

斑点叉尾鲷源海豚链球菌 DGX07 株 *simA* 基因 克隆、鉴定及分子特性分析

汪开毓^{1,2*}, 王均^{1,2}, 肖丹³, 陈德芳^{1,2},
黄锦炉^{1,2}, 黄凌远^{1,2}, 付希^{1,2}, 王浩丞^{1,2}

(1. 四川农业大学鱼病研究中心, 四川 雅安 625014;

2. 四川农业大学动物疫病与人类健康四川省重点实验室, 四川 雅安 625014;

3. 通威股份有限公司, 四川 成都 610041)

摘要: 以 Primer p14(5'-GATCAAGTCC-3')为引物对分离自患病斑点叉尾鲷海豚链球菌强毒株 DGX07 进行 RAPD 分析。同时参照 GenBank 中海豚链球菌 *simA* 基因序列设计特异性引物,以 DGX07 基因组 DNA 为模板,扩增出约 1 500 bp 的 *simA* 基因,并将其克隆到 pMD19-T 载体上,之后对重组质粒进行 PCR 和双酶切(*Bam*H I + *Xho* I)鉴定,鉴定正确后送测序公司测序。RAPD 分析结果显示,DGX07 和标准菌株均能扩增出 750 bp 大小的条带。通过生物信息学软件对测序结果分析显示,*simA* 基因全长 1 566 bp,由 521 个氨基酸组成,与海豚链球菌 *simA* 和 *simB* 亲缘性达 100%,存在 1 个由 41 个氨基酸组成的信号肽,具有 Bap31 和 Gram_pos_anchor 两个超家族的保守结构域;具有与蛋白翻译后修饰功能相关的磷酸化位点 22 个和 N-糖基化位点 2 个,编码多肽链中亲水区大于疏水区,是一种膜外蛋白,并具有多个抗原优势位点区域。密码子偏爱性分析表明,斑点叉尾鲷源海豚链球菌 *simA* 基因密码子使用频率差异较大,密码子偏爱性与酵母较为接近。获得 GenBank 登录号为 JF330100。

关键词: 斑点叉尾鲷; 海豚链球菌; *simA* 基因; 克隆

中图分类号: S 917; Q 785

文献标志码: A

海豚链球菌(*Streptococcus iniae*)为 β -溶血类链球菌,不属于任何 Lancefield 分群类型,目前报道有两种血清型(I和II)^[1-2]。1976年,Agnew等^[3]在亚马逊淡水河豚上首次分离到海豚链球菌,随后在以色列、日本、美国和中国香港等多个国家出现关于该菌引起养殖鱼类发病的报道^[3]。据统计,该菌引起美国水产养殖的经济损失约每年 1 000 万美元,全球的经济损失约每年 1 亿美元^[4];被认为是危害多种冷暖水鱼类的重要致病菌^[5];同时,该菌也能感染其他动物和人类,特别是老年人和免疫力低下的人群^[6],成为人们广泛关注的动物源传染病病原。

海豚链球菌 M-like 蛋白为一种表面蛋白,能

够粘附鱼上皮细胞,具有绑定纤维蛋白原和抵抗巨噬细胞吞噬菌体的作用,为海豚链球菌的主要毒力因子之一^[7]。目前关于海豚链球菌 M-like 蛋白编码基因的研究较少,Justice等^[8]对 50 株来源于不同宿主和地区的海豚链球菌 *sim* 基因进行了分析,发现大多数菌株的 *simA* 基因长度都为 1 566 bp,编码蛋白具有 521 个氨基酸,但菌株 QMA0072 的 *sim* 基因编码蛋白具有 522 个氨基酸,菌株 QMO141 的 *sim* 基因(编码基因 *simB*)编码蛋白具有 579 个氨基酸,且菌株 QMO141 绑定纤维蛋白原能力也远小于 *simA* 基因类型的菌株。可见不同地域和宿主来源的海豚链球菌 *sim* 基因序列存在不同的变异体,对其蛋白的大小和功能

收稿日期:2011-09-25 修回日期:2011-12-04

资助项目:教育部创新团队项目(IRT0848);通威股份有限公司重点资助项目(2006-2009)

通讯作者:汪开毓,E-mail:kywang@sicau.edu.cn

有一定的影响;同时这种变异体的出现也说明 M-like 蛋白具有显著的抗原变异性,其组成的变化可能成为该抗原疫苗研究的重要阻碍。因此,对海豚链球菌 *sim* 编码基因进行系统的生物信息学分析,可为不同来源的海豚链球菌 *sim* 基因分子特性和相关功能的比较分析提供重要理论支撑。本研究中,作者通过从我国某水产养殖链球菌病暴发疫区的患病斑点叉尾鲷(*Ictalurus punctatus*)分离到的海豚链球菌强毒株进行基因组 DNA 提取,参考 GenBank 中海豚链球菌 *simA* 核酸序列设计的引物进行 PCR 扩增,并将扩增产物与 pMD19-T 载体连接,经 PCR 和双酶切鉴定以及序列的进化关系比对验证所扩增的基因即为 *simA* 基因,随后应用生物信息学分析工具对其进行分子特性分析,旨在为丰富海豚链球菌 *simA* 基因生物信息学理论提供资料,同时为 M-like 蛋白的功能和抗原性研究提供分子生物学依据。

