痢疾志贺氏菌亚群的比较基因组学研究与进化分析*

宾 文[①③†] 杨 娥[②③†] 彭俊平[③] 张笑冰[③] 王 璟[③] 杨 剑[③] 董 杰[③]
楚雍烈[②] 张景海[①] 金 奇[③**]

(① 沈阳药科大学制药工程学院, 沈阳 110016; ② 西安交通大学医学院, 西安 710061; ③ 病毒基因工程国家重点实验室, 北京 100176)

摘要 利用大肠杆菌 K12 MG1655 株全基因组 ORFs 和痢疾志贺氏菌 A1 型 Sd51197 株特异性 ORFs 探针制备的芯片, 研究了痢疾志贺氏菌 13 个血清型代表株的基因组组成. 结果显示, 该血清群成员的基因组中包含有 2654 个保守的源于大肠杆菌 ORFs; 共同缺失了 219 个涉及前噬菌体基因、分子伴侣、特异性 O 抗原合成等大肠杆菌原有的基因; 并通过水平转移获得了一些特异性基因, 如 II 型分泌系统相关组分、铁转运相关因子等. 根据基因组组成所作的进化树, 发现 A1, A2, A8 和 A10 这四型菌与其他痢疾志贺氏菌亲源关系较远. 研究所得结果为进一步深入探索痢疾志贺氏菌的生理过程、致病性和进化奠定了基础.

关键词 痢疾志贺氏菌亚群 **DNA** 芯片 比较基因组学

目前, 志贺氏菌属感染引起的细菌性痢疾的发病率仍居我国肠道传染病之首, 依据O抗原结构不同, 分为 A, B, C, D 四群, 即志贺氏菌 (*Shigella dysenteriae*)、福氏菌(*S. flexneri*)、鲍氏 (*S. boydii*)及宋内氏菌(*S. sonnei*). 其中, A群有 13 个血清型, 痢疾志贺氏菌A1 型可导致非常严重的菌痢, 致死率约为 7%[1]. 不同血清型的痢疾志贺氏菌在生化特征和毒力表型上差异较大, 其基因组组成也存在着差异, 而 传统的分类方法不能反映菌株间遗传关系的远近及进化关系. 随着细菌全基因组序列测定的进行[2], 特别是本实验室对痢疾志贺菌(*S. dysenteriae*) A1 型 SD51197 株基因组核苷酸序列测定的完成(另文发表), 所得到的大量重要数据和信息为揭示痢疾志贺氏菌亚群内各型间亲源关系远近及进化提供了重要线索.

2004-11-10 收稿, 2005-01-10 收修改稿
* 国家重点基础研究发展规划项目(批准号: G1999054105)、国家高技术研究发展计划项目(批准号: 2001AA223011)及国际科技合作重点项目(批准号: 2001AA223116)资助
** 联系人, E-mail: mailto:zdsys@sina.com
† 并列第一作者

为了揭示A群痢疾志贺氏菌与*E. Coli* K12的进化关系，及A群痢疾志贺氏菌 13 个血清型间的基因组成和亲源关系，我们制备了包含有大肠杆菌K12 MG1655 株全基因组ORFs[3]、痢疾志贺菌(*S. dysenteriae*) A1型Sd51197株特异性ORFs的DNA芯片，与其他12型痢疾志贺氏菌代表株杂交，较深入地对痢疾志贺氏菌亚群内各型间进行了比较以及进化基因组学分析.

1 材料和方法

1.1 菌株与培养

痢疾志贺氏菌 A1 型 CN51197(Sd51197)由中国疾病预防控制中心传染病预防控制研究所提供；其余 12 个痢疾志贺氏菌血清型代表株由天津南开大学王磊提供(表 1). 所有的菌株均在液体 LB 培养基中37℃振荡培养过夜，用于分离基因组 DNA.

