

金黄色葡萄球菌蛋白质相互作用网络及功能

刘琦, 姜春雷, 许正超, 徐辉, 赵锐, 乔代蓉, 曹毅*

(四川大学生命科学学院, 四川省生物信息与代谢工程共享实验平台, 成都 610064)

摘要 【目的】金黄色葡萄球菌是一种革兰氏阳性菌,是目前最难以对付的病菌之一。它能引起多种感染,特别是在医院环境中。近年来,抗药性金黄色葡萄球菌传染更加严重,已成为公共卫生威胁。由于以前对于金黄色葡萄球菌的实验性研究大都是基于单个基因或者蛋白进行的,为了更好的研究这个物种,有必要从整体上把握金黄色葡萄球菌的蛋白作用机理。【方法】采用系统发生谱、操纵子法、基因融合法、基因邻近法、同源映射法等5种计算方法预测金黄色葡萄球菌蛋白质相互作用网络。【结果】从蛋白组的角度构建了金黄色葡萄球菌蛋白相互作用网络,并对网络进行功能分析。【结论】网络的分析表明金黄色葡萄球菌的蛋白质相互作用网络也服从 scale-free 属性,发现了 SA0939、SA0868、rplD 等重要的蛋白。通过对金黄色葡萄球菌的重要的细胞壁合成和信号转导调控蛋白局部网络分析,发现了一些对这两个系统十分重要的蛋白分子,这些信息将为更好的了解金黄色葡萄球菌的致病机理和开发新的药物靶点提供指导。

关键词: 金黄色葡萄球菌; 蛋白质相互作用; 系统发生谱; 基因融合法; 同源映射

中图分类号: Q816 **文献标识码:** A **文章编号:** 0001-6209(2009)01-0056-08

金黄色葡萄球菌(*Staphylococcus aureus*)是人类的一种重要病原菌,可引起肺炎、伪膜性肠炎、败血症和脓毒症等感染。由于该致病菌很容易获得抗生素耐药性,对青霉素及目前广泛使用的甲氧西林(methicillin)均具有抗性,能抵抗所有青霉素、甲氧西林及其他抗 β -内酰胺酶的青霉素。从 Jevons 首次发现耐甲氧西林金黄色葡萄球菌(methicillin resistant *Staphylococcus aureus*, MRSA)以来^[1],MRSA 的感染几乎遍及全球,已成为医院内和社区感染的重要病原菌之一。

尽管已有大量关于金黄色葡萄球菌 MRSA 的耐药性报道,但大都处于单个分子水平的研究。为了从整体上对金黄色葡萄球菌的蛋白分子作用机制进行研究,更好的为以后的实验性研究提供指导,本研究拟从蛋白组的水平上运用生物信息学的方法构建蛋白质相互作用网络,并对其进行网络功能结构分析。

生物体内的任何活动都要归结于蛋白质相互作用。蛋白质相互作用为了解生命活动的机制,提供了必要的信息,也为注释未知蛋白的生物学功能提供了线索。同时生物体内的任何蛋白分子都不是独立的在细胞中行使被赋予的功能,而是与其他蛋白质分子形成大的蛋白复合体,在特定的时间和空间完成特定的功能,因此构建蛋白质相互作用网络对于研究这些分子机制是很有效的,对于致病菌潜在的致病靶点的鉴定也是有帮助的。随着物种的基因组测序的完成,可以通过各种高通量实验(酵母双杂交(yeast two-hybrid)、免疫共沉淀(co-immunoprecipitation)等,以及计算生物学方法预测并构建蛋白质相互作用网络,从系统生物学的角度进行研究。到目前为止已经有很多物种的蛋白质相互作用网络被绘制出来,如酿酒酵母(*Saccharomyces cerevisiae*)^[2]、大肠杆菌(*Escherichia coli*)^[3]、幽门螺杆菌

基金项目: 国家自然科学基金(30871321),教育部新世纪人才项目(NCET-05-0785)

