

一株新铜绿假单胞菌噬菌体 SRT6 的分离以及全基因组序列分析^①

刘子辰 李 骑 张福康 张春龙 马鸿芮 伊廷旭 高晓萌 樊祥宇

(济南大学 生物科学与技术学院, 山东 济南 250022)

摘 要 分离一株新铜绿假单胞菌噬菌体,进行生物学特征分析,全基因组测序和比较基因组学分析.了解该噬菌体与其他假单胞菌噬菌体之间的亲缘关系以及它的潜在应用价值.利用双层平板法分离纯化噬菌体,利用酚氯仿抽提基因组,利用 Illumina Hiseq 测序平台对基因组测序,利用生物信息学进行基因组和比较基因组学分析.分离到一株新的烈性铜绿假单胞菌噬菌体 SRT6,其基因组全长 91 364 bp,G+C 含量 49.3%,存在 182 个蛋白质编码基因,其中的 96 个与已知功能的蛋白质具有相似性,无 tRNA 和 tmRNA.比较基因组分析发现,噬菌体 SRT6 属于一个新物种.分离鉴定出一株烈性铜绿假单胞菌噬菌体 SRT6,并发现它属于一个新种噬菌体,为未来利用噬菌体治疗耐药铜绿假单胞菌感染打下坚实的基础.

关键词 铜绿假单胞菌;铜绿假单胞菌噬菌体;SRT6;全基因组测序;比较基因组学分析

中图分类号 Q939

文献标识码 A

0 引言

铜绿假单胞菌又称绿脓杆菌,是土壤中常见的细菌之一,也是呼吸道疾病、创伤、免疫受损机体等的易感细菌^[1].该细菌广泛存在于各种空气、水、土壤以及动物的呼吸道和肠道中,对不良环境和抗生素的耐受力非常强,其临床治疗非常困难^[2],该细菌也常常是 ICU 病房中造成人死亡的重要致病菌.因此,对该细菌的研究显得尤其重要.现在自然界中新物种的分离还在如火如荼地开展^[3,4].噬菌体正是其中最热门的物种.噬菌体作为细菌、古菌或另外原核生物的病毒,是被 William Twort 和 Feliz dHerelle 分别于 1915 年和 1917 年所发现^[5].它们在自然界中广泛存在,有研究人员预测地球上噬菌体的数目可能达到 10^{13} .目前已经完成全基因组测序的噬菌体仅仅 2 000 个左右,该数值远远低于噬菌体总数.随着细菌耐药性问题的日益严重,作为可替代抗生素成为新型抗菌剂的噬菌体,引起了国内外研究者的重视^[6].铜绿假单胞菌噬菌体的遗传和分类具有显著多样性,将它们用于铜绿假单胞菌分型和治疗的研究已经有数十年的历史.截至 2018 年 5 月 17 日,已有 96 个不同的铜绿假单胞菌噬菌体全基因组数据上传至 NCBI 基因组数据库中.目前只有少数铜绿假单胞菌噬菌体的种属被明确划分,大多数噬菌体并未被明确分类.因此,新的铜绿假单胞菌噬菌体的分离和全基因组测序分析有助于加深我们对这种噬菌体进化和生理特征的理解^[7].

本研究介绍了一株新分离的铜绿假单胞菌噬菌体 SRT6.对其进行了全基因组测序和分析,发现它的基因组大小是 91 364 bp(G+C 含量为 49.3%),存在 182 个 ORF,其中的 96 个 ORF 所编码的蛋白质与已知功能的蛋白质有同源性,没有鉴定到 tRNA 编码基因.对其进行了比较基因组学分析发现,它和铜绿

① 收稿日期:2018-03-18

基金项目:国家自然科学基金青年科学基金资助项目(31600148);山东省优秀中青年科学家科研奖励基金(BS2014YY031);济南大学科研基金(XBS1519, XKY1633)资助

通讯作者:樊祥宇,男,汉族,博士,副教授,研究方向:噬菌体基因组学、噬菌体-宿主菌互做等,E-mail:fxysnd@126.com.

