



# 生物信息学法筛选志贺氏菌疫苗候选抗原

李钧<sup>1</sup> 邵铁娟<sup>2\*</sup> 濮小英<sup>1</sup>

(1. 杭州市疾病预防控制中心微生物检验科 浙江 杭州 310021)

(2. 浙江中医药大学基础医学院 浙江 杭州 310053)

**摘要:** 从 GOLD 基因组在线数据库中下载志贺氏菌 *S. flexneri* 2a301、*S. flexneri* 5b8401 和 *S. sonnei* Ss046 株的全基因组序列, 通过 SignalP、PSORT-B、Cell-Ploc、TMHMM、Phobius、LipoP 和 PRED-TMBB 等生物信息学工具, 筛选可能成为疫苗组份的表面抗原。结果分别从 3 株志贺氏菌的全基因组中得到了 96、158 和 107 个 ORF, 它们分别编码一系列的脂蛋白、 $\beta$ -桶型跨膜蛋白或分泌性蛋白, 组成了一个疫苗候选抗原的信息库。上述 ORF 编码的蛋白在 3 株志贺氏菌中呈现良好的同源性, 有 63 种候选抗原在同源性分析中被发现同时存在于 3 株志贺氏菌中。说明生物信息学方法可以作为高通量筛选志贺氏菌疫苗候选抗原的辅助工具, 且为志贺氏菌新疫苗研制提供了大量有用的数据信息。

**关键词:** 志贺氏菌, 生物信息, 疫苗, 候选抗原

## Screening Candidate Antigens for *Shigella* Vaccine by Bioinformatics Methods

LI Jun<sup>1</sup> SHAO Tie-Juan<sup>2\*</sup> PU Xiao-Ying<sup>1</sup>

(1. Department Microbiology, Hangzhou Center for Disease Control and Prevention, Hangzhou, Zhejiang 310021, China)

(2. School of Basic Medical Sciences, Zhejiang Chinese Medical university, Hangzhou, Zhejiang 310053, China)

**Abstract:** The bioinformatics analysis tools of SignalP, PSORT-B, Cell-Ploc, TMHMM, Phobius, LipoP and PRED-TMBB were used to screen putative surface-exposed or exported proteins for vaccine candidate antigens among three *Shigella* genome sequences of *Shigella flexneri* 2a strain 301, *S. flexneri* 2a strain 5b8401 and *Shigella sonnei* strain Ss046. As results, 96, 158 and 107 ORFs, putatively coding a series of lipoproteins,  $\beta$ -barrel outer membrane proteins and secreted proteins, respectively, were identified from these genome sequences. Among these ORFs, a total of 63 ORFs were conserved in all of these three genomes. In conclusion, bioinformatics methods could be used as an assistant tool in screening candidate antigens for the development of *Shigella* vaccines. The results might provide an useful basis for further studies of novel *Shigella* vaccines.

**Keywords:** *Shigella*, Bioinformatics, Vaccine, Candidate antigen

志贺氏菌属(*Shigella*)是一类大小约为(0.5–0.7)  $\mu\text{m} \times (2\text{--}3) \mu\text{m}$ 、无芽胞、无荚膜、有菌毛、无鞭毛的革兰氏阴性兼性厌氧杆菌,是人类细菌性痢疾最为常见的病原菌。每年大约有 1.69 亿人感染志贺氏菌,约 50 万–110 万人死亡,绝大部分病例来自于发展中国家,其中 69% 为 5 岁以下儿童<sup>[1]</sup>。在我国平均每年 1000 万的感染病例中,菌型分布以福氏志贺菌为主<sup>[2]</sup>,其中又以 2a 亚型(占 50%–70%)、3 型多见,其次为宋内氏志贺菌。近来抗生素的无节制滥用,使志贺氏菌的多重耐药性问题日益突现。一些地区的流行株对一线抗生素均产生不同程度的抗性,其中包括氟喹诺酮类(Fluoroquinolone),抗药率已经达到或超过 5%,并且具有明显的上升趋势<sup>[3–4]</sup>。在当前防治策略无法足够应对志贺氏菌大规模流行威胁的形势下,人类更亟待安全而高效的志贺氏菌疫苗的研制成功。

