

网络出版时间:2025-11-12 17:55:52 网络出版地址:<https://link.cnki.net/urlid/34.1065.R.20251202.1333.013>

艰难梭菌毒素 B 基因型生物信息学分析及制备

方一泰^{1,2}, 宁年智², 孙娅坤², 李德雨², 王慧², 罗德炎^{1,2}

(¹安徽医科大学基础医学院, 合肥 230032;

²军事医学研究院病原微生物生物安全全国重点实验室, 北京 100071)

摘要 目的 针对国内艰难梭菌毒素 B(TcdB)基因型的生物信息学分析,并通过原核表达制备 TcdB,为有效疫苗的开发提供数据支持。**方法** 利用多个生物信息学软件,如 Snippy, Blast, Muscle, R 语言下的 dist.alignment() 函数与 helust() 函数,对 NCBI Genbank 数据库中中国境内的 1 355 株艰难梭菌基因组进行比对、分析,构建系统发育树,并对其 TcdB 进行分群、建树。使用在线生物信息学分析网站,对占比最大的 TcdB1 与 TcdB2 亚群的代表性菌株的蛋白质结构及抗原表位进行预测并分析。通过构建重组质粒,使用原核表达系统,纯化抗原蛋白 TcdB。**结果** 根据 TcdB 的基因型,国内 1 355 株艰难梭菌大抵上可分为 12 个亚型,以 TcdB1 与 TcdB2 为主要亚型,占所有分离得到的菌株数量的 93.94% 以上;约 17.20% 的菌株为非产毒菌株或缺乏 TcdB,单独表达艰难梭菌毒素 A。TcdB1 与 TcdB2 的抗原表位分布基本一致,多分布于蛋白 C 端重复组合寡肽结构域外。**结论** 成功建立了一种专门针对国内 TcdB 的分型系统,可对其主要亚型的代表性菌株进行抗原表位的预测,并对筛选后的 TcdB 进行重组表达制备。

关键词 艰难梭菌; 毒素 B; 序列分析; 生物信息学分析; 表位预测; 重组表达

中图分类号 R 378.8

文献标志码 A 文章编号 1000-1492(2025)12-2273-08

doi:10.19405/j.cnki.issn1000-1492.2025.12.010

艰难梭菌是一类专性厌氧型细菌^[1], 艰难梭菌毒素 B(*Clostridioides difficile* toxin B, TcdB) 是其最重要的毒力因子^[2-3]。既往针对艰难梭菌疫苗的研究多停留在临床试验阶段或失败^[4-5]。笔者推测在菌株和抗原表位的选择上,是否存在不合理性。如,传统实验方法对复杂构象表位的识别能力有限,可能导致关键中和表位的遗漏,从而影响到对免疫原性的评估。得益于生物信息学技术的快速发展,多种在线分析工具可通过 AI 模型对抗原蛋白的理化性质、空间结构、抗原表位等进行分析与预测^[6-9],从而筛选出潜在的疫苗候选表位,加快疫苗研发进程。此外,已经有大量的艰难梭菌基因组数据被上传至多个公共数据库中以供查询,也有文献对这些数据作了系统性的分析^[10-11],但聚焦在中国,则缺少艰难梭菌的流行情况及详情信息。该研究计划通过生物信息学方法,对中国艰难梭菌的流行情况进行分析,对其分泌的 TcdB 进行分群,同时对主流亚

型的优势抗原表位进行预测,并对大分子毒素蛋白进行重组表达,为疫苗研究提供新思路。

1 材料与方法

1.1 基因组数据收集 登录 NCBI 网站,点击进入 Pathogen Detection (<https://www.ncbi.nlm.nih.gov/pathogens/>),输入关键词“*Clostridium difficile*”并检索,下载所有艰难梭菌基因组信息表,总计 28 575 株。根据表中所显示的“Location”信息,单独挑选出包括港澳台在内的中国地区,总计 1 485 株。由于部分菌株($n = 130, 9.59\%$)相关信息缺失,故不纳入后续分析。

1.2 艰难梭菌系统发育分析 使用 Snippy 提取单核苷酸多态性,以生成核基因组比对。使用 Fasttree2 将核基因组比对生成最大似然系统发育树(将 VPI10463 作为参考序列)。建树结果使用 iTOL (<https://itol.embl.de/>)^[12] 加工美化后展示。

1.3 多位点序列分型 (multi-locus sequence typing, MLST) 分析 MLST 是一种基于核酸序列测定的细菌分型方法。登录在线网站 PubMLST (<https://pubmlst.org/organisms/clostridioides-difficile>), 使用 Oxford 分类法,将每一株艰难梭菌氨基酸序列输入后,即可获得其对应的序列类型 (sequence

2025-08-11 接收

基金项目:病原微生物生物安全全国重点实验室自主研究课题(编号:SKLPBS2228)

作者简介:方一泰,男,硕士研究生;

罗德炎,女,副研究员,硕士生导师,通信作者, E-mail:
ldy612@126.com

type, ST)。

1.4 艰难梭菌 TcdB 氨基酸序列收集 将 VPI10463 TcdB 氨基酸序列作为参考序列, 使用 Blast, 默认参数, 分别比对下载的每一株艰难梭菌基因组氨基酸序列。依据“identity”获取每一株艰难梭菌 TcdB 编号, 使用 Seqkit, 根据序列名称提取每一株艰难梭菌 TcdB 氨基酸序列。

