人组织蛋白酶 L、K、S；其余为阴道毛滴虫半胱氨酸蛋白酶；TvCP2、TvCP3 和 TvCP4 基因为本试验克隆所得，其余 TvCP 分子来自 TIGR 数据库。

Note: The scale representing the genetic distance unit (percent divergence or percentage); Papain used as an outgroup. Human cathepsin L, K, S mean human-L, K, S, respectively; and the other sequences represent TVCPs from TIGR database. TvCP2, TvCP3 and TvCP4 gene from present study.

图 3 用邻接法构建的分子系统进化树
 Fig.3 Phylogenetic tree constructed using the neighbor-joining method by ClustalX

TvCP2 和 TvCP3 氨基酸序列相似性分别为 72%、65% 和 56%^[14], TvCP1 与 TvCP4 和 TvCP25 之间相似性超过 79%。因此有研究者推测, 这些基因分子可能是由一个共同祖先基因复制的结果^[25], 至少 *TvCP4* 基因家族各成员的形成可能是这种进化模式中的一个例子。在 G3 株 CP 基因中, 85377.m00025 (TVAG_151460) 与 TvCP25 分子只在催化区相差一个氨基酸残基, 相似性高达 98.2%, 与 *TvCP4* 基因家族成员间氨基酸水平上的相似性达 84%, 只是酶前体第 3~22 之间缺失一段多肽序列, TVAG_484350 与 TVAG_465470 和 TVAG_467970 相比, 也主要是 N 末端缺失一个由 14 个氨基酸残基组成的小片段, 这些均可看作典型缺失突变实例^[9], 充分体现阴道毛滴虫 CP 基因的多态性和复杂性, 可作为一种分子进化模式来研究。TvCP 中部分成员如 TVAG_146680、TVAG_228710 酶活性位点 Q 和 N 之间氨基酸序列与木瓜蛋白酶以及宿主——人的组织蛋白酶聚在一起, 表明两者在物种起源上的进化关系或寄生虫与宿主的协同进化。新近发表的阴道毛滴虫基因组序列草图显示, 阴道毛滴虫 CP 数量大且种类繁多, 仅 CA 类 C1 族成员多达 47 个, 其中主要为与本文研究直接相关的 L 样组织蛋白酶亚族各成员。CA 类 C19 族成员多达 117 个, 充分体现了 TvCP 的多样性^[9]。这些结果表明, 半胱氨酸蛋白酶可能在阴道毛滴虫的生存和对宿主致病方面具有重要的生物学功能, 可能与阴道毛滴虫生存和侵袭宿主细胞有重要关系。

总之, 对 *TvCP4* 基因的克隆, 序列检索、比对和聚类分析的结果表明, *TvCP* 基因具有明显的复杂性和多样性。*TvCP4* 基因家族有 3 个以上的成员, 各成员序列虽呈多态性, 组成不同的蛋白酶异形体, 但酶活性位点均高度保守。在此基础上, 它们还可能通过碱基缺失突变等形成新成员, 以增加阴道毛滴虫 CP 的多样性。此外, 根据 *TvCP4* 基因的变异分析结果可推测, *TvCP4* 基因形成较晚, 但进化速度快。本研究结果将为进一步研究 TvCP4 的细胞分布、分泌和功能等提供重要参考资料。