1 材料与方法

1.1 菌株

海豚链球菌 DGX07 分离自某水产养殖链球菌病暴发疫区的患病斑点叉尾鲷,由本实验室分离、鉴定和保存;海豚链球菌标准菌株购自 ATCC 菌种库,编号:29178; *Escherichia coli* DH5 α 菌株购自天根生化科技(北京)有限公司。

1.2 主要试剂

pMD19-T (Simple) 克隆载体、高保真酶 PrimeSTAR HS (Premix)、T₄ DNA 连接酶、限制性内切酶 *Bam*H I 和 *Xho* I、DNA Marker、DNA 凝胶回收试剂盒和质粒提取试剂盒均购自 TaKaRa 公司。细菌基因组 DNA 提取试剂盒 (DP302-02) 购自天根公司。

1.3 引物设计与合成

根据 Bachrach 等^[9]报道的 Primer p14 (5'-GATCAAGTCC-3') 为引物对 DGX07 进行随机扩增多态性 DNA 标记 (rapid amplified polymorphic DNA, RAPD) 分析。根据 GenBank 中海豚链球菌 *simA* 基因序列 (GenBank 登录号为 EU287919), 运用生物软件 Oligo 6.0 和 Primer Premier 5.0 对其上下游引物进行设计,引物序列分别为上游 5'-CGGGATCCATGGCTAAACAAATCAAAGC-3' 划线部分为添加的 *Bam*H I 酶切位点;下游 5'-CCCTCGAGTTATTCTTCCTCTTTGCGTTTAC-

3' 划线部分为添加的 *Xho* I 酶切位点,划线前端部分为保护性碱基。引物由生工生物工程(上海)有限公司合成。

1.4 菌株培养及 DNA 制备

在无菌条件下,取低温保存的海豚链球菌 DGX07 和海豚链球菌标准菌株菌种接种于脑心浸液营养肉汤中,37 °C 恒温摇床 120 r/min 振荡培养 24 h,取 2 mL 菌液 5 000 \times g 离心 2 min 收集菌体,并按天根细菌基因组 DNA 提取试剂盒 (DP302-02) 的操作说明提取菌体基因组 DNA。

1.5 RAPD 分析 DGX07

以 DGX07 DNA 为模板进行 RAPD 分析,并设阴性对照(无菌超纯水)和阳性对照(标准菌株 DNA)。PCR 反应体系为 50 μ L: PrimeSTAR HS (Premix) 25 μ L, DNA 模板 1.0 μ L, Primer p14 (10 pmol/L) 1 μ L, 无菌超纯水 23 μ L。反应条件:94 °C 预变性 5 min,94 °C 30 s,37 °C 1 min,72 °C 1 min,共 25 个循环,最后 72 °C 延伸 10 min。取 5 μ L 扩增产物在 1% 琼脂糖凝胶中进行电泳,通过凝胶成像系统检测结果。

1.6 *simA* 基因的 PCR 扩增

以 DGX07 DNA 为模板对 *simA* 基因进行 PCR 扩增。PCR 反应体系为 25 μ L: PrimeSTAR HS (Premix) 12.5 μ L, DNA 模板 1.0 μ L, 上、下游引物 (10 pmol/L) 各 1 μ L, 无菌超纯水 9.5 μ L。反应条件:94 °C 预变性 5 min,94 °C 30 s,55 °C 35 s,72 °C 1 min,共 30 个循环,最后 72 °C 延伸 10 min。取 5 μ L 扩增产物在 1% 琼脂糖凝胶中进行电泳,通过凝胶成像系统检测结果。

1.7 *simA* 基因的 TA 克隆、测序及序列分析

PCR 扩增产物经 1% 琼脂糖凝胶电泳分离后,回收目的片段,再与 pMD19-T 载体进行连接反应,构建 pMD19-T-*simA* 重组质粒,并将重组质粒转化到大肠杆菌 DH5 α ,筛选阳性克隆并进行 PCR 和酶切鉴定。同时将 pMD19-T-*simA* 重组质粒送生工生物工程(上海)有限公司测序。

1.8 核酸序列特征分析

利用 NCBI 的 ORF Finder 工具进行开放性阅读框分析^[10],再利用 NCBI 的 BLASTN 工具进行序列的相似性搜索^[11],并通过 New BlankIt 向 NCBI 提交该序列。