1.5 艰难梭菌 TcdB 进化分析 将下载得到的所有艰难梭菌 TcdB 氨基酸序列合并为一个“. fasta”格式文件。使用 Muscle, 默认参数, 将各条序列比对对齐, 再采用 fasttree2 绘制进化树, 建树结果使用 iTOL^[12] 展示。

1.6 序列聚类与分析 将对齐后的序列输入 R 环境中, 使用 seqinR package 中的 dist.alignment() 函数将 multiple sequence alignment 多序列比对转化为距离矩阵 distance matrix。使用 hclust() 函数, 采用平均连锁作为聚类方法, 并选择聚类阈值。根据阈值大小, 结合进化分析, 即可对 TcdB 进行分群, 生成 TcdB 亚型^[10-11]。将分型结果映射到进化树上, 使用 iTOL^[12] 加工美化后展示。

1.7 艰难梭菌 TcdB 生物信息学分析

1.7.1 理化性质预测 登录 EXpasy (<https://web.expasy.org/protparam/>)^[6], 输入 TcdB 的氨基酸序列, 即可得到其理化性质的分析预测信息。

1.7.2 信号肽及跨膜区预测 登录 DTU Health Tech (<https://services.healthtech.dtu.dk/>), 选择 SignalP-6.0 工具, 输入 TcdB 的氨基酸序列, 即可得到其信号肽及位置的分析预测信息。选择 TMHMM-2.0 工具 (<https://services.healthtech.dtu.dk/services/TMHMM-2.0/>), 即可对该蛋白跨膜区进行预测。

1.7.3 空间结构预测 登录 NPS@ (<https://npsa-pbil.ibcp.fr/>), 选择 SOPMA Protein Secondary Structure Prediction 功能, 输入 TcdB 的氨基酸序列, 即可对蛋白质的二级结构进行预测。

1.7.4 B、T 细胞表位预测 登录 ABCpred (https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html) 与 SYFPEITHI (<http://www.syfpeithi.de/bin/MHCServer.dll/EpitopePrediction.htm>)^[7-8], 输入 TcdB 的氨基酸序列, 即可对其 B、T 细胞表位进行预测。登录 VaxiJen v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>)^[9], 输入分析得到的优势表位序列, 可对其抗原性进行分析。

1.8 抗原蛋白 TcdB 的原核表达及纯化

1.8.1 主要试剂 氨苄青霉素储存液、异丙基-β-D-硫代半乳糖苷 (isopropyl β-D-1-thiogalactopyranoside, IPTG)、BCA 蛋白浓度测定试剂盒 (货号: A1170、I8070、PC0020) 购自北京索莱宝科技有限公司。BL21 (DE3) 化学感受态细胞 (货号: CD601) 购自北京全式金生物技术有限公司。HisTrapTM FF Crude (货号: 17528601) 购自美国 Cytiva 公司。兔源性抗 TcdB 单克隆抗体 (1 : 1 000)、HRP 偶联羊抗兔 IgG 抗体 (1 : 10 000) (货号: ab270452、ab205718) 购自英国 Abcam 公司。

1.8.2 主要仪器 化学发光成像仪 (型号: MiniChemi830) 购自北京赛智创业科技有限公司, 细胞破碎仪 (型号: Mini pro) 购自广州聚能纳米生物科技股份有限公司, 振荡培养箱 (型号: MQD-S3R) 购自上海曼泉仪器有限公司, PCR 仪 (型号: T100TM Thermal Cycler) 与半干转印槽 (型号: 221BR Trans-Blot® SD Cell) 购自美国 BIO-RAD 公司, CO₂ 培养箱 (型号: MCO-15AC) 购自日本 SANYO 公司, 电泳仪 (型号: JY600E) 购自北京君意东方电泳设备有限公司, 蛋白质纯化仪 (型号: AKTA pure) 购自美国 GE HealthCare 公司。

1.8.3 TcdB 重组质粒的设计、表达及蛋白纯化 登录在线网站 NCBI, 从数据库中下载艰难梭菌 VPI10463 TcdB 基因序列 (登录号: KC292162.1), 并在其 C 端插入 10 × His 序列; 经密码子优化后连接至 pET22b (+) 空载体中, 构建重组质粒。通过化学热激法将质粒转化入大肠埃希菌 BL21 (DE3) 细胞系中; 针对 pET22b (+) 设计特异性 PCR 引物, 扩增并验证重组质粒。挑取阳性菌落进行扩大培养, 发酵量为 2 L。37 °C, 220 r/min 振荡培养至菌液在 600 nm 波长处吸光度数值为 0.6 左右时, 加入终浓度为 0.1 mmol/L IPTG, 20 °C, 200 r/min 过夜诱导表达。收集菌体并进行超声破碎。离心后的上清液使用 HisTrap FF Crude 5 mL 预装填镍柱通过 AKTA 纯化仪进行纯化。梯度洗脱产物经透析后, 使用 BCA 试剂盒检测浓度。

1.8.4 Western blot 实验 将洗脱产物与上样缓冲液混合后进行 SDS-PAGE; 使用半干转印槽, 300 mA, 30 min 将条带转印到 PVDF 膜上。含 5% 脱脂奶粉的 TBST 室温封闭 2 h 后, 与抗 TcdB 单克隆抗体共孵育, 4 °C 过夜。使用 TBST 洗涤后, 与 HRP 偶联的抗兔 IgG 抗体共孵育, 37 °C, 30 min。重复洗涤, 加入 ECL 化学显色液, 并于凝胶成像系统上显