表 1 实验所用菌种

血清型	来源	编号
A2	NLEP[a)]	LSPQ3473
A3	临床分离	LSPQ2436
A4	NLEP	LSPQ3474
A5	NLEP	LSPQ3679
A6	NLEP	LSPQ3475
A7	NLEP	LSPQ3680
A8	NLEP	LSPQ3681
A9	NLEP	LSPQ3682
A10	NLEP	LSPQ3476
A11	NLEP	LSPQ3684
A12	NLEP	LSPQ3683
A13	NLEP	LSPQ3477

a) NLEP，加拿大国家肠道病原实验室

1.2 DNA 芯片的制备

采用 array designer 软件设计引物，分别以 MG1655，Sd51197 染色体为模板扩增探针，用 Cartisen 芯片机点样于玻片上. 紫外交联固定，1，2-二氯乙烷封闭备用.

1.3 探针的制备及杂交

杂交所用细菌 DNA 采用 Wizard® Genomic DNA Purification Kit 进行分离. A1 型的杂交以 MG1655 株为参考，其余 12 型痢疾志贺氏菌的比较基因组学杂交以 MG1655 和 Sd51197 株作为参考. 待测染色体 DNA，背景参考各 2 μg，分别以 Cy5，Cy3-DCTP (Amersham)过夜标记后，用 QIA-quick PCR purifcation kit 纯化. 与芯片 65℃杂交过夜.

1.4 数据采集和分析

芯片经2×SSC, 0.1% SDS, 1×SSC, 0.2×SSC溶液漂洗后，采用GenePix 4100A scanner (Axon Instruments, Inc.)扫描，所得到的TIFF图像文件中背景和点的亮度采用Genepix Pro 5.0 software (Axon Instruments, Inc.) 进行量化，并将两个通道中所有数值中位数比值调整为1，对样品数值进行了归一化，那些信号值小于背景值+2倍背景标准差的点不进入后续的数据分析. 为了方便分析所得实验结果，将全基因组芯片杂交结果表述为一个包含两种状态的矩阵: 基因的存在或缺失信号分别采用1和0表示，所得数据采用Spotfare软件进行分析.

1.5 进化分析

将基因组杂交所得数据导入 Phylip 软件包，使用最大简约法对 13 型痢疾志贺氏菌构建进化树，并对此结果进行 100 次的 bootstrap 检验(所得分值标于进化树分支处).

2 结果

2.1 信号标准的判定

本研究中用于制备芯片的 PCR 产物长度均在 150 bp 以上，去除了 MG1655 全部 4279 个 ORFs 中的 91 个小于 150 bp 的 ORFs，杂交后标化时去除了 23 个信号值小于背景值+2 倍背景标准差的 ORFs，最后进行数据分析的有 4165 个 MG1655 ORFs，代表了其基因组 ORFs 的 97%，并包含了 294 个 Sd51197 株特异性 ORFs.

根据已经测序的MG1655和Sd51197株与芯片杂交结果确定了杂交信号的判定标准：以点在两个通道信号值 Cy3，Cy5 比值≥0.3 为阳性；< 0.3 为阴性，所有的数据都根据 3 次杂交结果的信号来确定，3 次杂交结果中出现两次及两次以上为阳性，否则为阴

性．根据芯片内所设对照(SpotReport™ A. thaliana Array Validation Systems)及测序株杂交所得数据进行综合分析，发现芯片杂交的假阳性和假阴性均在 1%左右，对那些参考株的信号值小于阴性对照的 ORFs，采用制备芯片所用引物与进行了 PCR 验证，最终以 PCR 扩增的结果确定其存在与否.

2.2　痢疾志贺氏菌型基因组间共有骨架的构成

图 1 显示，MG1655 株 ORFs 在痢疾志贺氏菌各型间分布有一定差异，LSPQ3473 株(A2 型)与 MG1655 株基因组成较为接近，含有所检测的 4165 个 ORFs 中的 3663 个，而 Sd51197 株仅有 3213 个；有 2654 个 MG1655 ORFs 在痢疾志贺氏菌中是保守的，构成了痢疾志贺氏菌亚群的“共有骨架”，其中除 248 个 ORFs 所编码蛋白与目前已知的任何蛋白质都没有显著的同源性之外，其余 2406 个 ORFs 所编码的产物的功能涵盖了细菌生理的各种代谢途径和基本结构组成，如能量代谢、碳水化合物代谢等，以及参与细菌生长繁殖的各种蛋白质、酶及调控因子等，涉及核糖体结构、细胞分裂与染色体分离、核苷转运与代谢、翻译，翻译后修饰、蛋白质折叠等的 ORFs 与 MG1655 差别不大，其功能分类详见表 2.