目前处于开发阶段的各类志贺氏菌减毒活疫苗以及化学组份疫苗尚不能达到大范围推广的要求。为了寻求突破,新表面抗原的发掘已经成为疫苗研制的首要环节和最亟待解决的重大科学问题。近来,快速发展的生物信息学促成了反向疫苗学的诞生,使得人们可以在计算机上完成候选抗原的高通量筛选工作,大大加快了寻找新抗原的速度,并且极大地开阔了抗原的考察范围,涵盖了那些仅在病原体感染宿主时期才表达的蛋白<sup>[5]</sup>。此方法已成功地应用于脑膜炎奈瑟球菌 B 血清型<sup>[6]</sup>、牙龈卟啉菌<sup>[7]</sup>、肺炎衣原体<sup>[8]</sup>、肺炎链球菌<sup>[9]</sup>等疫苗的研制,并且完全打破了传统疫苗学概念的束缚;另一方面,至今已有多达 7 种、涵盖 4 个菌群的志贺氏菌基因组及其大质粒被完全测序并公布,为反向疫苗学在志贺氏菌新疫苗开发上的应用创造了条件。本项目拟在丰富的志贺氏菌基因组信息基础上,结合蛋白质组学、结构生物学和分子免疫学等理论依据为候选抗原筛选的主要指标,运用反向疫苗学思路,建立相应的免疫性保护抗原生物信息学法筛选体系,以期获得若干具有实验验证价值的特异性免疫抗原,为志贺氏菌新疫苗的开发设计奠定基础。

# 1 材料与方法

## 1.1 材料

本课题根据志贺氏菌血清型流行特点,选取了

福氏志贺菌 2a 亚型 301 株、5b 亚型 8401 株以及宋内氏志贺菌 046 株作为研究材料。三株菌的全基因组序列和所有开放阅读框(ORF)从 GOLD 基因组在线数据库 v3.0 (<http://www.genomesonline.org>)中下载。菌株信息如表 1 所示。

表 1 三株志贺氏菌的基因组基本信息 Table 1 General features of the <i>Shigella</i> genomes			
菌株及其相关质粒 Strains and related plasmids	来源 Host origin	基因组大小 Length of the genomes (Mb)	编码蛋白数量 Proteins
<i>Shigella flexneri</i> 2a str. 301	北京一病例 (1984)	4.6000	4176
Plasmid pCP301		0.2200	263
<i>Shigella flexneri</i> 5b str. 8401	中国流行分离株	4.6000	4114
<i>Shigella sonnei</i> Ss046	中国流行分离株 (50 年代)	4.8000	4218
Plasmid pSS_046		0.2100	238
pSS046_spA		0.0084	5
pSS046_spB		0.0052	8
pSS046_spC		0.0021	1

## 1.2 方法

1.2.1 疫苗候选抗原的筛选: 根据表面抗原的一般分子特点,对所有编码蛋白序列进行如下筛选:

(1) 应用 SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP>)对所有 ORF 进行信号肽预测<sup>[10]</sup>,选择含有信号肽的序列;

(2) 结合 PSORT-B<sup>[11]</sup>(<http://www.psort.org/psortb>)和 Cell-Ploc<sup>[12]</sup>(<http://www.csbio.sjtu.edu.cn/bioinf/Gneg>)对含有信号肽序列的 ORF 进行蛋白表达亚细胞定位分析,选择定位于外膜或菌毛上的以及分泌到胞外的蛋白;

(3) 通过 TMHMM<sup>[13]</sup>(<http://www.cbs.dtu.dk/services/TMHMM>)和 Phobius<sup>[14]</sup>(<http://phobius.sbc.su.se>)筛选多次跨膜的蛋白,以排除  $\alpha$ -螺旋形式的跨膜蛋白;

(4) 对上述筛选到的 ORF 分别用 LipoP<sup>[15]</sup>(<http://www.cbs.dtu.dk/services/LipoP>)和 PRED-TMBB<sup>[16]</sup>(<http://biophysics.biol.uoa.gr/PRED-TMBB>)鉴别出脂蛋白和  $\beta$ -桶型跨膜蛋白;

(5) 对 A、B、C 步骤后筛选到的 ORFs 进行汇集整理,排除编码脂蛋白和  $\beta$ -桶型跨膜蛋白的两类 ORF,并从余下的 ORF 中挑选出定位于胞外的分泌

性蛋白。

**1.2.2 同源性分析:** 利用 ClustalX v1.81 对获得的脂蛋白、 $\beta$ -桶型跨膜蛋白和分泌性蛋白分别进行序列比对, 分析 3 株志贺氏菌菌株中疫苗候选抗原的同源性。

筛选流程如图 1 所示。

2 结果与分析

2.1 信号肽分析

志贺氏菌 *S. flexneri* 2a301 株、*S. flexneri* 5b8401 株和 *S. sonnei* Ss046 株全基因组分别编码 4439、4114 和 4470 种蛋白。根据 GOLD 数据库提供的基因序列, 应用 SignalP 3.0 Server 对这 13023 个 ORF 进行 Hidden Markov 模型(HMM)分析。以 NP\_708110 为例(图 2A), 其含有一个信号肽序列, 酶切位点在 N 端 21-22 个氨基酸之间(NA-AE)。根据结果统计, 3 株志贺氏菌分别有 653、691 和 680 种蛋白具有信