假单胞菌噬菌体 C11 具有最大的相似性,但仍然有 4 处基因组水平上的不同. 可见本研究所分离出的铜绿假单胞菌噬菌体 SRT6 是一株新的噬菌体. 本研究有助于揭示噬菌体种群的演化规律,为病毒的多样性和进化提供了新见解,还为未来的噬菌体治疗和抗铜绿假单胞菌药物的研发打下了坚实的基础.

1 实验材料和方法

1.1 材料

1.1.1 土样及实验室所用菌株. 以铜绿假单胞菌临床菌株作为宿主菌. 在山东省济南市济南大学甲子湖畔,取土壤表层 1-2 cm 的表层土并将其放置于干燥的锥形瓶中,待用.

1.1.2 主要试剂的配置. 噬菌体 buffer^[8]: 分别用电子天平称取 1.21 g Tris-HCl, 1.75 g NaCl, 0.41 g $MgCl_2 \cdot 6H_2O$, 0.04 g $CaCl_2$, 均置于锥形瓶中,溶解并定容至 200 mL,调节 pH 到 7.5, 121℃ 高温高压蒸汽灭菌 20 min, 保存备用.

水琼脂: 用电子天平称取 0.70 g 琼脂粉, 加去离子水溶解并定容至 100 mL, 121℃ 高温高压蒸汽灭菌 20 min. 之后在超净台内添加灭菌后的 $CaCl_2$, 调节浓度至 2 mM, 保存备用.

LB 液体培养基: 用电子天平称取 1.0 g 蛋白胨, 0.5 g 酵母提取物, 1.0 g 氯化钠, 加去离子水溶解混匀, 并最终定容至 100 mL, 121℃ 高温高压蒸汽灭菌 20 min, 保存备用.

LB 固体培养基: 用电子天平称取 1.0 g 蛋白胨, 0.5 g 酵母提取物, 1.0 g 氯化钠, 2.0 g 琼脂, 加去离子水溶解混匀, 并最终定容至 100 mL, 121℃ 高温高压蒸汽灭菌 20 min, 保存备用.

1.2 实验方法

1.2.1 噬菌体分离. 称取 10 g 样品至已灭菌的锥形瓶内, 分别向其中加入 10 mL 噬菌体 buffer 和 10 mL 临床铜绿假单胞杆菌对数期悬浊液, 混匀, 37℃ 过夜培养. 将瓶内液体移至离心管中, 调节离心机温度为 4℃, 以 5 000 g 的转速离心 10 min, 用孔径为 0.22 μm 的过滤器对上清液进行过滤除菌, 得到噬菌体原液.

取 1 mL 铜绿假单胞杆菌对数期悬浊液和 10 μL 噬菌体原液到试管中, 再加入 4 mL 水琼脂, 混匀, 平铺在固体 LB 平板上. 37℃ 正置过夜培养.

挑取生长状况良好, 轮廓清晰的单一噬菌斑至 100 μL 噬菌体 buffer 中, 倍比稀释. 分别取每一种稀释浓度液 10 μL 与 1 mL 临床铜绿假单胞杆菌悬浊液以及 4 mL 水琼脂混合均匀, 平铺在固体 LB 平板上, 37℃ 过夜培养. 重复以上过程纯化 3-5 次, 直至同一平板内只出现单一噬菌斑形态. 将噬菌体命名为 SRT6.

1.2.2 噬菌体纯化与基因组提取. 挑取生长状况良好的单一噬菌斑至 100 μL 噬菌体 buffer 中, 倍比稀释. 分别取每一稀释浓度液 10 μL 与 1 mL 临床铜绿假单胞菌悬浊液以及 4 mL 水琼脂混合均匀, 平铺在固体 LB 平板上, 37℃ 过夜培养. 观察噬菌斑的生长状况, 选取生长良好且全部覆盖有噬菌斑的平板, 保留所用稀释度的噬菌体液, 用以扩大培养. 吸取适量的噬菌体液, 将其与 1 mL 临床铜绿假单胞杆菌悬浊液、4 mL 水琼脂混合均匀, 平铺在固体 LB 平板上(20 个平板), 37℃ 过夜培养. 之后在每一平板中加入 5 mL 噬菌体 buffer. 4℃ 条件下放置过夜后, 大量收集噬菌体液.