参 考 文 献

- [1] Petrin D, Delgaty K, Bhatt R, *et al.* Clinical and microbiological aspects of *Trichomonas vaginalis*[J]. Clin Microbiol Rev, 1998, 11 (2): 300-317.
- [2] Xie H, Huang HC, Pan CW, *et al.* The prevalence of *Trichomonas vaginalis* in male urethritis[J]. Chin J Parasit Dis Control, 2002, 15(3): 148-150. (in Chinese)
(谢辉, 黄慧聪, 潘长旺, 等. 男性尿道炎患者阴道毛滴虫感染情况调查[J]. 中国寄生虫病防治杂志, 2002, 15(3): 148-150.)
- [3] Van Belkum A, van der Schee C, van der Meijden WI, *et al.* A clinical study on the association of *Trichomonas vaginalis* and *Mycoplasma hominis* infections in women attending a sexually transmitted disease (STD) outpatient clinic[J]. FEMS Immunol Med Microbiol, 2001, 32(1): 27-32.
- [4] Ma L, Chen SY, Fu AH. Study on the infection situation of *Trichomonas vaginalis* in female STD patients[J]. Chin J Dermatovenereol, 2006, 20(1): 39-40. (in Chinese)
(马蕾, 陈世义, 付爱华. 女性性病患者阴道毛滴虫感染情况的分析[J]. 中国皮肤性病学杂志, 2006, 20(1): 39-40.)
- [5] Viikki M, Pukkala E, Nieminen P, *et al.* Gynecological infections as risk determinants of subsequent cervical neoplasia[J]. Acta Oncol, 2000, 39(1): 71-75.
- [6] Sorvillo F, Kerndt P. *Trichomonas vaginalis* and amplification of HIV-1 transmission[J]. Lancet, 1998, 351(9097): 213-214.
- [7] Jacobs B, Mayaud P, Chantalucha J, *et al.* Sexual transmission of hepatitis B in Mwanza, Tanzania[J]. Sex Transm Dis, 1997, 24(3): 121-126.
- [8] Neale KA, Alderete JF. Analysis of the proteinases of representative *Trichomonas vaginalis* isolates[J]. Infect Immun, 1990, 58 (1): 157-162.
- [9] Carlton JM, Hirt RP, Silva JC, *et al.* Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*[J]. Science, 2007, 315(5809): 207-212.
- [10] Rovenzano D, Alderete JF. Analysis of human immunoglobulin-degrading cysteine proteinases of *Trichomonas vaginalis*[J]. Infect Immun, 1995, 63(9): 3388-3395.
- [11] Hernández-Gutiérrez R, Avila-González L, Ortega-López J, *et al.* *Trichomonas vaginalis*: Characterization of a 39-kDa cysteine proteinase found in patient vaginal secretions[J]. Exp Parasitol, 2004, 107(3-4): 125-135.
- [12] Sajid M, McKerrow JH. Cysteine proteases of parasitic organisms [J]. Mol Biochem Parasitol, 2002, 120(1): 1-21.
- [13] Solano-González E, Alvarez-Sánchez ME, Avila-González L, *et al.* Location of the cell-binding domain of CP65, a 65 kDa cysteine proteinase involved in *Trichomonas vaginalis* cytotoxicity[J]. Int J Biochem Cell Biol, 2006, 38(12): 2114-2127.
- [14] Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: a Laboratory Manual[M]. Translated by Jin DY, Li MF. 2nd ed. Beijing: Science Press, 1999. 672-692. (in Chinese)
(萨姆布鲁克 J, 弗里奇 EF, 曼尼蒂斯 T. 分子克隆实验指南[M]. 金冬雁, 黎孟枫译. 第 2 版. 北京: 科学出版社, 1999. 672-692.)
- [15] León-Sicairos CR, León-Felix J, Arroyo R. *Tvcp12*: a novel *Trichomonas vaginalis* cathepsin L-like cysteine proteinase gene[J]. Microbiology, 2004, 150(5): 1131-1138.
- [16] Thompson JD, Higgins DG, Gibson TJ, *et al.* Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice[J]. Nucl Acid Res, 1994, 22(22): 4673-4680.
- [17] Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap[J]. Evolution, 1985, 39 (4): 783-791.
- [18] Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees[J]. Mol Biol Evol, 1987, 4(4): 406-425.
- [19] Sommer U, Costello CE, Hayes GR, *et al.* Identification of *Trichomonas vaginalis* cysteine proteases that induce apoptosis in human vaginal epithelial cells[J]. Biol Chem, 2005, 280(25): 23853-23860.

(下转第 202 页)