示条带并记录。

2 结果

2.1 艰难梭菌 TcdB 序列数据库的构建与 TcdB 的分群 1 355 株艰难梭菌基因组信息中, 1 122 株 (82.80%) 艰难梭菌表达 TcdB, 其余 233 株 (17.20%) 艰难梭菌为非产毒菌株或缺失 TcdB, 单独表达艰难梭菌毒素 A。使用平均连锁作为聚类方法, 通过层次聚类将 1 122 株艰难梭菌编码表达的 TcdB 分为不同的亚型, 并将 12 个不同的 TcdB 亚型标记为 TcdB1 ~ TcdB12 (图 1)。TcdB1 是其中最大的亚型, 包含了 834 条序列, 约占总序列数的 74.33%, 其中就包含了经典菌株 C. difficile 630 与 VPI10463; TcdB2 包含了 220 条序列, 占总序列数的 19.61%; TcdB3 占总序列数的 1.87%; TcdB4 占总序列数的 1.70%; TcdB5 中含有 16 条序列; TcdB6、7 中分别包含了 5 株与 2 株艰难梭菌; 其余 5 种亚型中的菌株数量极少。

通过比较每一种亚型所包含的所有菌株测序结果完整性的高低, 在每一种亚型中挑选一条作为该亚型下的代表性序列, 构建邻接系统发育树 (图 2)。使用 Blast 对各亚型下代表性菌株进行比对, 氨基酸差异大小如表 1 所示, 部分 TcdB 亚型间同源度较高, 如 TcdB1、TcdB8、TcdB11。其中, TcdB4 与 TcdB6, TcdB5 与 TcdB7 这两对亚群间氨基酸差异性较小。此外, 部分亚型如 TcdB2、TcdB9 与其它 TcdB 亚型间差异较大。

2.2 TcdB 各亚型在艰难梭菌系统发育中的分布

系统发育树 Oxford ST 分型结果显示: ST3 ($n = 182$, 13.43%)、ST35 ($n = 179$, 13.21%)、ST37 ($n = 172$, 12.69%)、ST54 ($n = 162$, 11.96%) 与 ST2 ($n = 125$,

9.23%) 是中国流行的主要 ST 型。毒素表达情况与 ST 分型也显示出一定的对应关系: 例如, ST39 型均为非产毒菌株; ST37/ST81 型均为 TcdB2; ST3 型中也存在一小部分非产毒菌株; TcdB1 主要为 ST3、ST35、ST54、ST2 型。2015 年后, 艰难梭菌感染

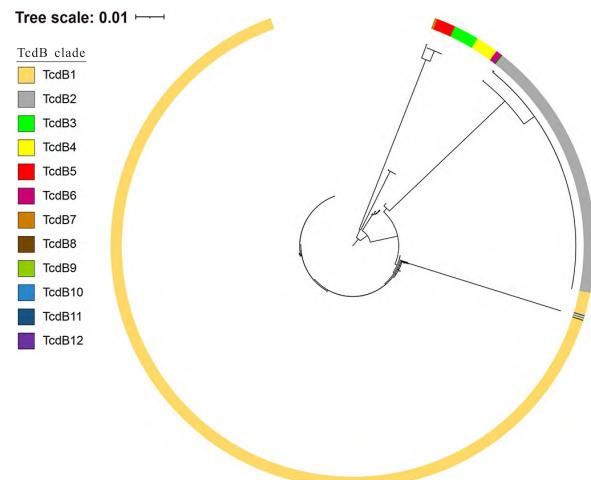


图 1 TcdB 亚型最大似然树

Fig. 1 Maximum likelihood tree of TcdB subtypes

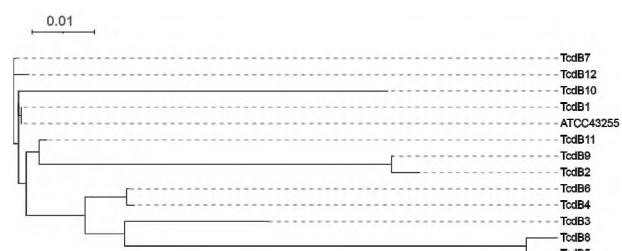


图 2 TcdB 亚型代表性序列的邻接系统发育树

Fig. 2 Neighbor-joining phylogenetic tree of representative sequences of each TcdB subtype

表 1 TcdB 各亚型间氨基酸组成同源度百分比

Tab. 1 Amino-acid identity percentage of representative strain between subtypes

Subtype	TcdB1	TcdB2	TcdB3	TcdB4	TcdB5	TcdB6	TcdB7	TcdB8	TcdB9	TcdB10	TcdB11
TcdB1											
TcdB2	93.71										
TcdB3	96.07	90.79									
TcdB4	98.06	92.82	96.32								
TcdB5	92.18	88.09	91.12	92.52							
TcdB6	98.14	92.90	96.41	99.83	92.60						
TcdB7	91.93	87.66	91.17	92.35	99.07	92.52					
TcdB8	99.66	93.54	96.32	98.23	92.27	98.31	92.01				
TcdB9	94.13	99.49	90.96	93.07	87.88	93.16	87.45	93.87			
TcdB10	97.70	85.88	94.24	96.22	93.83	96.22	94.00	97.62	86.45		
TcdB11	99.58	93.96	95.99	97.89	92.35	97.97	92.01	99.24	94.38	97.12	
TcdB12	99.62	93.66	96.03	97.93	92.05	98.01	91.80	99.54	93.92	97.53	99.28