在 4℃ 条件下将大量收集的噬菌体液 3 500 g 离心 5 min. 取上清液, 加入终浓度为 10% 的 PEG8000 和终浓度为 1 M 的 NaCl, 4℃ 下放置过夜. 之后在 4℃ 条件下 3 500 g 离心 5 min, 得到噬菌体的片状沉淀. 将其重悬于 5 mL 噬菌体 buffer 中, 4℃ 放置 5 h, 在此期间每隔 10 min 用枪轻轻吹打, 使其溶解. 在 40℃ 条件下 3 500 g 离心 6 min, 吸取上清液. 用孔径为 0.22 μm 的滤膜过滤后得到纯化的噬菌体液.

取纯化的噬菌体液, 用平衡酚对噬菌体颗粒连续抽提纯化, 直至有机相与水相界面无白色析出物. 用乙醇沉淀基因组, 并用超纯水再次将所得的基因组溶解. 得到噬菌体 SRT6 的基因组.

1.2.3 噬菌体的生理特征的分析. 将吸取的适量的纯化噬菌体液, 将其和 1 mL 临床铜绿假单胞杆菌悬浊液以及 4 mL 水琼脂混合均匀, 平铺在固体 LB 平板上, 37℃ 过夜培养. 选取生长状况良好的单一噬菌斑, 观察研究其形态^[9].

1.2.4 噬菌体 SRT6 基因组测序. 制备好噬菌体 SRT6 基因组, 利用 1% 琼脂糖凝胶电泳检测基因组 DNA. 确定基因组质量合格后, 送上海锐翌生物科技有限公司利用标准流程进行全基因组的 Illumina HiSeq 测序. 将原始下机数据过滤得到高质量的 Clean Data. 使用 SOAPdenovo 软件对 Clean Data 进行组装. 最终测序深度为 16,463×, 覆盖度为 100%. 铜绿假单胞菌噬菌体 SRT6 的基因组已上传至 Genbank, 登录号为 MH370478.

1.2.5 噬菌体基因组的序列分析. 利用软件 Geneious 11 (<http://www.geneious.com/default,83,home.sm>) 分析 SRT6 的核酸序列组成. 通过 Glimmer v3.02 软件预测噬菌体的基因编码序列^[10], 并通过 NCBI 中的 blastp 对每一个 orf 所编码的蛋白质进行功能预测. 通过 tRNAscan-SE Search Server 在线网站 (<http://lowelab.ucsc.edu/tRNAscan-SE/>) 预测噬菌体 SRT6 基因组上的 tRNA 以及 tmRNA. 通过在线软件 Berkeley Drosophila Genome Project (http://www.fruitfly.org/seq_tools/promoter.html) 来预测基因组上的启动子. 噬菌体基因组中的终止子由 FindTerm (<http://molbiol-tools.ca/Promoters.htm>) 程序来预测.

1.2.6 噬菌体的比较基因组学分析和系统进化树的构建. 登陆 NCBI 网站, 针对 SRT6 的全基因组序列, 用 blastn 进行序列搜索^[11]. 获得和 SRT6 具有较高同源性的噬菌体序列. 直接利用比对结果在 NCBI 上用 Neighbor Joining 方法构建进化树. 利用 NCBI 上 BLAST 模块的 blastn 功能对噬菌体 SRT6 和 C11 进行两两比对. 将比对结果用软件 ACT 进行展示^[12].

2 结果和讨论

2.1 噬菌体 SRT6 的分离

用铜绿假单胞菌的临床菌株作为宿主菌, 分离出一株铜绿假单胞菌噬菌体 (图 1). 我们将其命名为 SRT6. 其在铜绿假单胞菌临床菌株的菌苔上形成澄清的透明噬菌斑. 这说明噬菌体 SRT6 是一株烈性噬菌体, 具备用于噬菌体治疗的潜力.