1.9 蛋白质序列特征分析

应用瑞士蛋白质专家网在线翻译工具将

DGX07 *simA* 基因序列翻译成氨基酸序列^[12];通过 NCBI 的 BLASTP 程序进行氨基酸序列同源性检索^[13],利用 Clustal X 2.0 软件和 MEGA 4.1 软件构建进化树^[14-15];应用 DNASTAR Protean 软件进行二级结构预测,通过 Chou-Fasman 法从氨基酸残基的晶体结构预测蛋白质二级结构^[16],通过 Garnier-Robson 法计算特定氨基酸残基在特定结构内部的可能性^[17];应用 NCBI Conserved Domains 查找工具分析蛋白质结构域^[18];应用 Bioedit 7.0 软件进行疏水性分析^[19-20],并在 <http://www.biotech.ou.edu/> 进行 DGX07 M-like 蛋白溶解度预测;NetPhos 2.0 及 NetNGlyc 1.0 在线程序分析磷酸化位点和糖基化位点^[21-22],SignalP 3.0 Server 在线搜索查找信号肽序列^[23],同时 TMHMM 进行跨膜区预测^[24];根据 DNASTAR Jameson-Wolf 分析该蛋白的主要抗原决定簇位置;应用 EMBOSS 生物信息学在线分析工具的 chips 和 cusp 程序计算 DGX07 *simA* 基因中不同密码子在同一氨基酸编码中所占的比例,并通过外延法计算密码子在编码基因中出现的频率^[25],将结果与大肠杆菌、酵母和人的密码子偏爱性进行比较^[26]。

2 结果

2.1 RAPD 分析 DGX07

PCR 扩增结果显示,DGX07 和标准菌株均具有一条约 750 bp 大小的扩增产物,而阴性对照未扩增出条带(图 1)。

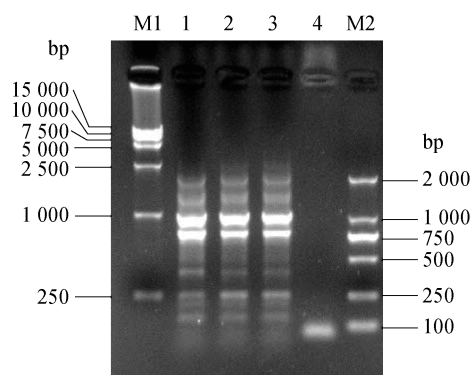


图 1 RAPD 分析电泳图

M1. DL15000 Marker; M2. DL2000 Marker; 1. 标准菌株; 2. DGX07-1; 3. DGX07-2; 4. 阴性对照。

Fig. 1 Electrophoretogram of RAPD

M1. DL15000 Marker; M2. DL2000 Marker; 1. type culture strain; 2. DGX07-1; 3. DGX07-2; 4. negative control.

2.2 *simA* 基因的扩增和 T 克隆鉴定

以 DGX07 基因组 DNA 为模板,根据所设计的 PCR 引物,经高保真 DNA 聚合酶特异性的扩增出一条位于 1 000 ~ 2 000 bp 目的片段(图 2)。PCR 产物经胶回收纯化后,与 pMD19-T 载体连接获得重组质粒 pMD19-T-*simA*,重组质粒经 *Bam*H I + *Xho* I 以双酶切和 PCR 鉴定正确(图 3)。

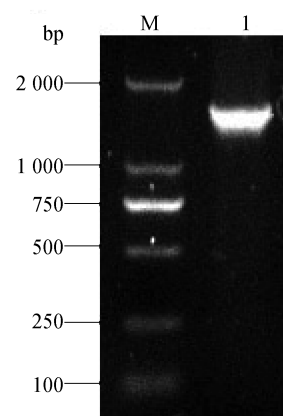


图 2 DGX07 *simA* 基因的 PCR 扩增结果

M. DL2000 Marker; 1. *simA* 基因扩增产物。

Fig. 2 PCR amplification result of DGX07 *simA* gene

M. DL2000 Marker; 1. PCR production of the *simA* gene.

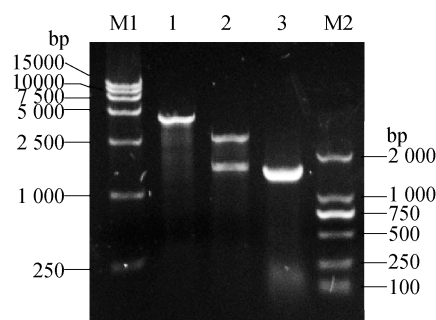


图 3 重组质粒 pMD19-T-*simA* 的酶切和 PCR 鉴定结果

M1. DL15000 Marker; M2. DL2000 Marker; 1. pMD19-T-*simA*/*Bam*H I; 2. pMD19-T-*simA*/*Bam*H I + *Xho* I; 3. pMD19-T-*simA* 重组质粒的 PCR 产物。

Fig. 3 Identification result of the recombinant plasmid pMD19-T-*simA* by restriction digestion and PCR amplification

M1. DL15000 Marker; M2. DL2000 Marker; 1. pMD19-T-*simA*/*Bam*H I; 2. pMD19-T-*simA*/*Bam*H I + *Xho* I; 3. PCR production of the recombinant plasmid pMD19-T-*simA*.