号肽序列, 各占全部基因组 ORFs 的 14.71% (653/4439)、16.80% (691/4114)和 15.21% (680/4470)。

2.2 亚细胞定位及螺旋跨膜域分析

对 2024 个含有信号肽序列的 ORF 结合 PSORT-B 和 Cell-Ploc 进行蛋白表达亚细胞定位分析。结果表明, *S. flexneri* 2a301 株中定位于外膜、胞外或者是菌毛上的蛋白有 120 种(外膜: 77; 胞外: 33; 菌毛: 10), 另外两株则分别为 182 种(73; 99; 10)和 124 种(84; 29; 11)。通过 TMHMM 和 Phobius 对这些蛋白进行  $\alpha$ -螺旋跨膜域分析, 3 株菌分别排除了 5、8 和 5 种  $\alpha$ -螺旋形式的跨膜蛋白。图 2B 和 2C 分别为 NP\_709330 的 TMHMM 和 Phobius 分析, 前者预测到 3 个螺旋跨膜域, 但无法区分信号肽序列, 而后者很好地判断了信号肽和 2 个螺旋跨膜域。因此, 若 TMHMM 预测到仅含 1 个跨膜区的 ORF, 要同时符合跨膜区位于 N 端 70 个氨基酸之外, 我们才推断其为  $\alpha$ -螺旋跨膜蛋白。

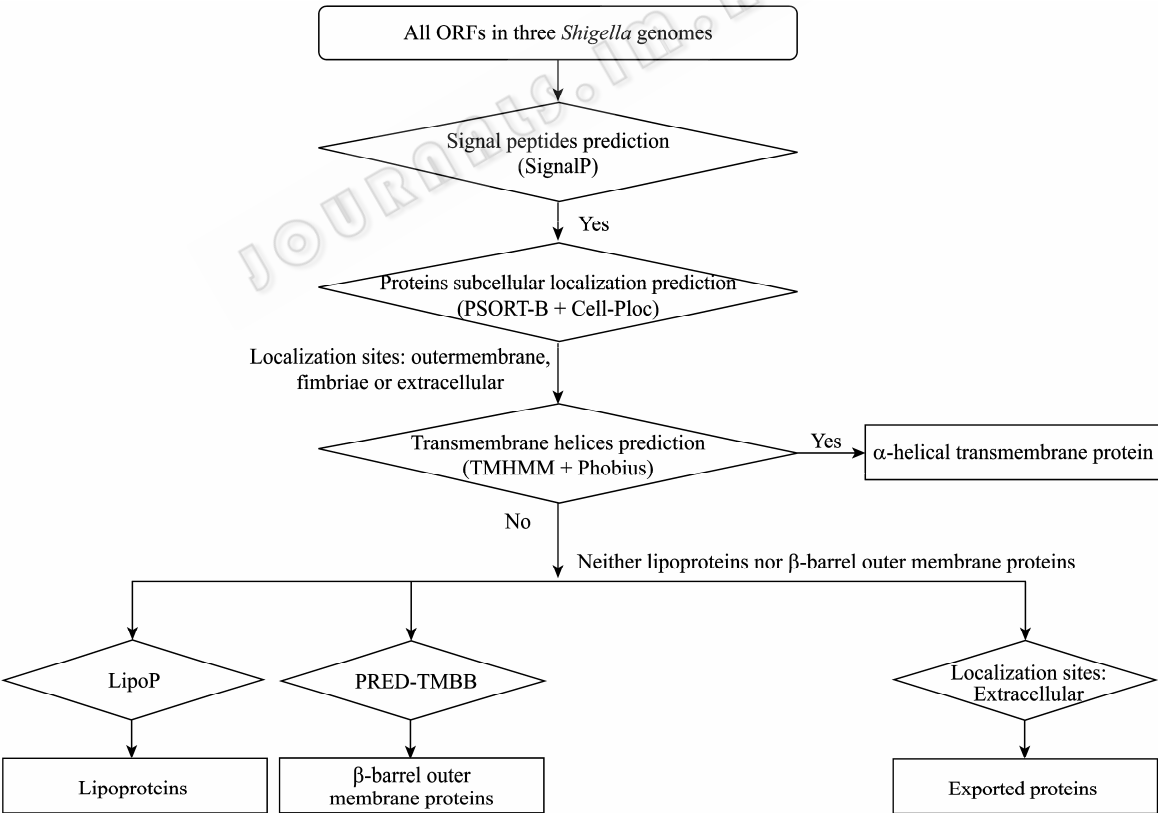


图 1 疫苗候选抗原的生物信息学预测流程图  
Fig. 1 Bioinformatics-based approach to vaccine candidate antigens screening