* 通信作者。Tel/Fax: +86-28-85412842; E-mail: caoyi_01@163.com

作者简介: 刘琦(1984-)男,湖北人,硕士,主要从事生物信息学研究。E-mail: liuqi_1984@yahoo.com.cn

收稿日期: 2008-07-25; 修回日期: 2008-09-28

(*Helicobacter pylori*)^[4]、钩端螺旋体(*leptospira*)^[5]等。

考虑到各种实验方法通常具有耗时、花费高等缺点,并且假阳性也不明显的低于生物信息学方法,因此,本研究采用生物信息学的方法预测金黄色葡萄球菌的蛋白质相互作用网络。现在的计算方法主要有系统发生谱(phylogenetic profile)^[6]、基因邻近法(gene neighbor method)^[7]、操纵子法(operon method)^[8]、基因融合法(gene fusion method)^[9]、同源映射法(interolog)^[10]等。这些计算方法都是基于基因组上下文的方法,每种方法都有自己的预测灵敏度和准确度,它们之间可以相互补充,因此,通过结合上面几种方法可以得到很高覆盖率和较理想可信度的预测结果^[11]。在构建得到金黄色葡萄球菌的蛋白质相互作用网络的基础上,对重要功能模块进行分析,发现金黄色葡萄球菌重要的蛋白分子,并可预测未知功能蛋白质的功能,为揭示金黄色葡萄球菌致病机理和耐药机制的实验研究提供更多的信息。

1 材料和方法

1.1 实验数据

选取 *Staphylococcus aureus* N315 菌株的基因组,其蛋白序列及其相关信息从 TIGR(<http://www.tigr.org/>)下载。系统发生谱法中选取参考物种时所需要用到的物种进化信息和基因组序列分别从 NCBI(<http://www.ncbi.nlm.nih.gov/>)的 taxonomy 和 TIGR 下载。基因融合法中用到的结构域信息从 PFAM(<http://pfam.sanger.ac.uk/>)数据库获得。同源映射法中需要的已知蛋白相互作用数据从 DIP(Database of Interacting Proteins,<http://dip.doe-mbi.ucla.edu/>) BIND(Biomolecular Interaction Network Database,<http://www.bind.ca/>) MINT(Molecular Interactions Database,<http://mint.bio.uniroma2.it/mint/>) IntAct(Protein interaction database and analysis system,<http://www.ebi.ac.uk/intact/>)下载。蛋白功能分类信息从 NCBI 的 COG 下载。

1.2 实验方法

采用系统发生谱、操纵子法、基因融合法、基因邻近法、同源映射法五种方法构建金黄色葡萄球菌的蛋白质相互作用网络。系统发生谱是基于相互作用蛋白在参考物种里协同表达的原理来实现的。参考物种是利用物种的进化信息,并用 Sun 等^[12]描述的方法进行选取,最后选取 186 个参考物种。系统发生谱的构建参考 Date 等^[13]的方法,最后用互信息值作为衡量两个蛋白是否具有相互作用的标准。操纵子法是基于功能相关蛋白基因在原核生物中通常

形成操纵子的原理,利用基因组上基因之间的距离和方向信息,如果两个蛋白基因在基因组上方向相同,并且它们之间的距离小于 100 bp,就被预测在同一个操纵子上^[8]。

基因融合法是指一个物种中的两个蛋白如果功能相关,具有相互作用,在进化过程中常常在另外一个物种中融合成一个蛋白质^[4]。首先通过将 *Staphylococcus aureus* N315 的蛋白序列与 PFAM 结构域数据进行 blast^[14] 比对,得到其功能结构域信息,再与 Swiss-prot 所有蛋白的 PFAM 功能结构域信息进行比较,得到蛋白相互作用信息。

基因邻近法是基于功能相关蛋白,由于相互之间的进化压力,在许多物种基因组上的相互距离都是保守的原理^[7],采用 WIT 系统(<http://wit.mcs.anl.gov/WIT2/>)的方法。同源映射法是通过物种之间的直系同源关系,将其他物种中已知的经过实验得到的蛋白相互作用信息映射到查询物种中来的方法^[10]。首先利用双向最佳匹配法(Bidirectional Best Hit)^[15]构建 *Staphylococcus aureus* N315 与大肠杆菌和幽门螺杆菌的直系同源对(ortholog),再把原核生物大肠杆菌和幽门螺杆菌中已知的相互作用信息映射过来,得到蛋白相互作用信息。