(*clostridioides difficile* infection, CDI)在中国呈爆发式增长。约 83.90% 的菌株都是自 2015 年后分离得到,且多集中于中国发达地区。其中,在浙江省分离得到的菌株占总数的 40.20%;香港地区与北京市分别报道了 129 例(9.52%)与 105 例(7.74%)菌株信息,剩余 22.06% 为国内各地报道的零星病例(图 3)。

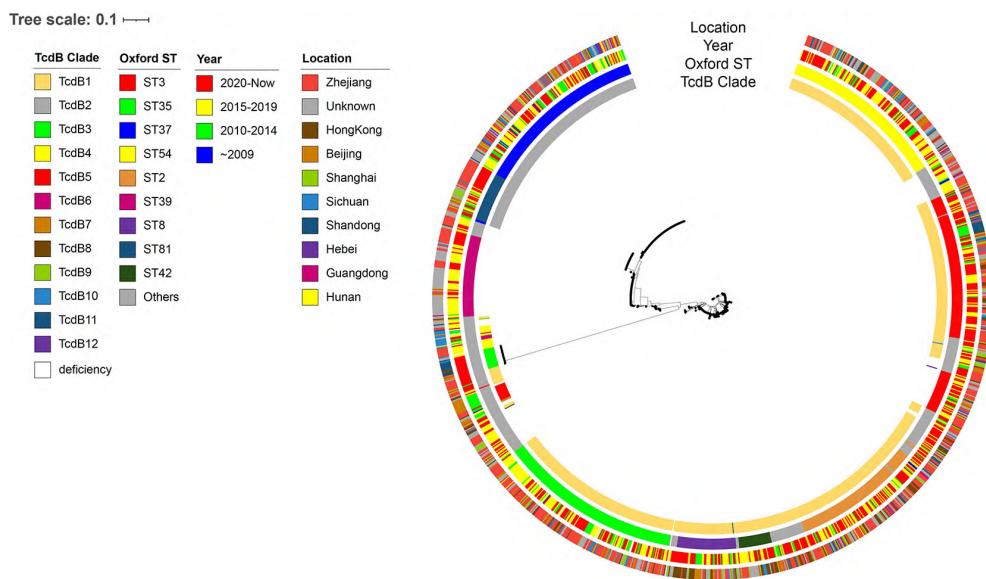


图 3 1 335 株艰难梭菌系统发育树
Fig.3 Phylogenetic tree of 1 335 *C. difficile* TcdB genome based on NCBI GenBank

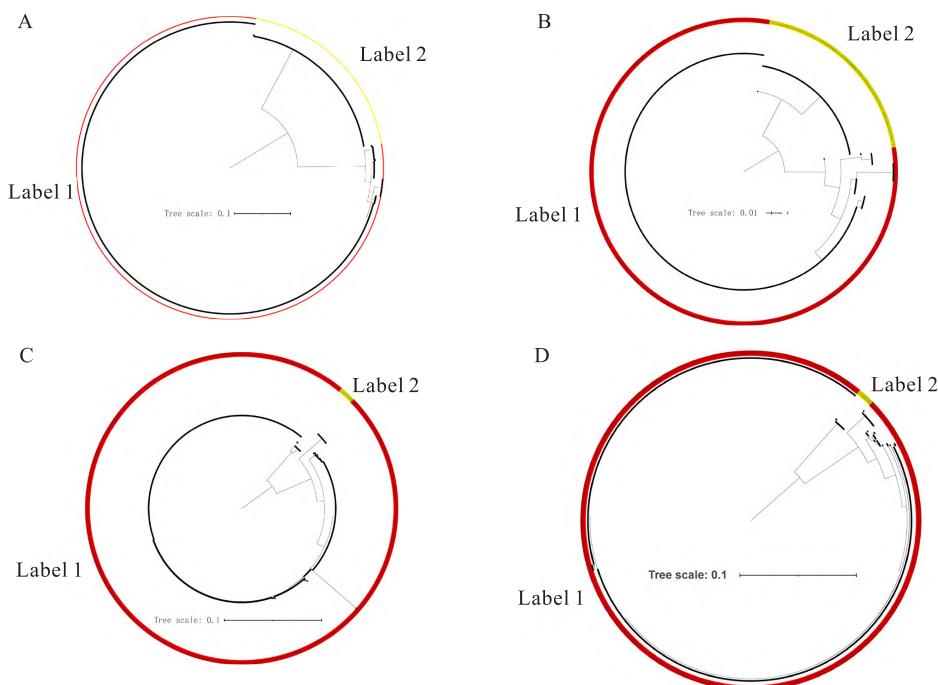


图 4 艰难梭菌 TcdB 各结构域系统发育树
Fig.4 Phylogenetic tree of individual domains in *C. difficile* TcdB

A: Glucosyltransferase domain; B: Cysteine protease domain; C: Delivery and receptor-binding domain; D: Combined repetitive oligopeptides domain.