2.3 核苷酸序列特征分析

对重组质粒 pMD19-T-*simA* 测序结果表明,DGX07 *simA* 基因是一个由 1 566 个碱基组成的完整开放性阅读框(ORF),编码一个长度为 521

<http://www.sexuebao.cn>

个氨基酸的蛋白。全序列中 GC 含量为 36.53%。

2.4 蛋白质序列特征分析

氨基酸序列分析 瑞士蛋白质分析专家系统 protparam 工具预测表明,DGX07 *simA* ORF 编码一条由 521 个氨基酸残基组成的多肽,相对分子量为 57.48 ku,等电点理论值为 5.78,分子式为 $C_{2481}H_{4159}N_{705}O_{827}S_{12}$ 。*simA* 编码蛋白的氨基酸组成成分分析表明(表 1),蛋白质中含量较高的

氨基酸为 Ala(14.4%)、Glu(13.6%)和 Lys(13.6%),含量较少的氨基酸为 Trp(0.4%),Tyr(0.8%)和 His(0.8%),而 *simA* 氨基酸组成中则不含有 Cys、Pyl 和 Sec。其中,疏水氨基酸 203 个,占氨基酸总数的 39.0%,亲水性氨基酸有 134 个,占 25.7%,酸性氨基酸(Asp + Glu)93 个,占 17.9%,碱性氨基酸(Arg + Lys)87 个,占 16.7%。

表 1 海豚链球菌 *simA* 基因序列推导肽链的氨基酸组分分析
Tab.1 Amino acid composition of *S. iniae simA* gene deduced amino acid

氨基酸 amino acid	数量 amount	摩尔百分比/% mol percentage	氨基酸 amino acid	数量 amount	摩尔百分比/% mol percentage
Ala	75	14.4	Lys	71	13.6
Arg	16	3.1	Met	12	2.3
Asn	31	6.0	Phe	6	1.2
Asp	22	4.2	Pro	6	1.2
Cys	0	0	Ser	23	4.4
Gln	24	4.6	Thr	37	7.1
Glu	71	13.6	Trp	2	0.4
Gly	19	3.6	Tyr	4	0.8
His	4	0.8	Val	22	4.2
Ile	21	4.0	Pyl	0	0
Leu	55	10.6	Sec	0	0

进化树分析 利用 Clustal X 2.0 软件对 20 个相关菌株的氨基酸序列(表 2)运行 Multiple Sequence Alignment 程序和 MEGA 4.1 程序构建系统发育树系图(图 4)。可见,DGX07 *simA* 编码氨基酸序列与海豚链球菌的 *simA* 和 *simB* 编码氨基酸序列亲缘关系均达 100%;与链球菌属的乳房链球菌(*S. uberis*)、停乳链球菌(*S. dysgalactiae*)、酿脓链球菌(*S. pyogenes*)和马链球菌(*S. equi*)等的亲缘关系较远。

二级结构预测 将 *simA* 基因编码氨基酸序列,共计 521 个氨基酸运用 DNASTAR Protean 软件模块进行二级结构预测,使用不同的预测方法分析,预测 *simA* 基因编码的氨基酸序列的二级结构不尽相同,其所含有的 α 螺旋、 β 折叠和下转角区段的数目及存在的位置表现出很大的差异。Garnier-Robson 法预测结果显示,鱼源海豚链球菌分离株的 *simA* 基因编码氨基酸序列有 10 个 α 螺旋中心,6 个 β 折叠区段,19 个 T 转角区段;Chou-Fasman 法预测结果显示,该蛋白有 15 个 α 螺旋中心,8 个 β 折叠区段,10 个 T 转角区段。两种方法所测的共有 α 螺旋中心为 Met1-Ala22、Leu46-Ala54、Trp77-Ile87、Arg92-Leu118、

Lys123-Leu135、Asn144-Asp155、Ile170-Arg177、Glu182-Ser187、Thr201-Ile234、Thr243-Leu262、Val269-Glu291、Ala299-Glu337、Thr340-Lys426、Ala431-Pro455、Arg515-Glu521;共有的 β 折叠区段为 Tyr67-Tyr67、Arg73-Phe74、Lys138-Gln139、Asn474-Asn474、Ser487-Asp490;共有的 T 转角区段 Ile26-Ile29、Thr37-Thr37、Ser194-Val196。

超家族结构域分析 NCBI 的 CDD (conserved domain) 工具分析发现,DGX07 *simA* 编码氨基酸序列在 N-末端的第 221 个和第 424 个氨基酸残基之间含有 1 个 B 细胞受体关联蛋白 31(Bap31)超家族的保守结构域和 N-末端的第 477 个和第 516 个氨基酸残基之间含有 1 个 Gram_pos_anchor 超家族的保守结构域。经一次迭代分析,Bap31 超家族结构域与斑马鱼(*Danio rerio*)、凹鼻鲀(*Tetraodon nigroviridis*)、光滑念珠菌(*Candida glabrata*)和白色念珠菌(*Candida albicans*)等具有极高相似性;Gram_pos_anchor 结构域序列与链球菌属的其他成员菌,如乳酸乳球菌(*Lactococcus lactis*)、酿脓链球菌和马链球菌具有极高相似性。

表 2 *sim* 基因氨基酸序列相关的菌株
Tab. 2 The related strains of the *sim* gene amino acid sequence