通过 5 种方法得到的所有 *Staphylococcus aureus* N315 的蛋白相互作用信息,利用 cytoscape 软件和 pajek 进行可视化,并联合 RBGL packages(R project,<http://www.r-project.org/>)进行网络拓扑学分析。由于 *Staphylococcus aureus* N315 的 GO 注释不完全,而 swiss-prot 注释较全,网络质量验证采用的是 keyword recovery 法^[16],即评估预测的相互作用的两个蛋白之间 swiss-prot 关键词注释的相似性。如果预测的相互作用的两个蛋白之间的关键词注释至少有一个相同,则认为相互作用具有很高的可信度。先构建与真实网络度分布相同(相同节点和连接数量)的随机网络,然后比较真实网络与随机网络样本之间的 keyword recovery 值来验证网络的质量。

2 结果和分析

2.1 整体蛋白质相互作用网络

整合五种算法,以 *Staphylococcus aureus* N315 菌株的基因组及其相关信息为材料,进行蛋白质相互作用网络预测,得到一个包含 2329 个节点,14496 个连接的蛋白质相互作用网络(图 1-A),网络节点覆盖 *Staphylococcus aureus* N315 基因组的 89%(2329/2619),在构成网络的 2329 个蛋白中有 1560 个功能已知蛋白(图 1-B 中彩色节点),769 个是属于功能未

知或功能为预测的蛋白(灰色节点),整个网络包括许多独立的子网络,其中最大的网络包含 2193 个蛋白,14355 个连接。在其他小的子网络中,最大的包含 8 个蛋白,最小的包含 2 个蛋白。将 *Staphylococcus aureus* N315 组成网络的所有蛋白进行 COG 功能分类,共可分成 22 个 COG 分类(图 2)。这些分类中假定蛋白(hypothetical protein)的含量最高,占 15.5%,其次是功能为预测的(9.1%)或未知的蛋白(8.4%),这三类功能注释差的蛋白在图中都用灰色

色来标示。含量最低的为染色质结构(0.1%)和细胞运动类蛋白(0.2%)。由网络图(图 1)可以看出颜色相同的节点表现出较明显的聚集特性,即功能相关蛋白在网络中通常聚集在一起而行使特定功能。在对网络进行具体分析时,首先采用 keyword recovery 法进行网络的质量评估,结果表明预测的网络的 keyword recovery 值(0.248)大于随机网络样本具有极显著性($P < 0.01$),证明预测的网络具有很高的可信度。