2.2 噬菌体 SRT6 的全基因组分析

2.2.1 噬菌体 SRT6 的基因组基本特征概述. 全基因组数据显示, 铜绿假单胞菌噬菌体 SRT6 包含一个 91 364 bp 的线性双链 DNA 分子, 它的 G+C 含量为 49.3%. 对其进行全基因组注释后发现, 基因组上包含 182 个蛋白质编码基因, 没有发现 tRNA 和 tmRNA (图 2). 通过 blastp 分析后发现, 其中 96 个基因的编码蛋白与已知功能的蛋白质具有序列同源性 (表 1). 该噬菌体的前 70% 基因的绝大部分是向左

转录的, 后 30% 基因绝大部分是向右转录的. 这个规律非常明显, 在前 133 个基因中仅有 24 个基因是向右转录的; 而在后 49 个基因中仅有 8 个基因是向左转录的, 转录方向和基因的功能并无明显相关性.

据我们所知噬菌体的基因组十分紧凑, 存在大量的基因重叠序列, 相应的非编码序列非常少见. 这些非编码序列可能在噬菌体基因组上起到基因表达调控等作用. 分析发现, 在该噬菌体基因组的 Gene128 和 Gene129 之间存在着一段约 900 bp 的非编码区; 在 Gene75 和 Gene76 之间存在一段大约 800 bp 的非编码区; 在 Gene1 和 Gene2 之间, Gene33 和 Gene34 之间以及 Gene127 和 Gene128 之间存在大约 600 bp 的非编码区; 在 Gene87 和 Gene88 之间存在一段约 500 bp 的非编码区; 在 Gene89 和 Gene90 之间, 以及 Gene132 和 Gene133 之间存在一段约 400 bp 的非编码区; 在 Gene22 和 Gene23 之间, Gene162 和 Gene163 之间以及 Gene164 和 Gene165 之间存在一段约 200 bp 的非编码区. 这些非编码序列可能存在于非常重要的基因外调控位点, 例如启动子、终止子和小 RNA 编码区等.

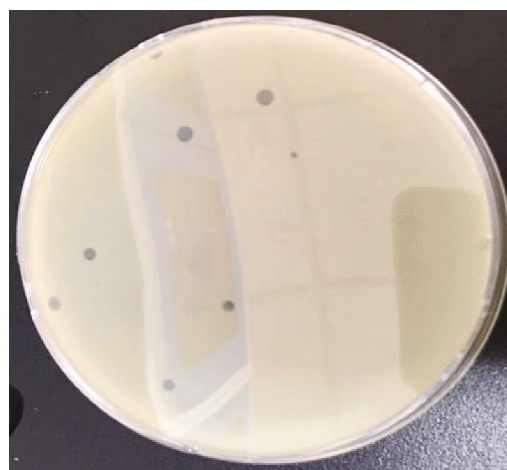


图 1 噬菌体 SRT6 的噬菌斑. 噬菌体 SRT6 在铜绿假单胞菌菌苔上显示澄清的圆状噬菌斑

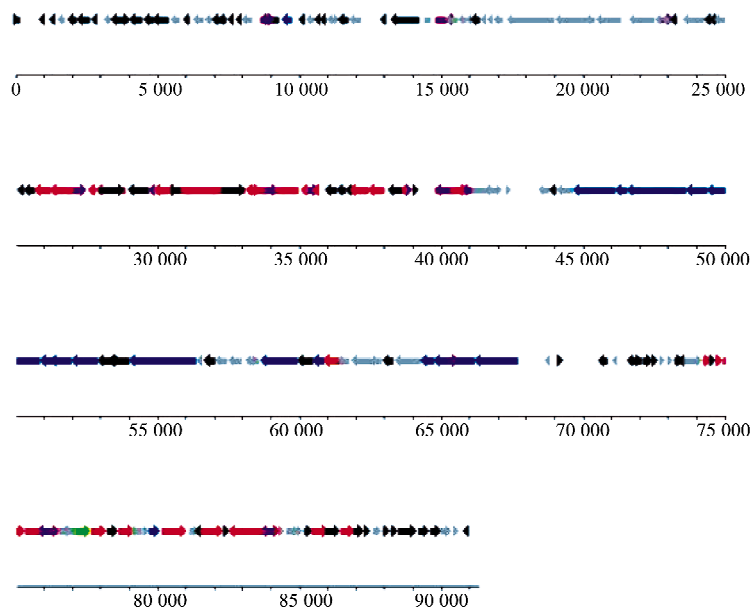


图2 铜绿假单胞杆菌噬菌体的基因组模式图(黑色箭头表示未知功能基因、紫色代表无功能蛋白;蓝色箭头表示编码噬菌体的结构蛋白的基因;红色箭头表示编码 DNA 代谢蛋白的基因;绿色箭头表示编码裂解酶的基因;灰色箭头表示有编码已知功能蛋白但除去上述功能蛋白的基因)