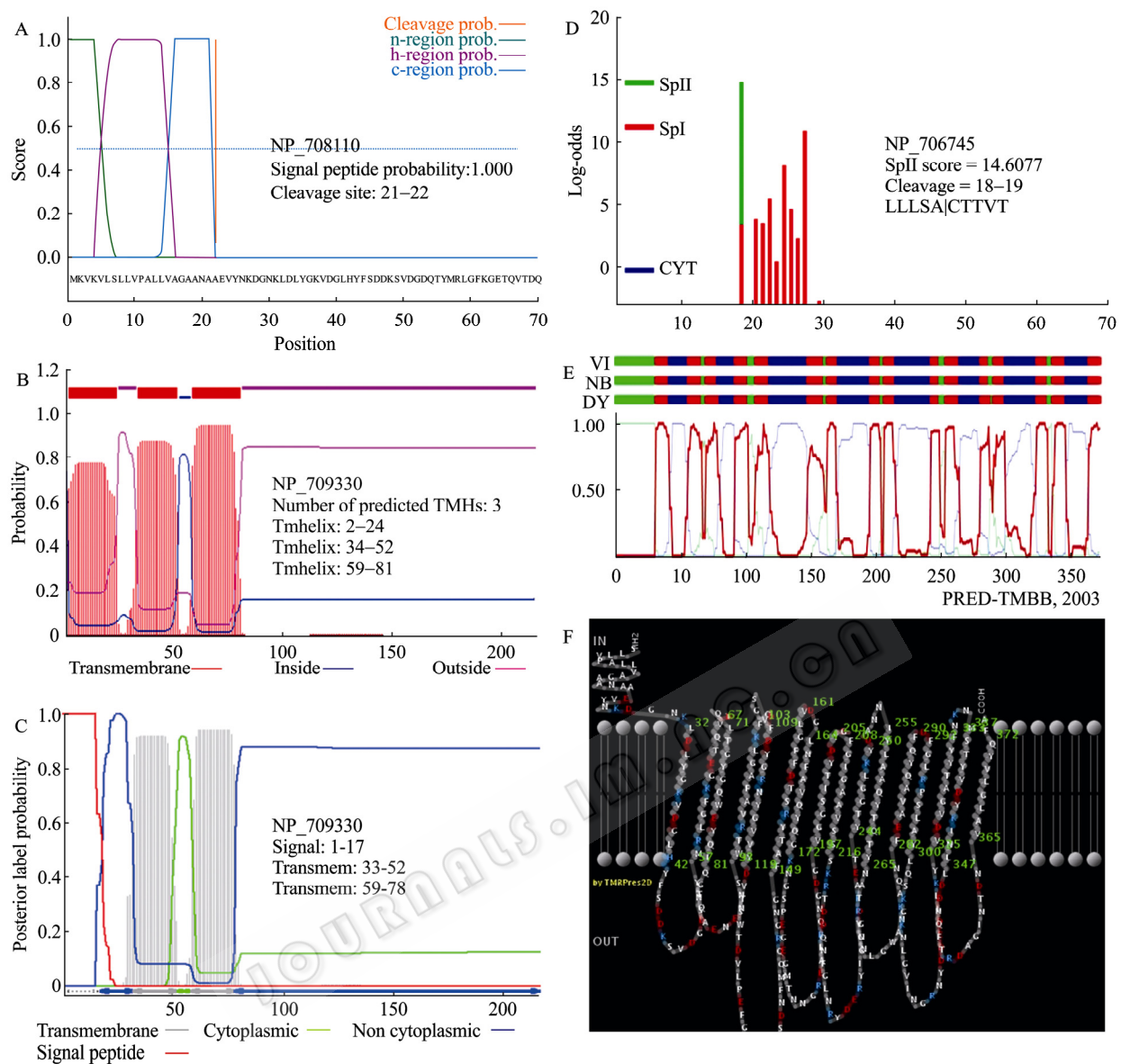


图 2 候选抗原的生物信息学分析

Fig. 2 Screening candidate antigens by bioinformatics methods

Note: A: SignalP; B: TMHMM; C: Phobius; D: LipoP; E-F: PRED-TMBB.