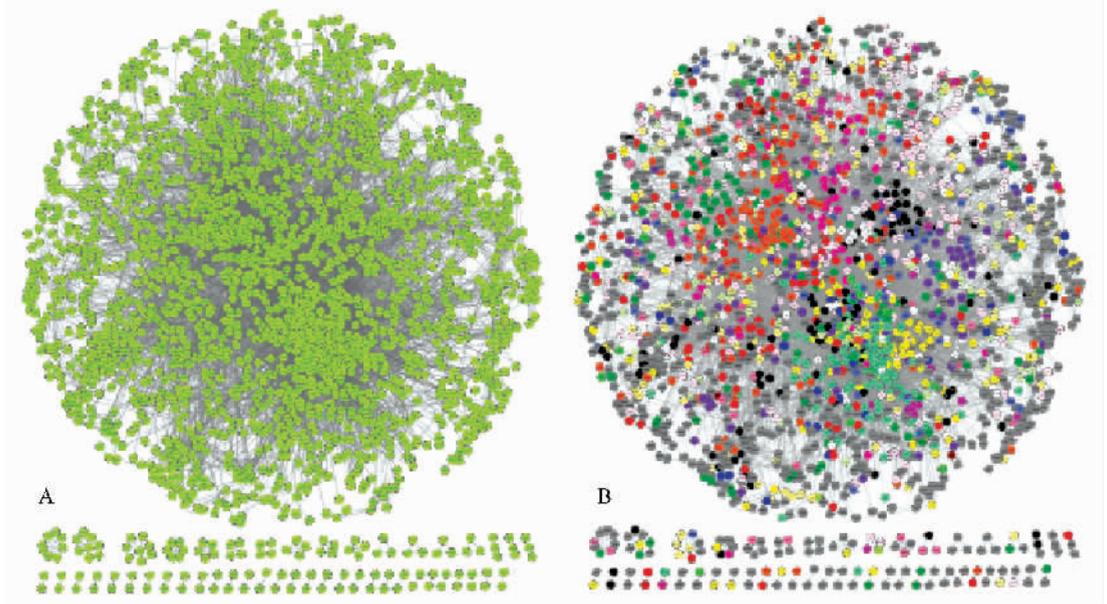


图 1 预测得到的金黄色葡萄球菌蛋白质相互作用网络

Fig.1 The predicted protein interaction network of *Staphylococcus aureus*. A: The whole network. B: The nodes of network are colored according to their COG classifications.

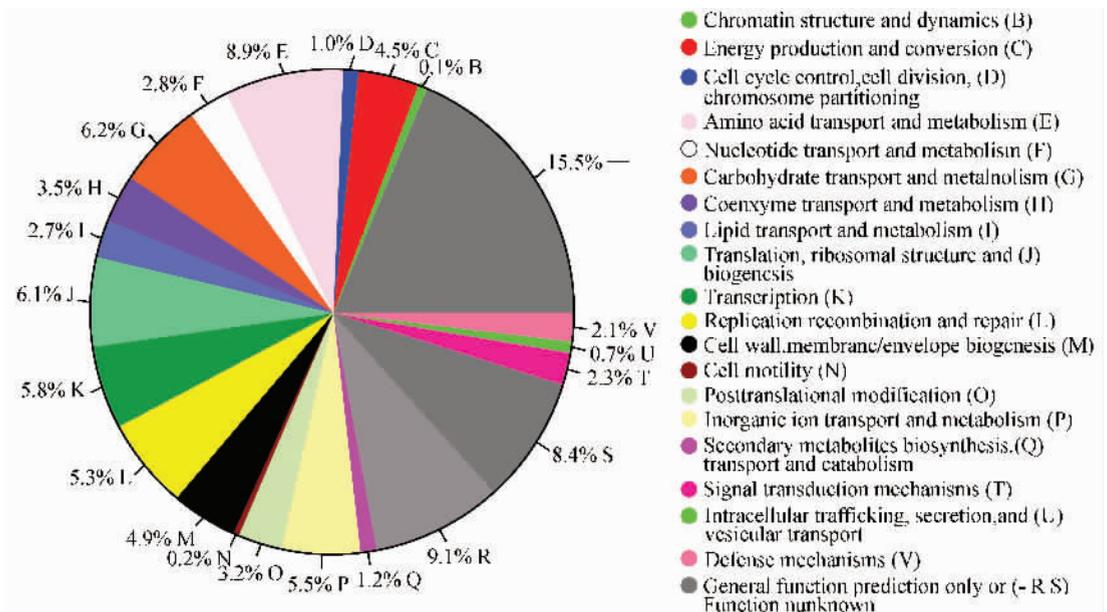


图 2 网络中蛋白在各类 COG 分类中的含量图

Fig.2 The content of each COG of the proteins in the protein interaction network. The kind represented by gray color has three COG classifications.

2.2 网络拓扑分析

从已经得到的物种的蛋白质相互作用网络可知,绝大部分生物学网络都服从 scale-free 网络的属性^[17],即网络中的少数节点具有大量的连接,大部分节点都只有很少的连接,这些少数节点是网络关键性节点(hub)。通过对本网络的拓扑分析,发现金黄色葡萄球菌的蛋白质相互作用网络也具有 Scale-free 属性(图 3-A),用幂次法则(power law)拟合得到 $y = 1558.2 * x^{-1.524}$, $r^2 = 0.880$ (图 3-A 中斜线),其中 x 为节点的度(k),也即与该节点直接相连的节点的个数, y 为度为 k 的节点的个数。图 3-B 为网络的聚类系数 $c(k)$ 分布图,平均聚类系数为 0.411,从图中可以看出度大的点具有较低的聚类的趋势,表现出一定的层次聚类的性质。对网络进行路径长度分析后,可以看出路径的长度大部分在 4 左右(图 3-C),服从 scale-free 网络的小世界(small world)效应。