进一步对每一株艰难梭菌 TcdB 的每一个结构域进行了单独的系统发育树分析,结果见图 4。每个结构域的系统发育树主要分为两个组(Label 1 与 Label 2),这与 TcdB 亚型分型也大致对应。其中, TcdB2 与其它亚型的氨基酸差异性主要位于葡萄糖基转移酶结构域与半胱氨酸蛋白酶结构域,提示来自 TcdB2 的 ST37/ST81 型菌株,与其它菌株,酶活性即

毒力高低存在较大差异。部分菌株 RBD 的氨基酸组成存在差异,这可能会影响到毒素分子与靶细胞的结合位点与亲和力。

2.3 艰难梭菌 TcdB 的生物信息学分析 为了比较 TcdB 各亚群间理化性质、空间结构与抗原表位有何异同,本研究选择了 TcdB1、TcdB2 两个主流亚群(总占比 93.94%),使用生物信息学在线软件软件对其代表性菌株的氨基酸序列进行分析并比较。其中,TcdB1 代表性菌株选择的是国际标准菌株 VPI10463;TcdB2 代表性菌株选择的是 2022 年上海交通大学医学院附属瑞金医院分离得到的菌株(GCA_024972115.1)。

2.3.1 理化性质 TcdB1 理化性质预测结果如下:该蛋白由 2 366 个氨基酸组成,分子式为 $C_{12135}H_{18574}N_{2984}O_{3857}S_{54}$,相对分子质量约为 269.71 ku;理论 PI 值为 4.41,组成该蛋白的负电荷残基(天冬氨酸与谷氨酸)与正电荷残基(精氨酸与赖氨酸)的数量分别为 392 与 205;半衰期具体信息如下:30 h(哺乳动物网织红细胞,体外);>20 h(酵母细胞,体内);>10 h(大肠埃希菌,体内)。经预测,该蛋白的稳定系数为 36.49,可归类为稳定蛋白;亲水指数为 -0.318,属于亲水性蛋白;脂肪族指数为 90,该值反映的是组成该蛋白的氨基酸中,非极性氨基酸的比例,与蛋白质的热稳定性相关。TcdB2 与 TcdB1 在理化性质上差异小:氨基酸组成相似,半衰期相近,同样属于稳定蛋白与亲水性蛋白。

2.3.2 信号肽及跨膜区 经预测,TcdB1 与 TcdB2 均无信号肽;这 2 种蛋白均为非跨膜蛋白,无跨膜区。

2.3.3 空间结构 TcdB1 二级结构预测结果如图 5A 所示: α -螺旋的数量为 393(16.10%); β -转角的数量为 272(11.50%);无规则卷曲的数量为 896(37.87%)。延伸链通常指的是在蛋白质的二级结构预测中完全扩展的链段,其数量为 805(34.02%),代表着组成蛋白质的氨基酸残基在空间中的排布方式。TcdB2 与之相比存在较大差异,结果如图 5B 所示: α -螺旋的数量为 798(33.71%);无 β -转角;无规则卷曲的数量为 1 119(47.28%),延伸链数量为 450(19.01%)。

2.3.4 B 细胞表位预测 登录在线网站 ABCpred,输入 TcdB1 与 TcdB2 氨基酸序列,默认参数,阈值设置为 0.51。经预测,TcdB1 评分 >0.90 的表位共有 9 条,其中 3 条具有抗原性(表 2)。TcdB2 评分 >0.90 的表位共有 8 条,其中 4 条具有抗原性(表

3)。TcdB1 与 TcdB2 的 B 细胞表位数量、序列起始位点与抗原性均相近,且表位序列全部位于半胱氨酸水解酶与递送与受体结合结构域。

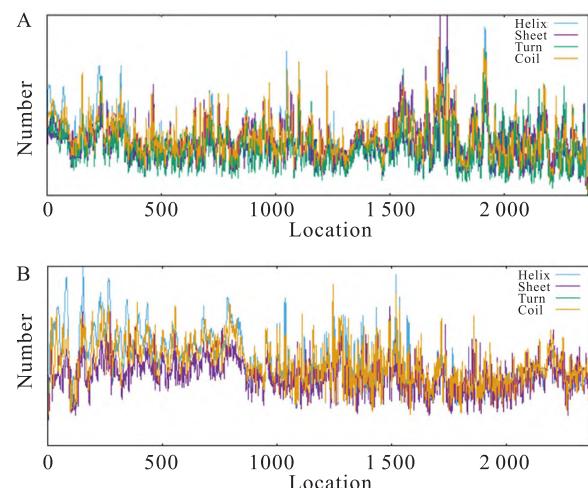


图 5 TcdB 二级结构预测

Fig. 5 Secondary structure prediction of TcdB

A: Secondary structure prediction of TcdB1; B: Secondary structure prediction of TcdB2.

表 2 TcdB1 的 B 细胞抗原表位预测

Tab. 2 Prediction of B cell epitopes of TcdB1

Method	Location	Sequence (5'-3')	Score
ABCpred	1 700 - 1 716	DEINITPVYETNNNTYP	0.92
	1 167 - 1 183	CEIWRMEGGSGHTVTD	0.93
	733 - 749	DSIIVSANQYEVRINS	0.95

表 3 TcdB2 的 B 细胞抗原表位预测

Tab. 3 Prediction of B cell epitopes of TcdB2

Method	Location	Sequence (5'-3')	Score
ABCpred	416 - 432	GPIISQGNDFNTTMNN	0.92
	1 701 - 1 717	DEINITPVYETNNNTYP	0.92
	1 168 - 1 184	CEIWRMEGGSGHTVTD	0.93
	734 - 750	DSIIVSANQYEVRINS	0.95

2.3.5 T 细胞表位预测

2.3.5.1 CD8⁺ T 细胞表位预测 登录在线网站 SYFPEITHI,输入 TcdB 氨基酸序列。调整参数:选择 MHC 类型为 HLA-A * 02:01;默认 mer 为 9 aa。经预测:TcdB1 评分 >26 的表位序列共有 14 条,其中 3 条有抗原性(表 4)。TcdB2 评分 >26 的表位序列共有 18 条,其中 5 条表位序列有抗原性(表 5)。与 TcdB1 比较,TcdB2 的 CD8⁺ T 细胞抗原表位较多,抗原性较强,且 TcdB2 位于 N 端的葡萄糖基转移酶结构域有一段可被识别的抗原表位,该结构域为毒素蛋白的酶活区。