2.3　缺失的基因

与 MG 1655 比较，痢疾志贺氏菌亚群中缺失的 ORFs，从 Sd51197 株的 953 个到 LSPQ3473 株的 502 个不等，其中包括整个操纵子的完全缺失和部分缺失两种类型. 这些 ORFs 涉及葡糖转移酶 wbbK，胞嘧啶脱氨酶 codA，渗透酶 codB，胆碱脱氢酶 betA，硫

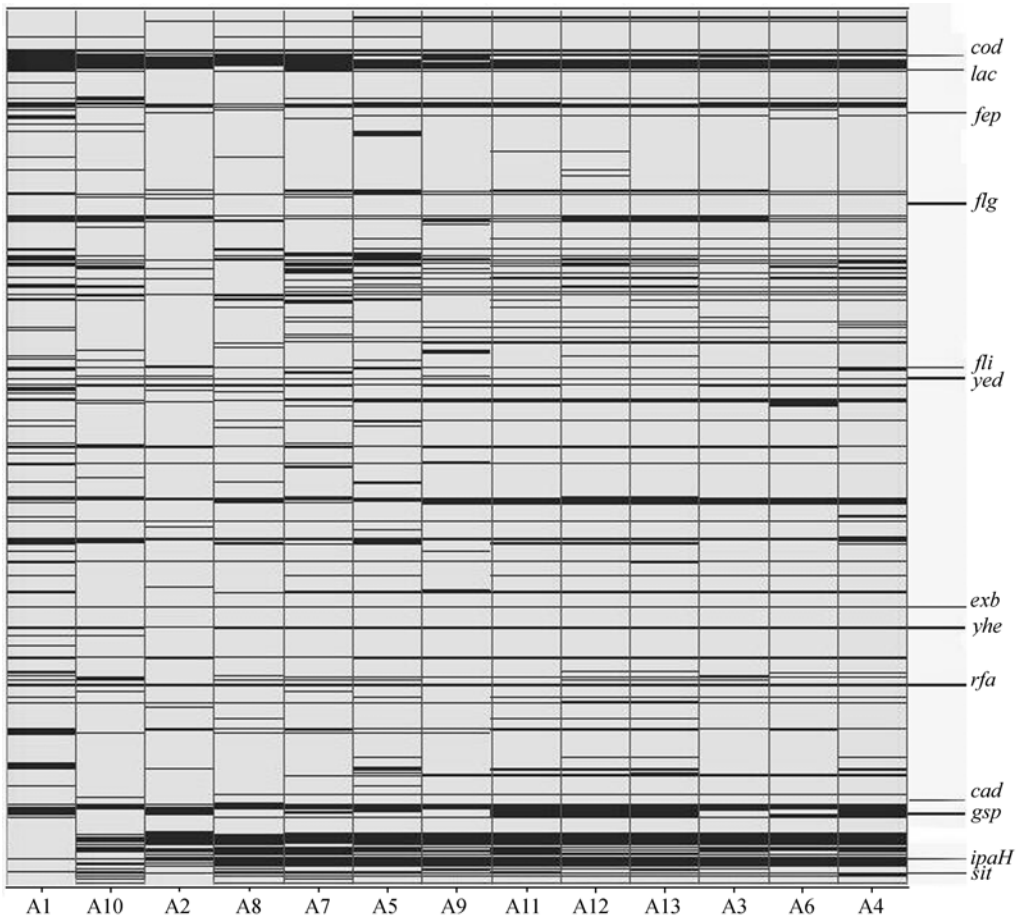


图 1　痢疾志贺氏菌遗传的多样性

13 型志贺痢疾杆菌染色体微阵列杂交结果的比较，柱状分枝图是依据芯片上 ORFs 存在或缺失所作的分层聚类分析图，由 Genespring 软件产生数据后用 Spotfare 软件生成，代表型与型之间亲源关系的远近. 图中灰色的线表示 ORFs 存在，黑色线表示 ORFs 缺失或高度变异