2.3 脂蛋白、 $\beta$ -桶型跨膜蛋白或分泌性蛋白的鉴别

经过上述方法筛选后余下的 ORF 用 LipoP 和 PRED-TMBB 分别进行脂蛋白和  $\beta$ -桶型跨膜蛋白的鉴别。如图 2D 表示 NP\_706745 编码脂蛋白, 其酶切位点位于 N 端 18-19 个氨基酸之间(LLLSA-CTTVT)。图 2E 和 2F 为 PRED-TMBB 分析 NP\_708110 的结果, 展示了该 ORF 编码的  $\beta$ -桶型跨膜蛋白的折叠概率以及二维结构预测。最后, 在 *S. flexneri* 2a301 株中共预测到 21 种脂蛋白和 67 种  $\beta$ -桶型跨膜蛋白, 而 *S. flexneri* 5b8401 株中则为 25 和 67 种, *S. sonnei* Ss046

株中则分别为 29 和 72 种。余下的非编码脂蛋白和  $\beta$ -桶型跨膜蛋白的 ORF, 其又定位于胞外的可推测为分泌性蛋白, 3 株菌中此类蛋白的种数分别为 8、66 和 6 种。

2.4 同源性分析

我们利用 ClustalX 对预测到的脂蛋白、 $\beta$ -桶型跨膜蛋白和分泌性蛋白 3 类候选抗原分别进行同源性序列比对, 对具有 80% 以上相似性的蛋白序列认定为同种蛋白。结果表明筛选到的 361 个 ORF 表现出良好的同源性, 88.37% (319/361) 的候选抗原在两株以上的志贺氏菌中被发现, 其中分别有 20 种脂蛋

表2 三株志贺氏菌全基因组中预测到的共有疫苗候选抗原列表  
Table 2 List of predicted common vaccine candidate antigens in the genomes of three *Shigella* strains

蛋白类型 Protein type	蛋白编号 Accession number			预测功能描述 Predicted function	蛋白编号 Accession number			预测功能描述 Predicted function
	Sj2a301	Sj5b8401	Ss046		Sj2a301	Sj5b8401	Ss046	
脂蛋白 Lipoproteins	NP_706347	YP_688001	YP_309452	Hypothetical protein	NP_708344	YP_689956	YP_311450	Putative outer membrane lipoprotein
	NP_706493	YP_688153	YP_309684	Peptidoglycan-associated outer membrane lipoprotein	NP_708351	YP_689963	YP_311457	Outer membrane protein assembly complex sub-unit YfgL
	NP_706816	YP_688293	YP_309729	Putative pectinesterase	NP_708453	YP_690247	YP_311614	Putative outer membrane lipoprotein
	NP_706745	YP_688382	YP_309828	Putative lipoprotein	NP_708646	YP_690304	YP_311843	Putative lipoprotein
	NP_706828	YP_688437	YP_309885	Putative heat shock protein	NP_708734	YP_690396	YP_312047	Murein transglycosylase C
	NP_707543	YP_689136	YP_310450	Putative outer membrane protein	NP_708948	YP_690536	YP_312103	Putative glycosylase
	NP_707576	YP_689174	YP_310417	Murein lipoprotein	NP_709855	YP_691422	YP_312991	Putative outer membrane efflux protein MdtP
	NP_707956	YP_689551	YP_311006	Putative polysaccharide export protein	NP_709890	YP_691474	YP_312941	Hypothetical protein
	NP_708228	YP_689829	YP_311281	Lipoprotein precursor	NP_706906	YP_688517	YP_309959	Putative function in exopolysaccharide production
	NP_708283	YP_689893	YP_311386	Hypothetical protein	NP_706997	YP_688616	YP_310059	Flagellar basal body L-ring protein
$\beta$ -桶型跨膜蛋白 $\beta$ -barrel outer membrane proteins	NP_706122	YP_687749	YP_309216	Outer membrane protein assembly factor YaeT	NP_708225	YP_689827	YP_311278	Long-chain fatty acid outer membrane transporter
	NP_706241	YP_687869	YP_309301	Outer membrane phosphoprotein E	NP_708846	YP_690448	YP_311992	Outer membrane channel protein
	NP_706299	YP_687953	YP_309399	Outer membrane protein Tsx	NP_709164	YP_690743	YP_312315	Outer membrane porin HofQ
	NP_706439	YP_688093	YP_309537	Outer membrane receptor FepA	NP_709458	YP_691165	YP_312391	Putative ferric siderophore receptor
	NP_706488	YP_688148	YP_309690	Putative homeobox protein	NP_709540	YP_691092	YP_312672	Putative major fimbrial subunit
	NP_706492	YP_688152	YP_309685	Hypothetical protein	NP_709627	YP_691007	YP_312752	Phospholipase A
	NP_706543	YP_688202	YP_309628	Hypothetical protein	NP_709770	YP_691345	YP_312887	vitamin B12/cobalamin Outer membrane transporter
	NP_706848	YP_688458	YP_309906	Outer membrane protein F	NP_709925	YP_691505	YP_313184	Hypothetical protein
	NP_706879	YP_688492	YP_309933	Outer membrane protein A	NP_709978	YP_691554	YP_313120	Hypothetical protein
	NP_706951	YP_688570	YP_310013	Curlin minor subunit	NP_710099	YP_691669	YP_313214	Hypothetical protein
	NP_706993	YP_688610	YP_310053	Flagellar basal body P-ring biosynthesis protein FlgA	NP_706680	YP_688322	YP_309764	Hypothetical protein
	NP_707165	YP_688780	YP_310813	Outer membrane protein W	NP_707513	YP_689108	YP_310480	Hypothetical protein
	NP_707332	YP_688931	YP_310326	Hypothetical protein	NP_709977	YP_691553	YP_313121	Hypothetical protein
	NP_707597	YP_689199	YP_310559	Hypothetical protein	NP_706865	YP_688478	YP_309918	Putative fimbrial-like protein
	NP_707620	YP_689221	YP_310581	Putative outer membrane porin protein	NP_708215	YP_689818	YP_311269	Putative minor fimbrial subunit
	NP_706647	YP_689335	YP_310683	Putative membrane protein precursor	NP_708219	YP_689822	YP_311273	Putative fimbrial-like protein
	NP_707731	YP_690125	YP_311287	Putative membrane protein precursor	NP_706692	YP_688334	YP_309776	Outer membrane protein X
	NP_707819	YP_689425	YP_311316	Putative outer membrane porin protein C precursor	NP_707025	YP_688644	YP_310090	Hypothetical protein
	NP_708110	YP_689709	YP_311155	Outer membrane porin protein C	NP_707997	YP_689594	YP_311048	Putative fimbrial-like protein
	NP_708114	YP_689713	YP_311166	Hypothetical protein	NP_707514	YP_689109	YP_310479	Membrane-associated protein
	NP_708136	YP_689737	YP_311191	Hypothetical protein	NP_709326	YP_690880	YP_312619	Putative lipase
分泌蛋白 Secreted proteins	NP_706515	YP_688173	YP_309662	Putative fimbrial-like protein				