- dia*[J]. Mol Biochem Parasitol, 1997, 86(1): 13-27.
- [7] Svard SG, Meng TC, Hetsko ML, *et al.* Differentiation-associated surface antigen variation in the ancient eukaryote *Giardia lamblia* [J]. Mol Microbiol, 1998, 30(5): 979-989.
- [8] Nash TE. Antigenic variation in *Giardia lamblia* and the host's immune response[J]. Philos Trans R Soc Lond B Biol Sci, 1997, 352 (1359): 1369-1375.
- [9] Lu SQ, Wang ZY, Yan G, *et al.* Four isolates of *Giardia lamblia* cultivated axenically in China and the restriction endonuclease analysis of their DNA[J]. J Parasitol, 1996, 82(4): 659-661.
- [10] Keister DB. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile[J]. Trans R Soc Trop Med Hyg, 1983, 77(4): 487-488.
- [11] Mowatt MR, Aggarwal A, Nash TE. Carboxy-terminal sequence conservation among variant-specific surface proteins of *Giardia lamblia*[J]. Mol Biochem Parasitol, 1991, 49(2): 215-217.
- [12] Erlich HA, Gelfand D, Sninsky JJ. Recent advances in the polymerase chain reaction[J]. Science, 1991, 252: 1643-1651.
- [13] Nash TE, Mowatt MR. Characterization of a *Giardia lamblia* variant-specific surface protein (VSP) gene from isolate GSM and estimation of the VSP gene repertoire size[J]. Mol Biochem Parasitol, 1992, 51(2): 219-227.
- [14] Ey PL, Khanna KM, Mayrhofer G, *et al.* A gene encoding a 69 kDa major surface protein of *Giardia intestinalis* trophozoites[J]. Mol Biochem Parasitol, 1993, 58(2): 247-257.
- [15] Monis PT, Andrews RH, Mayrhofer G, *et al.* Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin[J]. Infect Gene Evolu, 2003, 3(1):29-38.
- [16] Sara R, Hayman D, Nash TE. Genetic manipulation of *Giardia lamblia*[J]. Mol Biochem Parasitol, 2002, 122(1): 1-7.
- [17] Kulakova L, Singer SM, Nash TE, *et al.* Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation [J]. Mol Microbiol, 2006, 61(6): 1533-1542.

(Received: 2007-09-03 Editors: YU Sen-hai, GAO Shi)

(上接第 196 页)

- [20] Gal S, Gottesman MM. Isolation and sequence of a cDNA for human pro-(cathepsin L)[J]. Biochem J, 1998, 253(1): 303-306.
- [21] Joseph LJ, Chang LC, Stamenkovich D, *et al.* Complete nucleotide and deduced amino acid sequences of human and murine preprocathepsin L. An abundant transcript induced by transformation of fibroblasts[J]. J Clin Invest, 1998, 81(5): 1621-1629.
- [22] Cohen LW, Coghlan VM, Dihel LC. Cloning and sequencing of papain-encoding cDNA[J]. Gene, 1986, 48(2-3): 219-227.
- [23] Strausberg RL, Feingold EA, Grouse LH, *et al.* Generation and initial analysis of more than 15 000 full-length human and mouse cDNA sequences[J]. Proc Natl Acad Sci USA, 2002, 99(26): 16899-16903.
- [24] Shi GP, Munger JS, Meara JP, *et al.* Molecular cloning and expression of human alveolar macrophage cathepsin S, an elastolytic cysteine protease [J]. J Biol Chem, 1992, 267(11): 7258-7262.
- [25] Mallinson DJ, Lochwood BC, Coombs GH, *et al.* Identification and molecular cloning of four cysteine proteinase genes from the pathogenic protozoan *Trichomonas vaginalis*[J]. Microbiology, 1994, 140(10): 2725-2735.
- [26] Scott DA, North MJ, Coombs GH. The pathway of secretion of proteinases in *Trichomonas vaginalis*[J]. Int J Parasitol, 1995, 25 (6): 657-666.
- [27] Mendoza-Lopez MR, Ecerril-Garcia C, Fattel-Facenda LV, *et al.* CP30, a cysteine proteinase involved in *Trichomonas vaginalis* cytoadherence[J]. Infect Immun, 2000, 68(9): 4907-4912.
- [28] Alvarez-Sanchez ME, Avila-Gonzalez L, Becerril-Garcia C, *et al.* A novel cysteine proteinase (CP65) of *Trichomonas vaginalis* involved in cytotoxicity[J]. Microbiol Pathogiol, 2000, 28(4): 193-202.
- [29] Alderete JF, Newton E, Dennis C, *et al.* The vagina of women infected with *Trichomonas vaginalis* has numerous proteinases and antibodies to trichomonad proteinases[J]. Genitourin Med, 1991, 67(6): 469-474.

(收稿日期: 2007-11-08 编辑: 富秀兰)

**感谢寄生虫病科研、防治、教育工作者
多年来对本刊的大力支持！欢迎继续投稿！**