基因 gene	蛋白产物 protein product	细菌名称 bacterial name	GenBank 登录号 accession no.
<i>simA</i>	SiM protein	<i>Streptococcus iniae</i>	ABZ81624
<i>simA</i>	SiM protein	<i>Streptococcus iniae</i>	ABZ81629
<i>simB</i>	SiM protein	<i>Streptococcus iniae</i>	ABZ81638
<i>lbp</i>	Lactoferrin binding protein	<i>Streptococcus uberis</i>	AAQ83577
<i>demA</i>	M-like protein	<i>Streptococcus dysgalactiae</i>	CAB65411
<i>emm</i>	M protein	<i>Streptococcus dysgalactiae</i>	ADD82984
<i>stg480.0</i>	Antiphagocytic M protein	<i>Streptococcus dysgalactiae</i>	YP_002995961
<i>fbp</i>	Antiphagocytic cell surface-anchored fibrinogen-and IgG Fc-binding protein SzM	<i>Streptococcus equi</i>	YP_002743751
<i>fbp</i>	Fibrinogen-binding protein	<i>Streptococcus equi</i>	AAC38445
<i>cspZ.2</i>	Emm-like cell surface protein CspZ.2	<i>Streptococcus equi</i>	YP_002124236
<i>spa18</i>	Protective antigen	<i>Streptococcus pyogenes</i>	AAD42939
<i>spa</i>	Streptococcal protective antigen	<i>Streptococcus pyogenes</i>	ACD81471
<i>fcrA8</i>	fcrA8	<i>Streptococcus pyogenes</i>	CAA55235
<i>emm12</i>	M protein	<i>Streptococcus pyogenes</i>	P19401
<i>emm18</i>	M18 protein precursor	<i>Streptococcus pyogenes</i>	NP_608047
<i>emm9</i>	M protein type 9	<i>Streptococcus pyogenes</i>	CAA51281
<i>emm1</i>	M protein type 1	<i>Streptococcus pyogenes</i>	NP_269973
<i>emm3.1</i>	M protein type 3	<i>Streptococcus pyogenes</i>	AAA96960
<i>emm</i>	M protein	<i>Streptococcus</i> sp. 'group G'	ACL51838

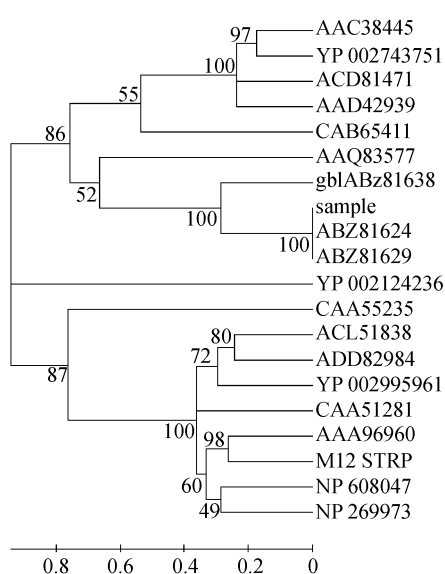


图 4 DGX07 *simA* 与参考菌株同源基因氨基酸序列进化树分析

“sample”代表 DGX07 *simA* 编码氨基酸序列。

Fig. 4 The phylogenetic tree analysis between the DGX07 *simA* and its homologous of the reference strains

“sample” represents the amino acid sequence of DGX07 *simA*.

疏水性分析 疏水性预测分析表明, DGX07 *simA* 多肽链有 7 个主要的疏水区, 分别位于氨基酸序列的 14 ~ 38 位、58 ~ 64 位、190 ~

201 位、296 ~ 306 位、325 ~ 339 位、402 ~ 407 位和 497 ~ 511 位。整个蛋白质疏水性最大值为 2.156, 最小值为 -3.222, 而且亲水区在蛋白质多肽链占据的区域大于疏水区域。重组蛋白可溶度预测结果显示, 当选择大肠杆菌为表达宿主时, DGX07 *simA* 重组蛋白的溶解度达 64.4%。

磷酸化位点和 N-糖基化位点分析

NetPhos2.0 程序分析表明, 当阈值为 0.5 时, DGX07 *simA* 氨基酸序列中共有 22 个潜在的磷酸化位点, 其中包含 10 个丝氨酸 (Ser) 磷酸化位点、11 个苏氨酸 (Thr) 磷酸化位点和 1 个酪氨酸 (Tyr) 磷酸化位点, 而且 N-末端氨基酸磷酸化的数量明显多于 C-末端。NetNGlyc1.0 程序预测表明, DGX07 *simA* 氨基酸序列中共有 2 个潜在的 N-糖基化位点, 当取阈值为 0.5 时, 第 166 位和第 293 位的天冬酰胺 (Asn) 可能被糖基化。

信号肽和跨膜区分析 SignalP V 3.0 程序的 SignalP-NN 和 SignalP-NNSIMA-HMM 两个模型预测发现, DGX07 *simA* 氨基酸序列 N 末端有一个由 41 个氨基酸组成的信号肽, 信号肽切割位点位于序列的第 41 和 42 个氨基酸残基之间。通过在线预测工具 TMHMM Server v. 2.0 对 DGX07 *simA* 基因编码氨基酸序列进行跨膜区预

测,预测结果表明第 1~16 个氨基酸位于膜内、第 17~39 个氨基酸为跨膜区、第 40~521 个氨基酸位于膜外。

抗原位点预测 Jameson-Wolf 分析显示, DGX07 *simA* 编码氨基酸序列具有多个抗原优势位点区域,分别位于 4~16 位、39~61 位、70~73 位、82~113 位、118~181 位、183~193 位、202~225 位、229~267 位、270~296 位、303~338 位、341~398 位、402~450 位、457~463 位、465~495 位、513~521 位。