注: Sj2a301: 福氏志贺菌 2a 亚型 301 株; Sj5b8401: 福氏志贺菌 5b 亚型 8401 株; Ss046: 宋内氏志贺菌 046 株。

Note: Sj2a301: *Shigella flexneri* 2a str. 301; Sj5b8401: *Shigella flexneri* 5b str. 8401; Ss046: *Shigella sonnei* Ss046.

白、42 种  $\beta$ -桶型跨膜蛋白和 1 种分泌性蛋白同时存在于 3 株菌中(表 2); 另一方面, 同一菌株全基因组内部也存在一些同源序列, 如推测为 IS1 编码的分泌性蛋白, 不仅共同存在于 *S. flexneri* 2a301 株和 *S. flexneri* 5b8401 株中, 且在两株菌中也各自存在 3 和 62 个同源序列, 可能为该基因在全基因组中的多拷贝形式。此外, 还有 5 种类似的  $\beta$ -桶型跨膜蛋白在 3 株志贺氏菌中普遍存在, 且经 BLASTp 分析该 15 个 ORF 序列之间的相似性在 55% 以上。

综上所述, 通过生物信息学法的筛选, 我们分别从 *S. flexneri* 2a301、*S. flexneri* 5b8401 和 *S. sonnei* Ss046 株全基因组中得到了 96、158 和 107 个 ORF, 它们分别编码一系列的脂蛋白、 $\beta$ -桶型跨膜蛋白或分泌性蛋白, 组成了一个疫苗候选抗原的信息库。上述 ORF 编码的蛋白在 3 株志贺氏菌中呈现良好的同源性, 有 63 种候选抗原在同源性分析中被发现同时存在于 3 株志贺氏菌中。

### 3 讨论

肠道感染每年在世界范围内造成约 170 万–250 万人死亡, 居所有疾病的第 3 位。引起腹泻的病原体及其抗原的多样性给肠道感染的病因诊断与有效疫苗的开发带来了巨大的困难。目前仅有伤寒沙门菌和霍乱弧菌疫苗的效力较好。志贺氏菌疫苗的研制始于 20 世纪 40 年代, 但始终没有成功的产品投放市场。目前正在研制的志贺氏菌减毒活疫苗大部分结合了营养缺陷与毒力基因突变, 如 T32 株、SC602 株等, 这种结合会导致菌株不同程度的减毒。虽然突变株已几经更替, 但安全性与保护效果仍难达到较好的结合点。为了避免由活疫苗产生的安全问题, 亚单位疫苗有望成为未来疫苗研制的更优策略。此类疫苗是基于对表面抗原的清晰认识, 通过单一抗原分子的递呈和识别来诱导机体免疫反应。也就是说, 其首要环节是理想表面抗原的发掘。然而单纯地以试验手段去捕获生物中重要蛋白并进一步深入研究是十分有限的。本项目中, 我们应用生物信息学方法, 首次在全基因组水平上分析了福氏志贺菌 2a 亚型 301 株、5b 亚型 8401 株以及宋内氏志贺菌 046 株的表面抗原, 尝试了解其分子免疫的整体概况, 试图从中发掘若干可进一步实验验证并应用于疫苗研制的新蛋白。