我们进一步通过基因预测网站,对该噬菌体全基因组的启动子进行了预测,共计发现 11 个启动子序列的预测位点,它们分别位于 12 213 bp-12 463 bp、14 142 bp-14 392 bp、16 855 bp-17 105 bp、27 550 bp-27 800 bp、29 611 bp-29 861 bp、29 370 bp-29 620 bp、48 612 bp-48 862 bp、50 960 bp-51 210 bp、51 416 bp-51 166 bp、76 839 bp-77 089 bp 和 80 797 bp-80 547 bp 之间。第四个启动子序列(27 550 bp-27 800 bp)和第五个启动子(29 611 bp-29 861 bp)后面存在数个 DNA 代谢酶相关基因的表达,可能是与 DNA 代谢密切相关的启动子。第七个启动子(48 612bp-48 862bp)和第八个启动子(50 960 bp-51 210 bp)以及第九个启动子(51 416 bp-51 166 bp)都位于结构基因编码区,这些启动子应该和噬菌体的结构基因的表达有关,参与了噬菌体的组装过程;同时,第十个启动子(76 839 bp-77 089 bp)后面存在一个细胞壁水解酶基因,这一启动子应该作为噬菌体基因组上的晚期启动子和噬菌体裂解宿主菌的过程密切相关。

2.2.2 噬菌体 SRT6 的基因组上的蛋白质功能预测. 我们利用 blastp 对噬菌体 SRT6 的所有的编码蛋白进行了功能预测. 结果发现,在 SRT6 基因组上有功能的基因有 96 个(表 1),占全部基因的 52.7%. 这些基因呈现了典型的模块化结构. 其中 14 个基因编码结构蛋白(占有功能基因的 14.6%),它们属于噬菌体结构基因模块;26 个基因编码 DNA 代谢蛋白(占有功能基因的 27.1%),它们属于噬菌体 DNA 代谢模块;1 个基因编码细胞壁水解酶,它属于噬菌体裂解酶模块;剩下的 55 个基因编码其他的功能(占有功能基因的 57.3%). 其中属于噬菌体结构基因模块的是 Gene25、Gene98、Gene99、Gene100、Gene103、Gene104、Gene105、Gene109、Gene116、Gene118、Gene125、Gene126、Gene128、Gene156,它们主要编码噬菌体尾部蛋白、头尾结合蛋白、尾部装配蛋白、尾部纤突蛋白等;属于噬菌体 DNA 代谢模块的是 Gene60、Gene62、Gene64、Gene65、Gene69、Gene70、Gene73、Gene74、Gene76、Gene78、Gene82、Gene83、Gene85、Gene88、Gene119、Gene142、Gene144、Gene145、Gene146、Gene150、Gene152、Gene156、Gene160、Gene163、Gene168、Gene171,它们主要编码的蛋白有 DNA 聚合酶、DNA 解旋酶、RNA 基甲基转移酶、RNA 连接酶、NAD(P)H 氧化还原酶、FAD 氧化还原酶等;属于噬菌体裂解酶模块的是编码细胞壁水解酶的 Gene150;另外的一些有功能的基因则主要编码磷酸异构酶、转录调节因子、丙氨酸-乙二醛转氨酶家族蛋白等各种类家族蛋白以及各种结构域蛋白等。

值得我们关注的是编码细胞壁水解酶的基因 Gene150. 这一基因被预测为可以裂解铜绿假单胞菌的细胞壁,并进一步使得宿主菌死亡、子代噬菌体释放. 我们可以进一步的利用该基因的裂解功能,对其进行有效的改造,将其或其衍生物作为新型的抗菌药物,进而更加有效的针对铜绿假单胞杆菌引发的各种疾病进行治疗。