表 2 共有骨架编码蛋白在 COG 系统中的分类

		功能分类	Sd51197	MG1655
COGs 分类功能明确				
	信息存储和加工			
	J	翻译，核糖体结构等	157	168
	K	转录	174	279
	L	DNA 复制，重组与修复	115	175
	细胞加工			
	D	细胞分裂与染色体分离	30	34
	O	翻译后修饰，蛋白质折叠，伴侣	112	127
	M	细胞包膜合成，外膜	150	235
	N	细胞运动与分泌	52	135
	P	无机离子转运与代谢	127	190
	T	信号传导	85	134
	代谢			
	C	能量产生与转换	191	273
	G	糖类转运与代谢	194	368
	E	氨基酸转运与代谢	276	353
	F	核苷转运与代谢	71	87
	H	辅酶代谢	112	122
	I	脂代谢	55	82
	Q	次级代谢产物的生物合成，转运及分解代谢	51	90
功能不明确				
	R	仅有功能提示	226	342
	S	功能未知	228	319
未在 COGs 分类中			248	652
合计			2654	4165

基半乳糖苷乙酰转移酶 lacA，黄嘌呤脱氢酶 yagT，甲基转移酶 yjhP 等. 相当数量的糖转运或代谢基因发生缺失，乳糖分解代谢的 lac 操纵子中，lacA，Y 全部缺失，而 lacZ 仅在 sd51197, LSPQ3680(A7 型), LSPQ-3681(A8 型), LSPQ3682(A9 型), LSPQ3684(A11 型)中存在. 此外，有 219 个 ORFs 编码前噬菌体，分子伴侣及特异性的 O 抗原合成等相关基因如 rfa 等发生缺失，但这 219 个基因中不包括鞭毛组装相关基因，这类操纵子所涉及的基因仅在各型中部分缺失或全部保留.

2.4 获得的基因

与 Sd51197 株特异性 ORFs 杂交结果显示，不同痢疾志贺氏菌获得的 Sd51197 株特异性 ORFs 数量和种类有很大差异：在 294 个痢疾志贺氏菌 Sd51197 特异性 ORFs 中，有 164 个在 LSPQ3476 株(A10 型)中呈阳性，而在 LSPQ3681 株和 LSPQ3477 株(A13 型)中仅有 103 个，只有 62 个为痢疾志贺氏菌所共有，这 62 个 ORFs 涉及 II 型分泌系统相关组分、IS 元件、铁转运相关因子以及侵袭性质粒抗原 *ipa H*.

2.5 进化分析

图 2 为分别根据共同骨架 ORFs 和全部 ORFs 所构建的进化树. 两个树的拓扑结构仅有细微的差异，但在两个进化树中，sd51197，LSPQ3473，LSPQ3681 和 LSPQ3476 均偏离其他痢疾志贺氏菌.

3 讨论

以往的进化分析是在检测某些保守基因的变异、rRNA/rDNA和序列标记分析的基础上进行的，菌株基因组间组成从另外一个层面反映进化关系. 基于一株已测序细菌制备的全基因组DNA芯片为亲缘关系相近的未测序菌间的比较基因组学研究提供了一个非常重要的工具[5]. 它可以扩大和延伸测序株所得到的信息，使我们快速地了解与其关系相近的菌间的相互关系，从而可提供一个关于它们基因组的“快