密码子偏爱性分析 用 EMBOSS 的 chips 程序计算 DGX07 *simA* 基因的 $Nc = 41.924$, 该值是一个基因的密码子使用频率与同义密码子平均使用频率偏差的量化值。 Nc 值的范围为 20 (每个氨基酸只使用一个密码子的极端情况) 到 61 (各个密码子均被平均使用)。然后运行 EMBOSS 的 cusp 程序,得到密码子使用频率。其中, DGX07 *simA* 基因有 10 个密码子出现频率为 0, 终止密码子仅出现 TAA。分析表明, DGX07 *simA* 基因在同一氨基酸的不同密码子的选择上存在一定的偏向性,特别是对于 A、D、E、G、H、I、K、P 和 Q 氨基酸的密码子使用的选择上,存在较大的偏向性,分别偏爱使用 GCT (Fract 值为 0.467)、GAT (Fract 值为 0.727)、GAA (Fract 值为 0.873)、GGT (Fract 值为 0.526)、CAT (Fract 值为 1.000)、ATT (Fract 值为 0.571)、AAA (Fract 值为 0.845)、CCA (Fract 值为 0.833)、CAA (Fract 值为 1.000)。1/1000 代表该密码子在编码基因中出现的频率。*simA/E. coli*、*simA/Yeast*、*simA/Human* 分别表示 DGX07 *simA* 与大肠杆菌、酵母及人的密码子出现频率的比值。DGX07 *simA* 与大肠杆菌、酵母及人的密码子使用频率比值 (*simA/E. coli*、*simA/Yeast*、*simA/Human*) 在 0.5~2.0 之间的表示对该密码子的偏爱性较为接近,反之则差别较大。结果表明, DGX07 *simA* 与大肠杆菌等 3 种密码子使用频率差值较大的在大肠杆菌有 44 个,酵母有 33 个,人有 40 个。由此可见, DGX07 *simA* 的密码子使用偏爱性模式与酵母的基因最相似,真核表达系统更适合于 DGX07 *simA* 基因编码蛋白的外源表达。

3 讨论

M-like 蛋白为海豚链球菌的主要毒力因子,

与酿脓链球菌、马链球菌和无乳链球菌的 M-like 蛋白一样具有绑定人纤维蛋白原和保护菌体不被吞噬的作用,且在酿脓链球菌中与其血清型多样性分析有关^[7]。新型血清型的出现主要由细菌表面抗原发生变异引起,海豚链球菌的 M-like 蛋白有多种变异体。Justice 等^[8] 研究显示,该蛋白为螺旋盘绕蛋白,分子量约 53 ku,主要具有两个变异体,其中一个是在原氨基酸序列中插入了一个氨基酸,另一个分子量约 59 ku,而 Locke 等^[7] 发现罗非鱼源的海豚链球菌分离株具有另一变异体 SiMA1,其氨基酸序列出现移码或提前终止。因此对海豚链球菌的 M-like 蛋白的核酸序列和氨基酸序列进行分析,能够为该蛋白抗原变异性研究提供分子生物学理论依据。

病原菌在不同遗传背景下,其特异性毒力基因可能发生变异,而这些毒力基因可能与疾病的发病机制之间存在密切的联系^[28]。目前研究认为,海豚链球菌在感染宿主的过程中,*sim* 基因出现两种突变体,共有 3 种类型^[7-8]。本试验中,测序结果显示, DGX07 *sim* 基因序列全长 1 566 bp,通过比对为 *simA*,推导的氨基酸序列含有 521 个氨基酸,其中不同氨基酸的含量差异较大。与已报道的海豚链球菌 *simA* 基因同一性为 100%,与 *simB* 差异较大。进化关系分析表明, DGX07 *simA* 氨基酸序列与海豚链球菌 *sim* 基因的亲缘关系达到 100%,提示 DGX07 *simA* 氨基酸序列仍具有极高的保守性。通过 New BlankIt 将序列提交到 NCBI,获得基因登录号为 JF330100。

氨基酸的疏水性是每种氨基酸所固有的特性,它是决定一个蛋白质最终折叠成的三维构象的一个重要因素^[29],可以作为判定该蛋白是否为跨膜蛋白的指标,特别是在一段多肽中如果能够发现几个亲、疏水区相间隔存在的情况,一般可作为跨膜蛋白的标识,而跨膜蛋白往往是受体等的作用位点。因此,分析疏水区和跨膜区对于分析一个蛋白质的性质功能具有一定的意义。分析 DGX07 *simA* 肽链的疏水区分布情况,亲水区在蛋白质多肽链占据的区域大于疏水区域,表现出一定的亲水性。当选择大肠杆菌为表达宿主时, *simA* 重组蛋白的溶解度达 64.4%,这提示建构表达质粒时,在 *simA* 基因两端选择加入合适的融合蛋白标签显得尤为重要,而且也有利于提高重组蛋白的可溶解度和重组蛋白的纯化水平^[28]。跨