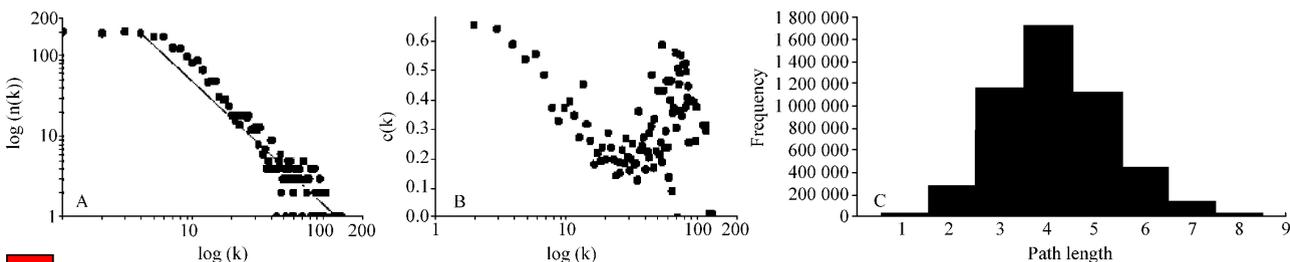


图 3 金黄色葡萄球菌蛋白质相互作用网络的拓扑学分析

Fig.3 The topology analysis of *Staphylococcus aureus* protein interaction network. A degree distribution analysis. B cluster analysis. C path length analysis.

表 1 网络中几种关键蛋白的注释

Table 1 The annotations of some essential proteins

ORF name	Gene name	COG	Degree	Function
SA0939	- *	P	138	hypothetical protein
SA0868	-	P	133	hypothetical protein
SA2046	rplD	J	124	50S ribosomal protein L4
SA1109	nusA	K	103	transcription elongation factor NusA
SA0993	uvrC	L	92	excinuclease ABC subunit C
SA1513	polA	L	92	DNA polymerase I
SA0729	tpi	G	90	triosephosphate isomerase
SA0460	pth	J	89	peptidyl-tRNA hydrolase

* - :The protein has only ORF name.

2.3 重要网络模块分析

通过对蛋白质相互作用网络的分析,结合实验性的资料能够在金黄色葡萄球菌蛋白组的水平挖掘出很多有用的信息。金黄色葡萄球菌的致病过程是一个复杂的过程,其中在不用的感染时期表达成分不同的细胞壁合成系统以及起到信号转导和调控作用的信号转导调控系统发挥着核心的作用^[19]。下面着重分析这两个与金黄色葡萄球菌耐药性相关度

计算 *Staphylococcus aureus* N315 的蛋白质相互作用网络中的每个节点的度,发现有 SA0939, SA0868, rplD 节点的度(> 100)都远大于其他节点,说明它们在 *Staphylococcus aureus* N315 蛋白相互作用网络中对于网络的功能的执行和稳定都是很十分重要的节点,可能在细胞中发挥很大的作用,可能是潜在的药理学靶点,它们的注释见表 1。Soo 等通过比较基因组学和等位基因替换突变分析得到了 *Staphylococcus aureus* N315 菌株的 122 个关键基因^[18],这些关键基因产物在本蛋白网络的度平均值为 24,通过比较这些节点与在网络中多次随机挑选相同数量的节点组成的随机样本的度分布,发现 Soo 研究得到的 122 个关键基因的度大于本研究挑选的随机节点样本也具有极显著性($P < 0.01$)因此这 122 个关键基因也具有很高的参考价值,其中度最大的前 5 个蛋白的注释见表 1。

比较高的功能分类系统,进行局部蛋白相互作用网络分析。

2.3.1 细胞壁合成系统 金黄色葡萄球菌的致病性主要是由不同的细胞表面和细胞壁成分、毒素和酶共同作用产生的^[20],有研究指出对于单个毒力基因的突变并不会减少菌株的毒性^[21],这就表明是基因产物相互之间在共同行使功能。 β -内酰胺类抗生素是目前主要用于治疗金黄色葡萄球菌 MRSA 感染的