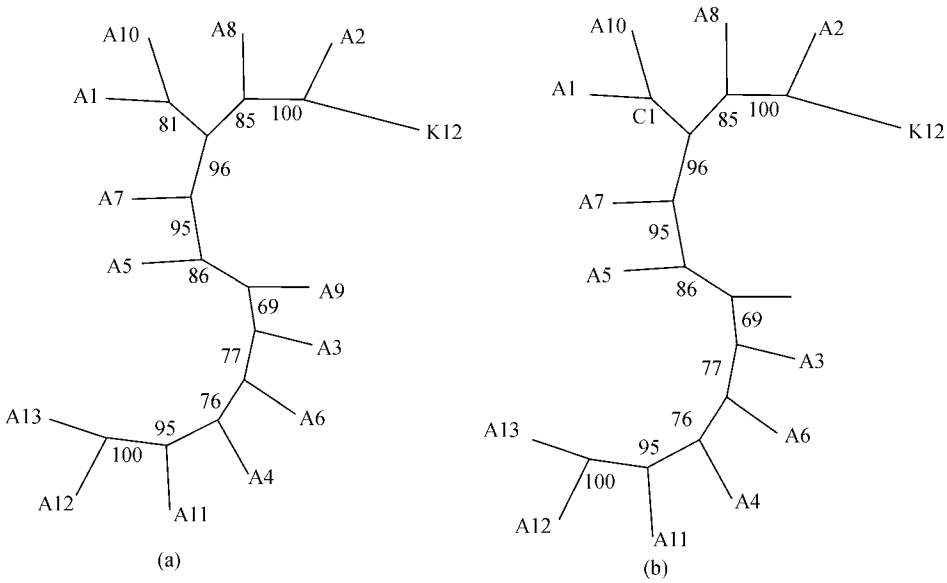


图 2　13 型痢疾志贺氏菌的种系进化树

为杂交数据输入Phylip软件包，使用最大简约法构建，bootstrap值对应于 100 次重复，bootstrap值高于 70%表明对应点的概率值大于 95%[4].

图 2(a)和(b)分别根据共有骨架 ORFs 和芯片上的全部的 ORFs 的数据所作

照”，是研究细菌的遗传多样性、进化的有力工具. 由于芯片的数据是基于杂交的结果得到的，所以，它不能为我们提供关于有无基因变异、有无染色体重排或是否是假基因等情况，但其高通量优势所产生的海量数据为鉴定细菌基因和等位基因的功能和生物学意义提供了帮助，因此破译基因组数据非常重要. 目前 DNA 芯片已应用于研究基因的功能和相互作用，可供我们研究致病菌的进化和从大体上了解关系相近的菌间差异.

志贺氏菌被认为在起源和进化关系上与大肠杆菌关系密切[6]，目前普遍认为痢疾杆菌是从多个独立的大肠杆菌祖先起源后经历了 7~8 次趋同进化而形成的. 在共有骨架中，作为基本代谢元件的核糖体蛋白基因如*frr*，*rpl*，*rpm*和*rps*等在痢疾志贺氏菌亚群中完整存在，嘌呤代谢的*gua*系统、芳香族氨基酸合成的*aro*系统及DNA修复酶*radC*等对痢疾志贺氏菌保存生命活力至关重要的基因均为阳性. 编码毒力表型调控因子OmpR/EnvZ双组分调节系统的基因在痢疾志贺氏菌亚群中均存在，当痢疾志贺氏菌暴露于高渗的肠道微环境时，双组分调节系统可调节质粒编码的侵袭表型的表达，与其致病性密切相关. 同样，编码毒力决定因子的 *dsbA* 基因也在全部痢疾志贺氏菌存在，其编码产物对细菌表面的多聚体结构(菌毛和黏附分子等)以及毒素的生物合成非常关键.