膜区预测结果显示,该序列具有一个跨膜区,提示它可能作为膜受体起作用。信号肽是引导翻译出的前体蛋白通过细胞膜分泌到细胞外的一段序列,经信号肽酶切除信号肽序列后的前体蛋白才可能成为具有正常功能的成熟蛋白^[31]。信号肽预测结果表明,DGX07 *simA* 信号肽由 41 个氨基酸组成,信号肽切割位点位于序列的第 41 和 42 个氨基酸残基之间。而 Justice 等^[8]研究发现,不同类型的 *sim* 基因的前 41 个氨基酸残基具有 100% 的相似性,与本实验预测结果一致,表明该信号肽具有很高的保守性。

编码天然蛋白质 20 种氨基酸的密码子共 61 种,每种氨基酸可由 1 个至多个密码子(最多的有 6 个)编码。密码子的这种兼并性特征造成了不同物种密码子使用偏爱性不同,即编码同一个氨基酸的密码子在不同物种中具有不同的使用频率。分析密码子的偏爱性对于外源基因选择合适的宿主表达系统,进行基因的体外表达具有重要意义。本试验使用 EMBOSS 直接通过网络递交序列进行分析。从 CHIPS 程序返回的结果可以看出,DGX07 *simA* 在密码子的选择可能存在一定的偏向性,因此决定运行 CUSP 程序,检查该蛋白的密码子使用的详细情况,从返回的结果可以看出,该蛋白在 A、D、E、G、H、I、K、P 和 Q 等氨基酸在密码子的选择上存在一定的偏向性,而关于 *simA* 在密码子的使用上的偏爱性问题,目前还未见有报道。通过与大肠杆菌、酵母和人的密码子偏爱性进一步比较可以看出,*simA* 的密码子与大肠杆菌和人的相差较大,其表达系统选择更适合在真核系统(如酵母)中高效作用。当然,外源基因的表达受到多种因素的共同作用,如要表达载体的选择、表达蛋白的种类、是否会对表达宿主菌产生毒性、是否需大量表达、如何纯化等,而密码子偏爱性的影响只是其中一种因素。通过密码子偏爱性分析不仅可为目的基因表达选择合适的表达系统,同时也可改造密码子从而提高基因的表达量提供依据。

参考文献:

- [1] Weinstein M, Litt M, Kertesz D, et al. Invasive infections due to a fish pathogen, *Streptococcus iniae* [J]. The New England Journal of Medicine, 1997, 337(9): 589-594.
- [2] Facklam R, Elliott J, Shewmaker L, et al. Identification and characterization of sporadic isolates of *Streptococcus iniae* isolated from humans [J]. Journal of Clinical Microbiology, 2005, 43(2): 933-937.
- [3] Agnew W, Barnes A C. *Streptococcus iniae*: An aquatic pathogen of global veterinary significance and a challenging candidate for reliable vaccination [J]. Veterinary Microbiology, 2007, 122(1-2): 1-15.
- [4] Shoemaker C, Klesius P, Evans J. Prevalence of *Streptococcus iniae* in tilapia, hybrid striped bass, and channel catfish on commercial fish farms in the United States [J]. American Journal of Veterinary Research, 2001, 62(2): 174-177.
- [5] Bromage E S, Owens L. Infection of barramundi *Lates calcarifer* with *Streptococcus iniae*: Effects of different routes of exposure [J]. Diseases of Aquatic Organisms, 2002, 52(3): 199-205.
- [6] Weinstein M R, Litt M, Kertesz D A, et al. Invasive infections due to a fish pathogen, *Streptococcus iniae* [J]. The New England Journal of Medicine, 1997, 337: 589-594.
- [7] Locke J B, Aziz R K, Vicknair M R, et al. *Streptococcus iniae* M-Like protein contributes to virulence in fish and is a target for live attenuated vaccine development [J]. PLoS ONE, 2008, 3(7): 1-13.
- [8] Justice B, Reiny T, Aarti U, et al. Identification and molecular characterisation of a fibrinogen binding protein from *Streptococcus iniae* [J]. BMC Microbiology, 2008, 8(67): 7-16.
- [9] Bachrach G, Zlotkin A, Hurvitz A, et al. Recovery of *Streptococcus iniae* from diseased fish previously vaccinated with a streptococcus vaccine [J]. Applied and Environmental Microbiology, 2001, 67(8): 3756-3758.
- [10] Rombel I T, Sykes K F, Rayner S, et al. ORF-FINDER: a vector for high-throughput gene identification [J]. Gene, 2002, 282(1-2): 33-41.
- [11] Eklund A C, Friis P, Wernersson R, et al. Optimization of the BLASTN substitution matrix for prediction of non-specific DNA microarray hybridization. [J]. Nucleic Acids Research, 2010, 38(4): 27.
- [12] Wilkins M R, Gasteiger E, Bairoch A, et al. Protein identification and analysis tools in the ExPASy server [J]. Methods in Molecular Biology, 1999, 112: 531-552.
- [13] Jacob A, Lancaster J, Buhler J, et al. Mercury