表 1 噬菌体 SRT6 的基因组成及基因功能

Gene	Strand	Left End	Right End	Score	Function
Gene1		148	171	171	hypothetical protein
Gene2	—	797	997	237	hypothetical protein
Gene3	—	1 181	1 375	228	hypothetical protein
Gene4	—	1 475	1 741	185	GLUG domain protein
Gene5	—	1 835	2 137	202	hypothetical protein
Gene6	—	2 154	2 615	470	hypothetical protein
Gene7	—	2 680	2 823	146	hypothetical protein
Gene8	—	3 096	3 377	259	sugar phosphate isomerase/epimerase
Gene9	—	3 377	3 979	549	hypothetical protein
Gene10	+	3 566	4 072	254	hypothetical protein
Gene11	—	3 990	4 469	261	hypothetical protein
Gene12	—	4 505	4 888	180	hypothetical protein
Gene13	—	4 831	5 412	539	hypothetical protein
Gene14	—	5 455	5 811	304	PAS domain S-box protein
Gene15	—	5 889	6 143	185	hypothetical protein
Gene16	—	6 248	6 640	312	nicotinamide riboside transporter PnuC
Gene17	—	6 637	6 903	186	alanine-glyoxylate aminotransferase family protein
Gene18	—	6 936	7 223	103	hypothetical protein
Gene19	—	7 157	7 453	255	unnamed protein product
Gene20	—	7 469	7 663	240	hypothetical protein
Gene21	—	7 736	7 930	131	hypothetical protein
Gene22	—	7 927	8 352	275	restriction endonuclease
Gene23		8 620	9 141	327	—
Gene24	—	8 623	9 294	544	uncharacterized protein
Gene25	—	9 370	9 753	255	phage tail tape measure protein
Gene26	—	9 990	10 136	169	hypothetical protein
Gene27	—	10 211	10 519	340	penicillin-binding protein
Gene28	—	10 519	10 647	176	hypothetical protein
Gene29	—	10 718	10 951	179	hypothetical protein
Gene30	—	10 941	11 234	208	polynucleotide kinase-phosphatase
Gene31		11 273	11 464	126	KR domain-containing protein
Gene32	—	11 386	11 733	182	hypothetical protein
Gene33	—	11 730	12 206	247	sprT domain-containing protein
Gene34	—	12 835	13 008	160	hypothetical protein
Gene35	—	13 231	14 217	644	hypothetical protein
Gene36	—	14 385	14 636	161	MOSC domain-containing protein
Gene37		14 814	15 593	363	—
Gene38	—	15 093	15 578	509	lipoprotein-releasing ABC transporter permease subunit
Gene39	—	15 590	15 892	181	class II fumarate hydratase
Gene40	—	15 879	16 073	214	rab effector Noc2 isoform X3
Gene41	—	16 054	16 368	252	hypothetical protein
Gene42	—	16 368	16 604	260	CarboxypepD_reg domain-containing protein