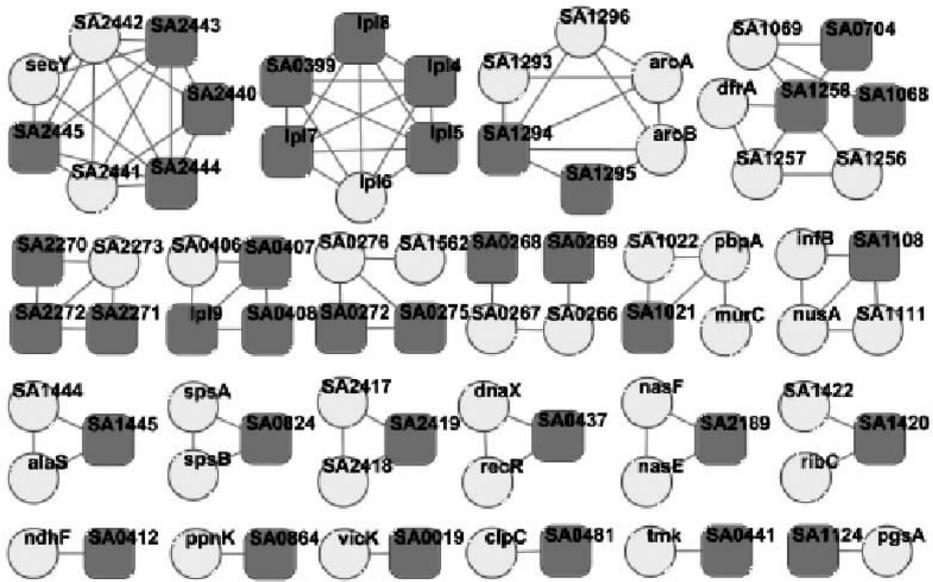


图6 通过相互作用预测蛋白功能

Fig.6 Function prediction of protein based on protein interaction.

通过分析整体网络并着重分析了与金黄色葡萄球菌耐药性和致病性相关度高的细胞壁合成蛋白的局部相互作用网络和信号转导和调控蛋白局部相互作用网络,发现了一些十分重要的蛋白分子,可以作为以后潜在的药物靶点分子,但还需通过进一步的生物学实验予以验证,这将为研究金黄色葡萄球菌的耐药性以及寻找新的更好的药物靶点提供资料。

参考文献

- [1] Jevons MP. Celbenin resistant staphylococci. *British Medical Journal* ,1961 242 :124 - 125.
- [2] Schwikowski B ,Uetz P ,Fields S. A network of protein-protein interactions in yeast. *Nature Biotechnology* 2000 ,18 (12) :1257 - 1261.
- [3] Butland G ,Alvarez JM ,Li J ,et al. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* 2005 433 (7025) 531 - 537.
- [4] Rain JC ,Selig L ,De Reuse ,et al. The protein-protein interaction map of *Helicobacter pylori*. *Nature* ,2001 ,409 : 211 - 215.
- [5] Sun JC ,Xu JL ,Cao JP ,at al. Prediction and systematic study of protein-protein interaction networks of *Leptospira interrogans*. *Chinese Science Bulletin* 2006 51(11) :1296 - 1305.
- [6] Pellegrini M ,Marcotte EM ,Thompson MJ ,et al. Assigning protein functions by comparative genome analysis :protein phylogenetic profiles. *Proceeding of the National Academy of Sciences of the United States of America* ,1999 ,96 :4285 - 4288.
- [7] Overbeek R ,Fonstein M ,D 'Souza M ,et al. The use of gene clusters to infer functional coupling. *Proceeding of the National Academy of Sciences of the United States of America* , 1999 96 2896 - 2901.
- [8] Strong M ,Mallick P ,Pellegrini M ,at al. Inference of protein function and protein linkages in *Mycobacterium tuberculosis* based on prokaryotic genome organization :a combined computational approach. *Genome Biology* 2003 4 :R59.
- [9] Marcotte EM ,Pellegrini M ,Ho-Leung N ,at al. Detecting protein function and protein-protein interactions from genome sequences. *Science* ,1999 285 :751 - 753.
- [10] Haiyuan Yu ,Nicholas M. Luscombe ,at al. Annotation Transfer Between Genomes :Protein-Protein Interologs and Protein-DNA Regulogs. *Genome Research* 2004 ,14 :1107 - 1118.
- [11] Chen Y ,Xu D. Computational analysis of high throughput protein-protein interaction data. *Current Protein and Peptide Science* 2003 4(3) :159 - 181.
- [12] Sun JC ,Li YX ,Zhao ZM ,Phylogenetic profiles for the prediction of protein-protein interactions :How to select reference organisms ? . *Biochemical and Biophysical Research Communications* 2007 353 985 - 991.
- [13] Date SV ,Marcotte EM. Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nature Biotechnology* 2003 21(9) :1055 - 1062.
- [14] Altschul SF ,Gish W ,Miller W ,at al. Basic local alignment search tool. *Journal of Molecular Biology* ,1990 ,215(3) : 403 - 410.
- [15] Tatusov R ,Koonin E ,Limpan D. A genomic perspective on