- BLASTP: accelerating protein sequence alignment [J]. ACM Transactions on Reconfigurable Technology and Systems,2008,1(2).
- [14] Tamura K, Dudley J, Nei M, *et al.* MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0 [J]. Molecular Biology and Evolution,2007,24(8):1596-1599.
- [15] Thompson J D, Gibson T J, Plewniak F, *et al.* The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. Nucleic Acids Research,1997,25(24):4876-4882.
- [16] 徐超,李欣然,信洪一,等. 鸭瘟病毒 gC 基因的克隆及其分子特性分析 [J]. 中国兽医科学,2008,38(12):1038-1044.
- [17] Garnier J, Osguthorpe D J, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins [J]. Journal of Molecular Biology,1978,20(1):97-120.
- [18] Marchler-Bauer A, Anderson J B, Cherukuri P F, *et al.* CDD: a Conserved domain database for protein classification [J]. Nucleic Acids Research,2005,33(Database issue):D192-D196.
- [19] Hopp T P, Woods K R. Prediction of protein antigenic determinants from amino acid sequences [J]. Proceedings of the National Academy of Sciences,1981,78(6):3824-3828.
- [20] Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein [J]. Journal of Biology,1982,157(1):105-132.
- [21] Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites [J]. Journal of Molecular Biology,1999,294(5):1351-1362.
- [22] Gupta R, Jung E, Gooley A A, *et al.* Scanning the available *Dictyostelium discoideum* proteome for O-linked GlcNAc glycosylation sites using neural networks [J]. Glycobiology,1999,9(10):1009-1022.
- [23] Bendtsen J D, Nielsen H, von Heijne G, *et al.* Improved prediction of signal peptides: SignalP 3.0 [J]. Journal of Molecular Biology,2004,340(4):783-795.
- [24] Moller S, Croning M D, Apweiler R. Evaluation of methods for the prediction of membrane spanning region [J]. Bioinformatics,2001,17(7):646-653.
- [25] 王艳,马文丽,郑文岭. SARS 冠状病毒的密码子偏爱性分析 [J]. 生命科学研究,2003,7(3):219-223.
- [26] Nakamura Y, Gojobori T, Ikenura T. Condon usage tabulate form the international DNA sequence databases: status for the year 2000 [J]. Nucleic Acids Research,2000,28(1):292.
- [27] Barnes A C, Young F M, Horne M T, *et al.* *Streptococcus iniae*: serological differences, presence of capsule and resistance to immune serum killing [J]. Diseases of Aquatic Organisms,2003,53:241-247.
- [28] Brochet M, Couve E, Zouine M, *et al.* Genomic diversity and evolution within the species *Streptococcus agalactiae* [J]. Microbes and Infection,2006,8(5):1227-1243.
- [29] 曹海燕,王丹,张耀洲. 家蚕 BmLITAF 基因的克隆表达及生物信息学分析 [J]. 浙江理工大学学报,2010,27(1):149-154.
- [30] Wilkinson D L, Harrison R G. Predicting the solubility of recombinant proteins in *Escherichia coli* [J]. Biotechnology(N Y),1991,9(5):443-448.
- [31] Nielsen H, Krogh A. Prediction of signal peptides and signal anchors by a hidden Markov model. [J]. Proceedings of International Conference on Intelligent Systems for Molecular Biology,1998,6:122-130.

Cloning, identification and molecular characteristics analysis of *simA* gene of a *Streptococcus iniae* strain DGX07 from channel catfish

WANG Kai-yu^{1,2*}, WANG Jun^{1,2}, XIAO Dan³, CHENG De-fang^{1,2},
HUANG Jin-lu^{1,2}, HUANG Ling-yuan^{1,2}, FU Xi^{1,2}, WANG Hao-cheng^{1,2}

(1. Fisheries Department of Sichuan Agricultural University, Ya'an 625014, China;

2. Key Laboratory of Animal Disease and Human Health of Sichuan Province,

Sichuan Agricultural University, Ya'an 625014, China;

3. Tongwei Co. Ltd., Chengdu 610041, China)

Abstract: The RAPD analysis of *Streptococcus iniae* virulent strain DGX07 isolated from channel catfish (*Ictalurus punctatus*) was conducted by PCR with Primer p14 (5'-GATCAAGTCC-3'). The *simA* gene of DGX07 was amplified by PCR with specific primers and cloned into pMD19-T vector. The positive plasmid was selected and identified by PCR and digestion of double restriction enzyme (*Bam*H I + *Xho* I). Molecular characterization analysis of *simA* gene was performed by bioinformatics tools. The results showed that *S. iniae simA* amino acid sequence was highly conservative and with 100% homology to the *simA* and *simB* of *S. iniae* isolated. The polypeptide contained a signal peptide comprising 41 amino acids and Bap31 and Gram_pos_anchor functional domains. The polypeptide had some important sites related to post-translational modification, including 22 phosphorylation sites and 2 N-glycosylation sites. The hydrophilic regions were larger than hydrophobic regions in polypeptide chain. The *S. iniae simA* was extramembranous protein with multiple antigen advantaged regions. The analysis of codon bias indicated the codon usage frequency of *S. iniae simA* different, and the codon bias of *S. iniae simA* preferred the yeasts. Gene accession number: JF330100.

Key words: *Ictalurus punctatus*; *Streptococcus iniae*; *simA* gene; clone

Corresponding author: WANG Kai-yu. E-mail: kywang@sicau.edu.cn