表 4 TcdB1 的 CD8⁺ T 细胞抗原表位预测Tab. 4 Prediction of CD8⁺ T cell epitopes of TcdB1

Method	Location	Sequence (5'-3')	Score
SYFPEITHI	1 807 - 1 813	GLIGYDLGL	28
	1 098 - 1 107	GISAGIPSL	28
	1 091 - 1 100	ILLVPLAGI	30

表 5 TcdB2 的 CD8⁺ T 细胞抗原表位预测Tab. 5 Prediction of CD8⁺ T cell epitopes of TcdB2

Method	Location	Sequence (5'-3')	Score
SYFPEITHI	1 684 - 1 693	GIDSCVNKV	26
	1 808 - 1 817	GLIGYDLGL	28
	1 099 - 1 108	GISAGIPSL	28
	78 - 87	YLVIEIEL	29
	1 092 - 1 101	ILLVPLAGI	30

2.3.5.2 CD4⁺ T 细胞表位预测 调整参数:选择 MHC 类型为 HLA-DRB * 0701;mer 为 15 aa。经预测,TcdB1 评分 > 26 的表位序列共有 26 条,其中 13 条表位序列有抗原性(表 6)。TcdB2 评分 > 26 的表位序列共有 22 条,其中 9 条表位序列有抗原性(表 7)。与 TcdB1 比较,TcdB2 的 CD4⁺ T 细胞抗原表位较少,抗原性较差,这可能是由于 β-转角易形变,利于抗体结合。

2.4 TcdB 蛋白的原核表达及鉴定 根据 T、B 淋巴细胞表位预测结果发现,TcdB1 与 TcdB2 亚群代表序列间的抗原表位差异性较小。所以,课题组选择来自 TcdB1 的 VPI10463 菌株作为模式菌,对其合成分泌的 TcdB 进行基因工程制备。从 NCBI 数据中获取了艰难梭菌 VPI10463 TcdB 的基因序列后,构建重组质粒,质粒构建图谱如图 6A 所示。将质粒转化入工程菌中,并使用 PCR 对转化产物进行鉴定。琼脂糖凝胶电泳结果如图 6B 所示,重组质粒

已成功转化入感受态细胞中。扩增后,加入 IPTG 进行诱导表达并纯化。纯化后样品经 SDS-PAGE 与 Western blot 鉴定,实验结果如图 6C 所示:270 ku 处有清晰条带;TcdB 为上清液可溶性表达。配置缓冲液:每升缓冲液中含 0.05 mol HEPES、0.15 mol NaCl

表 6 TcdB1 的 CD4⁺ T 细胞抗原表位预测Tab. 6 Prediction of CD4⁺ T cell epitopes of TcdB1

Method	Location	Sequence (5'-3')	Score
SYFPEITHI	1 792 - 1 807	VSEHILSFTPSYYED	28
	1 749 - 1 764	FILMSTSEENKVSQV	28
	1 424 - 1 439	SYKLLISGELKILML	28
	1 333 - 1 348	GINIELSESVDVVIID	28
	505 - 520	KTNISQSTEQEMASL	28
	407 - 422	YKILNNNSLNPAISED	28
	339 - 354	DEEVQSSFESVLASK	28
	1 533 - 1 548	KDDIKISLSLTLQDE	30
	1 264 - 1 279	WRYFAFIADALITTL	30
	1 184 - 1 199	IDHFFSAPSITYREP	30
	933 - 948	KGTIFDTVNGKLVKK	30
	826 - 841	INVISNIIDTQIVEER	30
	421 - 436	DNDNFNTTTNTFIDSI	32

表 7 TcdB2 的 CD4⁺ T 细胞抗原表位预测Tab. 7 Prediction of CD4⁺ T cell epitopes of TcdB2

Method	Location	Sequence (5'-3')	Score
SYFPEITHI	1 750 - 1 765	FILMSTSEENKVSQV	28
	1 425 - 1 440	SYKLLISGELKILML	28
	1 334 - 1 349	GINIELSESVDVVIID	28
	340 - 355	DEEVQSSFESVLASK	28
	1 534 - 1 549	KDDIKISLSLTLQDE	30
	1 265 - 1 280	WRYFAFIADALITTL	30
	1 185 - 1 200	IDHFFSAPSITYREP	30
	934 - 949	KGTIFDTVNGKLVKK	30
	827 - 842	INVISNIETQVVEER	30

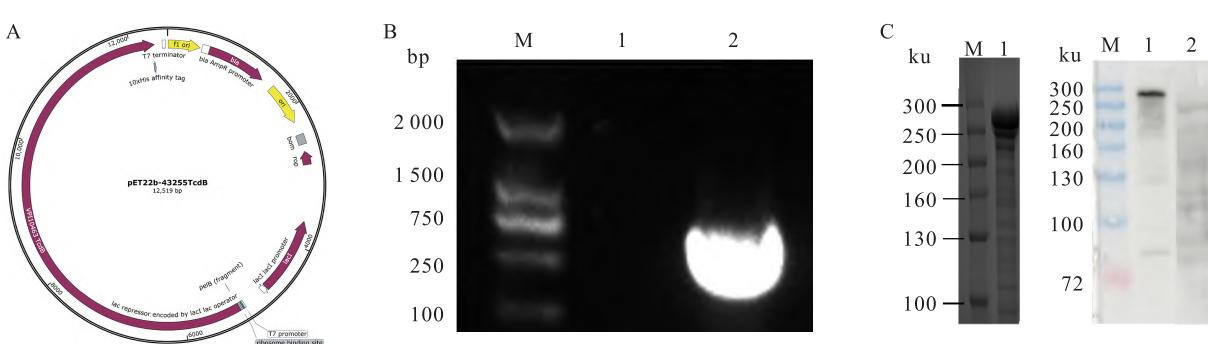


图 6 TcdB 蛋白的构建、表达、纯化与鉴定结果

Fig. 6 The Construction, expression, purification and identification of TcdB protein