- protein families. *Science* ,1997 ,278 :631 – 637.
- [16] Marcotte EM ,Pellegrini M ,Thompson MJ ,et al. A combined algorithm for genome-wide prediction of protein function. *Nature* ,1999 ,402 :83 – 86.
- [17] Albert R. Scale-free networks in cell biology. *Journal of Cell Science* ,2005 ,118 :4947 – 4957.
- [18] Soo KK ,Lee JY ,Song JH ,at al. Screening of Essential Genes in *Staphylococcus aureus* N315 Using Comparative Genomics and Allelic Replacement Mutagenesis. *Journal of Microbiology and Biotechnology* ,2006 ,16(4) :623 – 632.
- [19] Cheung AS ,Bayer G ,Zhang H ,et al. Regulation of virulence determinants in vitro and in vivo in *Staphylococcus aureus* ,*FEMS Microbiology Letters* ,2004 ,1649 :1 – 9.
- [20] Lowy FD. *Staphylococcus aureus* infection. *The New England Journal of Medicine* ,1998 ,339 :520 – 532.
- [21] Nilsson IM ,Hartford O ,Foster T ,at al. Alpha-toxin and gamma-toxin jointly promote *Staphylococcus aureus* virulence in murine septic arthritis. *Infection and Immunity* ,1999 ,67 :1045 – 1049.
- [22] De Lencastre H ,Oliveira D ,Tomasz A ,Antibiotic resistant *Staphylococcus aureus* : a paradigm of adaptive power. *Current Opinion Microbiology* ,2007 ,10 :428 – 435.
- [23] Arvidson S , Tegmark K. Regulation of virulence determinants in *Staphylococcus aureus*. *International Journal of Medical Microbiology* ,2001 ,291 :159 – 170.

Analysis of protein interaction network and function of *Staphylococcus aureus*

Qi Liu ,Chunlei Jiang ,Zhengchao Xu ,Hui Xu ,Rui Zhao ,Dairong Qiao ,Yi Cao *

(College of Life Science ,Sichuan University ,Sichuan Public Experiment Platform of Bioinformatics and Metabolic Engineering ,Chengdu 610064 ,China)

Abstract [Objective] *Staphylococcus aureus* is a member of Gram positive bacteria ,but is also one of common pathogens that are most difficult to treat. It infects human skin ,soft tissue ,mucous membrane ,bone and joint ,especially in the nosocomial environment. Because studies on *Staphylococcus aureus* before were largely based on a single gene or protein ,it is necessary to study this organism from the whole genome. **[Methods]** We used bioinformatics methods including five computational methods (phylogenetic profile ,gene neighbor method ,operon method ,gene fusion method ,interolog) to predict the protein interaction network of *Staphylococcus aureus* . **[Results]** We constructed the protein interaction network of *Staphylococcus aureus* and studied its function. **[Conclusion]** Through the network analysis ,we found that the protein interaction network of *Staphylococcus aureus* was subject to scale-free property and a number of very important proteins ,such as SA0939 ,SA0868 ,rplD. Through the analysis of the important cell wall synthesis and signal transduction and regulation local networks ,we also found some very important proteins. Such information will help us better understand pathogenic mechanism and develop new drug targets of *Staphylococcus aureus* .

Keywords : *Staphylococcus aureus* ; protein interaction ; phylogenetic profile ; gene fusion method ; interolog

(本文责编 张晓丽)