已有报道表明, 志贺氏菌的膜表面蛋白、分泌性蛋白存在大量的免疫抗原<sup>[17–19]</sup>, 可能是由于它们

与机体的免疫系统最近, 因此易激发机体产生免疫反应。在革兰氏阴性菌的双膜结构中, 这些蛋白质分子占据了约 50% 的外膜区域。不同于内膜蛋白跨膜区的  $\alpha$  螺旋结构, 外膜蛋白多为富含  $\beta$  片层结构的  $\beta$ -桶型跨膜蛋白<sup>[20]</sup>, 或是以 N 端的脂类分子锚定在膜上的脂蛋白。无论是脂蛋白还是  $\beta$ -桶型跨膜蛋白, 它们与分泌性蛋白一样都有一个穿膜过程, 即含有信号肽或前导肽序列。基于上述理论, 我们从 3 株志贺氏菌全基因组中分别筛选了 96、158 和 107 个候选抗原, 它们分别编码一系列的脂蛋白、 $\beta$ -桶型跨膜蛋白或分泌性蛋白, 占了各自全基因组 ORFs 的 2.16%、3.84% 和 2.39%, 大大缩小了筛查的范围, 使得后续实验具有一定的针对性。

根据对我国志贺氏菌病流行地区的观察, 其血清型呈现多样性的趋势。而普通的疫苗通常存在血清型特异性免疫保护的问题。如何通过进一步的疫苗改良来覆盖更广泛的血清型是目前多价疫苗研制中最关键的问题。本文报道了 20 种脂蛋白、42 种  $\beta$ -桶型跨膜蛋白和 1 种分泌性蛋白, 它们分别在 3 株菌中存在非常高相似性的同源序列(表 2)。此外, 统计中发现了一种  $\beta$ -桶型跨膜蛋白, 它不仅在 3 株志贺氏菌中普遍存在, 而且在各流行菌株的全基因组中分别找到了 5 个相似的 ORF, 经 BLASTp 分析表明此 15 个 ORF 序列之间的相似性高达 55% 以上。因此, 该蛋白也许能在交叉免疫实验中产生良好的效果。综合以上信息, 所筛选的候选抗原的同源性分析可以为多价疫苗的研制提供一定的信息数据。

Ying 等人利用双向电泳等手段对 *S. flexneri* 2a 2457T 株自然表达的膜蛋白和分泌性蛋白进行的免疫实验中, 得到了一些具有免疫活性的抗原蛋白<sup>[18–19]</sup>。通过序列比对发现, 本次生物信息学法筛选并整理的一批候选抗原中, 已有 4 种蛋白在 *S. flexneri* 2a 2457T 株的免疫实验中得到了验证, 包括外膜蛋白装配因子 YaeT (NP\_835905)、外膜蛋白 A (NP\_836665)、外膜通道蛋白(NP\_838556)以及丝氨酸蛋白酶 SepA (NP\_838462)。前 3 种蛋白在 3 株不同血清型志贺氏菌 *S. flexneri* 2a301、*S. flexneri* 5b8401 和 *S. sonnei* Ss046 株中都存在相似度极高的同源序列(外膜蛋白装配因子 YaeT: NP\_706122、YP\_687749、YP\_309216; 外膜蛋白 A: NP\_706879、YP\_688492、YP\_309933; 外膜通道蛋白: NP\_708846、YP\_690448、YP\_311992), 后一种蛋白也与 *S. flex-*

neri 2a301 株中筛选到的胞外丝氨酸蛋白酶 SepA (NP\_858203)高度同源。除此之外, 还有大量蛋白在普通培养条件下表达量较少或者根本不表达, 因此用传统方法难以验证其免疫原性, 有待在后续工作中通过人工克隆表达、蛋白分离纯化以及免疫实验加以补充。