A: Plasmid mapping of TcdB protein; B: Result of plasmid transformation identification, M: DNA marker, 1: Negative control, 2: TcdB plasmid PCR identification; C: Result of Western blot identification, M: protein marker, 1: TcdB antigen protein, 2: Negative control.

和 5% 蔗糖 ($C_{12}H_{22}O_{11}$)。将洗脱液透析至缓冲液中, 每次 4 h, 重复 4~6 次。使用 BCA 试剂盒, 检测本次纯化的蛋白质浓度为 0.7 mg/mL。

3 讨论

2003 年起, 随着高毒力、高耐药性菌株 NAP1/BL/027 型的爆发, CDI 已经成为医疗卫生相关性感染的主要威胁之一, 目前在全球范围内仍然是最常见的克隆株^[13]。与欧美国家比较, 艰难梭菌 027 型菌株在亚洲地区(尤其是中国)很少被提及。直到 2014 年, 中国内陆才首次分离发现艰难梭菌 027 型菌株; 2016 年, 第 1 例由艰难梭菌 NAP1/BL/027 型菌株爆发引起院内感染的案例首次被报道^[14]。在 1 355 株艰难梭菌中, ST1/027 型菌株仅仅只有 16 株(1.2%)。由于中国艰难梭菌的流行情况与其它地区不符, 因此, 如果沿用传统文献报道的分类方法对 TcdB 进行聚类分析, 不仅无法准确反映出 TcdB 亚型的具体信息, 在治疗及预防方面也存在难题。

目前, 国际上仍没有确定对 TcdB 的分群标准, 也没有所谓的标准菌株^[10]。因此, 在不同的研究中, 往往会出现同一种毒素变体有着不同命名的情况, 这在一定程度上阻碍了研究的深入^[11]。由于 TcdB 存在诸多亚型, 这些亚型在不同的氨基酸位点存在着单核苷酸多样性, 导致了各种 TcdB 突变株引起的艰难梭菌相关性疾病, 无论在致病性、致死率、GTD 活性、受体结合位点与结合亲和度方面均存在着差异。研究^[15]报道, 经 FDA 批准认证的针对 CDI 的单克隆药物 Bezlotoxumab, 在治疗由 TcdB1 与 TcdB2 引起的感染时, 会有完全不同的作用。在疫苗研究上, 世界上各大知名疫苗生产商多聚焦于毒素的受体结合区, 但无一家公司可以顺利通过Ⅲ期临床试验^[16]。因此, 对于该菌的基因流行病学研究及抗原表位的预测, 选择有效的疫苗抗原, 显得尤为重要。

本研究整理归纳了 1 355 株于中国分离得到的艰难梭菌基因组相关信息, 包括其毒素表达情况、ST 分型、分离年份及分离地区; 建立了一个专门针对中国艰难梭菌 TcdB 的分型系统, 并分别对其结构域也同时进行了建树分析; 预测分析了包含菌株数量最多的 2 个分群, TcdB1 与 TcdB2 亚型的代表性菌株的 B、T 细胞抗原表位, 旨在观察分群之间的异同。为未来针对 TcdB 某些与预防或治疗相关区域而制备的单克隆抗体及亚单位疫苗的研究提供新的思路。同时, 课题组尝试对 TcdB1 亚群中

VPI10463 合成分泌的 TcdB 进行重组制备。尽管毒素全长的原核表达存在诸多困难, 但仍然得到了纯度较高的抗原蛋白, 接下来的工作应聚焦于表达量的提高。

展望未来, 随着对临床分离株进行基因组测序能力的提高, 将越来越有可能根据其全基因组系统发育关系及其毒素亚型对菌株进行分类。TcdB 作为艰难梭菌最重要的毒力因子, 对其结构与功能进行更加深入的研究分析, 对未来预防与治疗 CDI 将会起到至关重要的作用。