续表 1 噬菌体 SRT6 的基因组及基因功能

Gene	Strand	Left End	Right End	Score	Function
Gene43	—	16 604	16 837	211	WD40 repeat-like protein
Gene44	—	16 837	17 178	150	Uncharacterised protein
Gene45	—	17 282	19 027	1 152	putative ribonucleoside-diphosphate reductase alphachain
Gene46	—	19 020	20 168	949	ribonucleotide-diphosphate reductase beta subunit
Gene47	—	20 083	20 427	268	gp071
Gene48	—	20 430	21 377	572	thymidylate synthase
Gene49	—	21 584	22 534	734	3'-phosphatase, 5'-polynucleotide kinase
Gene50	—	22 527	22 739	230	Glycerol-3-phosphate dehydrogenase
Gene51		22 740	23 210	256	—
Gene52	—	22 743	23 078	374	alpha/beta hydrolase
Gene53	—	23 097	23 306	271	hypothetical protein
Gene54	—	23 497	24 291	574	constituent protein
Gene55	—	24 273	24 512	184	hypothetical protein
Gene56	—	24 457	24 624	207	hypothetical protein
Gene57	—	24 621	25 058	339	PAS domain S-box protein
Gene58	—	25 055	25 297	195	hypothetical protein
Gene59	—	25 282	25 680	305	hypothetical protein
Gene60	—	25 677	26 240	347	HNH endonuclease
Gene61	+	25 983	26 393	171	hypothetical protein
Gene62	—	26 237	27 289	862	exonuclease
Gene63	+	27 017	27 502	268	—
Gene64	—	27 563	27 796	232	lysylphosphatidylglycerol synthetase family protein
Gene65	—	27 865	28 869	1 020	DNA primase/helicase
Gene66		27 946	28 887	704	hypothetical protein
Gene67	—	28 972	29 676	529	hypothetical protein
Gene68	—	29 678	29 887	171	—
Gene69	—	29 868	30 374	238	NAD(P)II-dependent oxidoreductase
Gene70	—	30 356	33 070	2 040	putative DNA Polymerase
Gene71		30 803	31 726	330	hypothetical protein
Gene72	+	32 225	33 103	463	hypothetical protein
Gene73	—	33 131	34 993	1 369	DNA primase/helicase
Gene74	+	33 239	33 655	301	sodium/glutamate symporter
Gene75	+	33 707	34 279	133	—
Gene76	—	35 047	35 232	169	UPF0600 protein C5orf51 homolog
Gene77	+	35 253	35 702	253	—
Gene78	—	35 484	35 675	274	putative RNA methyltransferase
Gene79	—	35 908	36 330	247	hypothetical protein
Gene80	—	36 331	36 633	306	hypothetical protein
Gene81	—	36 636	36 800	214	hypothetical protein
Gene82	—	36 787	37 443	527	prevent-host-death protein
Gene83	—	37 440	37 988	247	HAD-like domain protein
Gene84	—	38 119	38 631	328	unnamed protein product

续表 1 噬菌体 SRT6 的基因组组成及基因功能

Gene	Strand	Left End	Right End	Score	Function
Gene85	—	38 607	38 804	242	preprotein translocase subunit SecD
Gene86		38 869	38 988	101	—
Gene87	+	39 112	39 255	131	hypothetical protein
Gene88	+	39 781	40 926	1 028	RNA ligase
Gene89	—	39 805	40 362	323	—
Gene90	—	40 717	41 151	168	—
Gene91		41 024	41 899	709	putative phage protein
Gene92	—	41 516	41 938	231	gp431
Gene93		41 847	42 221	127	PAS/PAC sensor protein
Gene94	+	42 232	42 540	294	DUF768 domain-containing protein
Gene95	—	43 381	43 818	176	vimentin
Gene96	—	43 805	44 044	228	hypothetical protein
Gene97	—	44 062	44 661	298	DUF3380 domain-containing protein
Gene98	—	44 639	46 138	866	putative tail fiber protein
Gene99	—	46 152	46 535	281	phage tail protein
Gene100	—	46 570	48 627	1 274	putative tail fiber protein
Gene101	—	48 638	49 369	534	DUF2612 domain-containing protein
Gene102	—	49 388	50 851	1 050	putative baseplate related protein
Gene103	—	50 853	51 431	104	putative tail lysozyme
Gene104	—	51 235	51 975	340	baseplate protein
Gene105	—	51 972	52 889	478	putative structural protein
Gene106	—	52 886	53 242	302	Uncharacterised protein
Gene107		53 218	53 709	157	unnamed protein product
Gene108	—	53 248	54 009	541	hypothetical protein
Gene109	—	54 006	56 372	1 650	putative tape measure protein
Gene110	—	56 369	56 521	167	conserved hypothetical protein
Gene111	—	56 629	57 000	336	hypothetical protein
Gene112	—	57 014	57 493	255	DUF3277 domain-containing protein
Gene113	—	57 493	57 993	287	DUF3277 domain-containing protein
Gene114	—	58 068	58 592	425	DUF3277 domain-containing protein
Gene115		58 158	58 577	253	YbhB/YbcL family Raf kinase inhibitor-like protein
Gene116	—	58 623	59 909	1 032	putative tail sheath protein
Gene117	—	59 922	60 485	284	hypothetical protein
Gene118	—	60 482	60 862	201	putative head-tail joining protein
Gene119	—	60 862	61 404	283	putative RNA polymerase
Gene120	—	61 313	61 789	375	gp023
Gene121	—	61 840	62 874	1 024	major capsid protein
Gene122		62 182	62 853	564	efflux RND transporter periplasmic adaptor subunit
Gene123	—	62 918	63 328	473	hypothetical protein
Gene124	—	63 356	64 273	802	73 gene product
Gene125	—	64 270	64 740	429	putative methyltransferase
Gene126	—	64 750	66 189	973	putative portal protein