痢疾志贺氏菌亚群在漫长的进化过程中，通过基因水平转移，其基因组正经历着种内平行进化过程. 外源基因的获得以及基因的不断缺失导致痢疾志贺氏菌亚群内的剧烈分化，推动了其基因组的进化，如 II 型蛋白分泌系统的演变. 本研究中发现在大肠杆菌中作为“隐匿基因”存在的*yheDEFGHIJK-hofFGHD-pshM* II 型蛋白分泌系统[7]的基因，在除 A2 外的其他痢疾志贺氏菌亚群中均全部缺失；而在 A1 中发现的可能的 II 型分泌系统*gsp*系统，在痢疾志贺氏菌亚群中大部分呈阳性. 此系统存在于多种致病菌中，如产毒型大肠杆菌中的 II 型蛋白分泌系统与分泌外毒素有关.

一些在肠道中生存必需的基因在没有生存压力的选择下，逐渐退化，如鞭毛相关基因. 痢疾志贺氏菌亚群中有关鞭毛生物合成基因*fliFJP*，*flgCEFKL*在 A1，A4 和A5 中半数呈阴性，其余型中仅少数基因呈

阴性，以往的研究表明由于点突变、IS序列的插入等使这些基因不能表达[8]，从而其可被宿主免疫系统辨认的抗原减少，但这些框架基因的存在可用以探索痢疾志贺氏菌的进化起源. 同时，外源基因的获得使痢疾志贺氏菌亚群适应新的生态微环境. 在A1，这表现在获得侵袭性大质粒和可分泌志贺氏毒素[9]，编码志贺氏毒素的基因*stx*在A1 以外的痢疾志贺氏菌中仅A7存在*stx1A*，这与以往的研究中发现仅A1可产生志贺氏毒素一致. 由于A1 的*stx*基因是通过溶原性转换从噬菌体中获得的[10]，A7 中存在*stx1A*这一事件及众多痢疾杆菌趋同进化[11]的证据提示我们，其他痢疾志贺氏菌也可以经历类似的事件，可能在将来会得到产生痢疾毒素的稳定遗传，从而加强其致病能力，这需要我们密切关注.

在进化过程中，痢疾志贺氏菌不仅获得了质粒上的毒力相关基因，而且还通过缺失丢掉了某些基因[12]，在染色体上形成不可逆转的“黑洞”[13]，从而增强毒力，例如痢疾志贺氏菌亚群中都缺失了*ompT*基因. *ompT*基因编码一种表面蛋白水解酶，可降解外膜表面蛋白*vir G*，从而影响细菌在细胞间的扩散. 同样的例子还有赖氨酸脱羧酶基因*cadA*，其催化反应产物是尸胺，后者可降低细菌的毒力而抑制痢疾杆菌肠毒素. 实验中发现*cadABC*在痢疾志贺氏菌亚群中仅A8 完整存在, Maureli等人的研究表明, *cadA*这一抗毒力基因的失活在不同型别的痢疾杆菌中是独立发生的，是趋同进化的.

痢疾志贺氏菌亚群在进化过程中的多样性反映在很多方面，如铁离子转运系统的差别. 铁离子对细菌的生存和致病性非常重要，它是细菌多种抗氧化防御组分的重要辅助因子. 在宿主的细胞内外，游离铁是非常有限的，芯片杂交结果显示，痢疾志贺氏菌亚群可通过不同的铁载体来攫取铁，它们中均有铁蛋白受体基因*fhuE*，但A2 缺失了*fepA*，A1，A10 和A8 缺失了*fhuA*. 痢疾志贺氏菌亚群中均有为Fe(III)-铁载体复合物从细胞质膜进入细胞周质提供能量的*TonB-ExbB-ExbD*系统和细胞周质转运系统之*fhuBCD*，仅A2 缺失了另一细胞周质转运系统*fepBCDG*，同时A2 也缺失了帮助从铁载体释放Fe^{3+}的fes. 涉及二价铁转运的*feoAB*在痢疾志贺氏菌中都存在，另一二价铁转运系统sitABCD在A1, A2, A5, A9, A1 和A11 中完整存在，其他痢疾志贺氏菌亚群则缺失了*SitD*. 同时我们还发现, A1 和A10 还拥有*ShuA*，后者可以结合宿主细胞内溶解的血红素，帮助细菌获得铁离子[14]. 痢疾志贺氏菌亚群存在多种铁离子获得系统，从侧面反映了铁离子对该菌生长和致病的重要性，同时也反映了不同痢疾志贺氏菌在进化过程中的多样性，而A2 较其他痢疾志贺氏菌亚群缺失了一些铁转运元件这一事件也值得我们进一步探索.