参考文献

- [1] Lim S C, Knight D R, Riley T V. *Clostridium difficile* and one health[J]. Clin Microbiol Infect, 2020, 26(7): 857–63. doi: 10.1016/j.cmi.2019.10.023.
- [2] Stieglitz F, Gerhard R, Honig R, et al. TcdB of *Clostridioides difficile* mediates RAS-dependent necrosis in epithelial cells[J]. Int J Mol Sci, 2022, 23(8): 4258. doi: 10.3390/ijms23084258.
- [3] Luo J, Yang Q, Zhang X, et al. TFPI is a colonic crypt receptor for TcdB from hypervirulent clade 2 *C. difficile*[J]. Cell, 2022, 185(6): 980–94. doi: 10.1016/j.cell.2022.02.010.
- [4] Di Bella S, Sanson G, Monticelli J, et al. *Clostridioides difficile* infection: history, epidemiology, risk factors, prevention, clinical manifestations, treatment, and future options[J]. Clin Microbiol Rev, 2024, 37(2): e0013523. doi: 10.1128/cmr.00135–23.
- [5] Donskey C J, Dubberke E R, Klein N P, et al. Clover (CLOstridium difficile vaccine efficacy tRial) study: a phase 3, randomized trial investigating the efficacy and safety of a detoxified toxin a/b vaccine in adults 50 years and older at increased risk of *Clostridioides difficile* infection[J]. Clin Infect Dis, 2024, 79(6): 1503–11. doi: 10.1093/cid/ciae410.
- [6] Duvaud S, Gabella C, Lisacek F, et al. Expasy, the swiss bioinformatics resource portal, as designed by its users[J]. Nucleic Acids Res, 2021, 49(W1): W216–27. doi: 10.1093/nar/gka225.
- [7] Hashemzadeh P, Nezhad S A, Khoshkhabar H. Immunoinformatics analysis of *Brucella melitensis* to approach a suitable vaccine against brucellosis[J]. J Genet Eng Biotechnol, 2023, 21(1): 152. doi: 10.1186/s43141–023–00614–6.
- [8] Monterrubio-López G P, Llamas-Monroy J L, Martínez-Gómez A A, et al. Novel vaccine candidates of *Bordetella pertussis* identified by reverse vaccinology[J]. Biologicals, 2024, 85: 101740. doi: 10.1016/j.biologicals.2023.101740.
- [9] Liu J, Chen X, Wang J, et al. Prediction and identification of CD4⁺ T cell epitope for the protective antigens of *Mycobacterium tuberculosis*[J]. Medicine (Baltimore), 2021, 100(6): e24619. doi: 10.1097/MD.00000000000024619.
- [10] Mansfield M J, Tremblay B J, Zeng J, et al. Phylogenomics of 8 839 *Clostridioides difficile* genomes reveals recombination-driven evolution and diversification of toxin A and B[J]. PLoS Pathog,

- 2020, 16 (12) : e1009181. doi: 10.1371/journal.ppat.1009181.
- [11] Shen E, Zhu K, Li D, et al. Subtyping analysis reveals new variants and accelerated evolution of *Clostridioides difficile* toxin B [J]. *Commun Biol*, 2020, 3(1) : 347. doi: 10.1038/s42003-020-1078-y.
- [12] Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation [J]. *Nucleic Acids Res*, 2021, 49 (W1) : W293 – 6. doi: 10.1093/nar/gkab301.
- [13] Bednarska A, Sosińska-Bryla I, Hackiewicz M, et al. Risk factors for complicated course and recurrence of *Clostridioides difficile* infections during COVID-19 pandemic [J]. *Przegl Epidemiol*, 2022, 77(1) : 23 – 33. doi: 10.32394/pe.77.02.
- [14] Zhang R, Man Y, Bai Y, et al. Molecular characterization of *Clostridioides difficile* ribotype 027 in a major Chinese hospital [J]. *J Microbiol Immunol Infect*, 2021, 54 (6) : 1179 – 83. doi: https://doi.org/10.1016/j.jmii.2021.01.003
- [15] Thandavaram A, Channar A, Purohit A, et al. The efficacy of bezlotoxumab in the prevention of recurrent *Clostridium difficile*: a systematic review [J]. *Cureus*, 2022, 14(8) : e27979. doi: 10.7759/cureus.27979.
- [16] Kelly C R, Allegretti J R. Review article: Gastroenterology and *Clostridium difficile* infection: past, present, and future [J]. *Clin Infect Dis*, 2023, 77 (Supplement_6) : S463 – 70. doi: 10.1093/cid/ciad644.

Bioinformatics analysis and expression of the major genetic type of TcdB

Fang Yitai^{1,2}, Ning Nianzhi², Sun Yakun², Li Deyu², Wang Hui², Luo Deyan^{1,2}

(¹School of Basic Medical Sciences, Anhui Medical University, Hefei 230032;

²State Key Laboratory of Pathogen and Biosecurity, Academy of Military Medical Sciences, Beijing 100071)

Abstract Objective To analyze the bioinformatics of domestic *Clostridium difficile* toxin B (TcdB) and prepare it to provide data support for the development of effective vaccines. **Methods** Using bioinformatics software such as Snippy, Blast, Muscle, and the dist.alignment() and hclust() functions in R, 1 355 strains of *Clostridium difficile* from NCBI GenBank in China were compared and analyzed, and TcdB were grouped. The maximum likelihood tree and phylogenetic tree were beautified and displayed using iTOL. An online bioinformatics analysis website was used to predict and analyze the spatial structure and antigenic epitopes of the two largest subgroups, TcdB1 and TcdB2. The antigen protein TcdB was expressed and purified by prokaryotic system. **Results** According to the genotype of toxin B, the 1 355 prevalent strains of *Clostridium difficile* in China could be roughly divided into 12 subtypes, among which TcdB1 and TcdB2 were the main subtypes, accounting for more than 93.94% of all isolated strains, and about 17.20% of the strains were nontoxigenic or lack TcdB. The antigen epitope prediction of TcdB1 and TcdB2 showed that their antigen epitope distributions were basically the same, and many of them were distributed outside the C-terminal combined repetitive oligopeptides domains. **Conclusion** A specialized typing system for *C. difficile* TcdB in China has been established, and its main subtypes have been predicted for antigenic epitopes. The screened TcdB has been expressed for recombinant preparation.

Key words *Clostridioides difficile*; Toxin B; sequences analysis; bioinformatics analysis; epitopes prediction; recombinant expression

Fund program Independent Research Project of State Key Laboratory of Pathogen and Biosecurity (No. SKLPBS2228)

Corresponding author Luo Deyan, E-mail: ldy612@126.com