高通量、高效、高速是生物信息学筛选疫苗候选抗原的几大优点。然而, 有利必有弊, 该方法存在的缺陷也显而易见。本筛选体系的首个步骤即是对信号肽序列的预测。虽然绝大多数蛋白质的穿膜途径都遵守普通蛋白质转运分泌原则, 但是蛋白的分泌过程也不是绝对的, 有些蛋白质的分泌并不需要信号肽的存在<sup>[21]</sup>。其次, 我们在蛋白表达亚细胞定位分析中结合了 PSORT-B 和 Cell-Ploc 两种在线预测工具, 在分析螺旋跨膜域时也综合了 TMHMM 和 Phobius 两种方法的预测结果。在提高准确性的同时, 却可能会漏掉若干有用的蛋白。再者, 生物信息学预测完全依赖于全基因组序列以及现有生物学知识, 难免会产生部分假阳性的结果, 其可靠性必须通过后续实验的验证。

通过以上对志贺氏菌疫苗候选抗原信息的开发和整理, 不仅大大缩小了实验验证的范围, 减轻了筛查的工作量, 而且使疫苗组份的研发对象从点扩大到了面, 并可以充分利用其他项目的研究成果, 同时为其他项目提供数据平台和理论基础。

## 参 考 文 献

- [1] Kotloff KL, Winickoff JP, Ivanoff B, *et al.* Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ*, 1999, **77**(8): 651–666.
- [2] Mei Y, Liu H, Xu J. Cloning and application of genus specific DNA probes for *Shigella*. *Chinese J Epidemiol*, 1989(10): 167–170.
- [3] Pu XY, Pan JC, Wang HQ, *et al.* Characterization of fluoroquinolone-resistant *Shigella flexneri* in Hangzhou area of China. *J Antimicrob Chemother*, 2009, **63**(5): 917–920.
- [4] Yu HL, Chang ZR, Zhang LS, *et al.* Analysis on the status of *Shigella* spp. antimicrobial resistance through data from the National Shigellosis Surveillance System in China, in 2005. *Zhonghua Liu Xing Bing Xue Za Zhi*, 2007, **28**(4): 370–373.
- [5] Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine*, 2001, **19**(17/19): 2688–2691.
- [6] Pizza M, Scarlato V, Masignani V, *et al.* Identification of vaccine candidates against serogroup B *meningococcus* by whole-genome sequencing. *Science*, 2000, **287**(5459): 1816–1820.
- [7] Montigiani S, Falugi F, Scarselli M, *et al.* Genomic approach for analysis of surface proteins in *Chlamydia pneumoniae*. *Infect Immun*, 2002, **70**(1): 368–379.
- [8] Ross BC, Czajkowski L, Hocking D, *et al.* Identification of vaccine candidate antigens from a genomic analysis of *Porphyromonas gingivalis*. *Vaccine*, 2001, **19**(30): 4135–4142.
- [9] Wizemann TM, Heinrichs JH, Adamou JE, *et al.* Use of a whole genome approach to identify vaccine molecules affording protection against *Streptococcus pneumoniae* infection. *Infect Immun*, 2001, **69**(3): 1593–1598.
- [10] Bendtsen JD, Nielsen H, von Heijne G, *et al.* Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*, 2004, **340**(4): 783–795.
- [11] Gardy JL, Laird MR, Chen F, *et al.* PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, 2005, **21**(5): 617–623.
- [12] Chou KC, Shen HB. Large-scale predictions of gram-negative bacterial protein subcellular locations. *J Proteome Res*, 2006, **5**(12): 3420–3428.
- [13] Krogh A, Larsson B, von Heijne G, *et al.* Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*, 2001, **305**(3): 567–580.
- [14] Kall L, Krogh A, Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*, 2004, **338**(5): 1027–1036.
- [15] Juncker AS, Willenbrock H, Von Heijne G, *et al.* Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci*, 2003, **12**(8): 1652–1662.
- [16] Bagos PG, Liakopoulos TD, Spyropoulos IC, *et al.* PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res*, 2004, **32**(Web Server issue): W400–W404.
- [17] Liao X, Ying T, Wang H, *et al.* A two-dimensional proteome map of *Shigella flexneri*. *Electrophoresis*, 2003, **24**(16): 2864–2882.
- [18] Ying T, Wang H, Li M, *et al.* Immunoproteomics of outer membrane proteins and extracellular proteins of *Shigella flexneri* 2a 2457T. *Proteomics*, 2005, **5**(18): 4777–4793.
- [19] Ying TY, Wang JJ, Wang HL, *et al.* Immunoproteomics of membrane proteins of *Shigella flexneri* 2a 2457T. *World J Gastroenterol*, 2005, **11**(43): 6880–6883.
- [20] Wimley WC. The versatile beta-barrel membrane protein. *Curr Opin Struct Biol*, 2003, **13**(4): 404–411.
- [21] Sargent F, Stanley NR, Berks BC, *et al.* Sec-independent protein translocation in *Escherichia coli*. A distinct and pivotal role for the TatB protein. *J Biol Chem*, 1999, **274**(51): 36073–36082.