根据芯片的数据，我们对痢疾志贺氏菌亚群作了进化分析，其中，无论采取共同骨架ORFs还是采用全部ORFs所构建的进化树，A1，A2，A8 和A10 这四型均偏离其他痢疾志贺氏菌亚群. 这与Gulietta等人[15]利用持家基因对痢疾杆菌所作进化分析的结果基本一致，提示在痢疾杆菌分化为四个血清群之前，它们可能与其他痢疾志贺氏菌来源于不同的祖先菌，但这尚需进一步深入研究.

参 考 文 献

1 Bennish M L, Wojtyniak B J. Mortality due to shigellosis: community and hospital data. Rev Infect Dis, 1991, 4(Suppl): S245~251

2 Jin Q, Yuan Z H, Xu J G, et al. Genome sequence of *Shigella flexneri* 2a: Insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucl Acid Res, 2002, 30(20): 4432~4441[DOI]

3 Blattner F R, Plunkett G, Bloch C A, et al. The complete genome sequence of Escherichia coli K-12. Science, 1997, 277: 1453~1462[DOI]

4 Hittis D M, Bult J J. An empirical test of bootsrrapping as a method for assessing confidence in phylogentic analysis. Syst Boil, 1993, 42: 182~192

5 刘 红，彭俊平，杨 剑，等. 志贺氏菌属各亚群菌株基因组“共有骨架”组成的分析. 科学通报，2003，48(23): 2451~2456[摘要][PDF]

6 Rolland K, Lambert-Zechovsky N, Picard B, Denamur E. Shigella and enteroinvasive *Escherichia coli* strains are derived from distinct ancestral strains of *E. coli*. Microbiology, 1998, 144: 2667~2672

7 Mark J P, Roy R C, Lan R H, et al. Genomic analysis of secretion system. Curr Opin Micobioly, 2003, 6: 519~527[DOI]

8 Akira Tominagal, Mohamed A, Mahmoud1 H, et al. Characterization of cryptic flagellin genes in *Shigella boydii* and *Shigella dysenteriae*. Genes Genet Syst, 2001, 76: 111~120[DOI]

9 Alexandra Unkmeir, Herbert Schmidt. Structural analysis of phage-

borne stx genes and their flanking sequences in shiga toxin-producing *Escherichia coli* and *Shigella dysenteriae* type 1 strains. Infectiong and Immunity, 2000, 68(9): 4856~4864[DOI]

10 Herold S, Karch H, Schmidt H. Shiga toxin-encoding bacteriophages-genomes in motion. Int J Med Microbiol, 2004, 294(2-3): 115~121[DOI]

11 Satoru Fukiya, Hiroshi Mizoguchi, Toru Tobe, et al. Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. J Bacteriology, 2004, 186(12): 3911~3921[DOI]

12 张笑冰, 刘　红, 杨　帆, 等. 福氏2a志贺氏菌301株基因组水平的缺失基因分析. 科学通报, 2003, 48(6): 587~592[摘要] [PDF]

13 Maurelli A T, Fernandez R E, Bloch C A, Rode C K, Fasano A. “Black holes” and bacterial pathogenicity: a large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc Natl Acad Sci USA, 1998, 95: 3943~3948[DOI]

14 O’Malley S M, Mouton S L, Occhino D A, Deanda M T, Rashidi J R, Fuson K L, Rashidi C E, Mora M Y, Payne S M, Henderson D P. Comparison of the heme iron utilization systems of pathogenic vibrios. J Bacteriol, 1999, 181: 3594~3598

15 Pupo G M, Lan R, Reeves P R. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proc Natl Acad Sci USA, 2000, 97(19): 10567~10572[DOI]