续表 1 噬菌体 SRT6 的基因组及基因功能

Gene	Strand	Left End	Right End	Score	Function
Gene127		65 167	6 5637	194	—
Gene128	—	66 202	67 722	1 053	putative terminase large subunit
Gene129	—	68 643	68 789	131	conserved hypothetical protein
Gene130	+	69 185	69 316	118	hypothetical protein
Gene131	—	70 524	70 847	212	hypothetical protein
Gene132	—	70 990	71 178	109	JNK
Gene133	—	71 560	72 054	284	unnamed protein product
Gene134		71 653	72 120	411	hypothetical protein
Gene135		72 117	72 473	396	hypothetical protein
Gene136	+	72 433	72 642	145	hypothetical protein
Gene137	+	72 677	72 955	233	etratricopeptide repeat protein
Gene138	+	72 952	73 230	264	ABC transporter substrate-binding protein
Gene139	+	73 211	73 558	245	hypothetical protein
Gene140	—	73 334	73 831	141	hypothetical protein
Gene141		73 560	74 219	235	AP2 domain protein
Gene142		74 228	74 497	285	FAD-dependent oxidoreductase
Gene143	+	74 494	74 706	219	hypothetical protein
Gene144	+	74 716	74 943	249	DNA-binding protein WhiA
Gene145	+	74 946	75 362	307	putative CMP/dCMP deaminase, zinc-binding
Gene146	+	75 359	76 567	1 105	putative DNA ligase
Gene147	—	75 746	76 471	441	—
Gene148	—	76 517	77 077	195	nodal modulator 1
Gene149		76 580	77 041	381	DUF4081 domain-containing protein
Gene150		77 106	77 666	305	cell wall hydrolase
Gene151	+	77 668	78 222	457	metal dependent phosphohydrolase HD region
Gene152	+	78 212	78 643	427	hypothetical protein
Gene153		78 643	79 197	443	putative phosphoesterase
Gene154	—	79 006	79 473	274	probable ubiquitin-like-specific protease 2A
Gene155		79 448	79 723	259	transcriptional regulator
Gene156	+	79 720	80 127	187	Phage protein
Gene157		80 139	81 056	684	putative RNA ligase/tail attachment protein
Gene158	+	81 067	81 483	298	Phi92_gp095
Gene159	—	81 310	81 906	303	hypothetical protein
Gene160	+	81 493	82 359	651	phosphoribosylpyrophosphate synthetase
Gene161	+	82 356	82 574	201	hypothetical protein
Gene162	—	82 532	83 368	252	RNA recognition motif
Gene163		82 631	84 319	1 448	putative nicotinate phosphoribosyltransferase
Gene164	—	83 636	84 163	352	—
Gene165	+	84 331	84 528	155	prolactin-releasing peptide receptor-like
Gene166	+	84 525	85 004	440	macro domain protein
Gene167	+	84 890	85 165	136	DUF1776 domain-containing protein
Gene168	+	85 167	85 484	156	hypothetical protein

续表 1 噬菌体 SRT6 的基因组成及基因功能

Gene	Strand	Left End	Right End	Score	Function
Gene169		85 442	86 062	207	putative DNA recombination-mediator protein A
Gene170	+	85 887	86 459	241	hypothetical protein
Gene171	+	86 446	86 949	252	helix-turn-helix domain-containing protein
Gene172	+	86 918	87 292	344	hypothetical protein
Gene173	+	87 289	87 504	227	hypothetical protein
Gene174		87 561	87 923	257	wall-associated receptor kinase-like 20-like protein
Gene175		87 923	88 234	247	unnamed protein product
Gene176		88 222	88 500	110	hypothetical protein
Gene177	+	88 503	89 165	465	hypothetical protein
Gene178	+	89 208	89 612	235	hypothetical protein
Gene179	+	89 609	90 025	392	hypothetical protein
Gene180	+	89 994	90 419	314	DUF4238 domain-containing protein
Gene181		90 419	90 685	208	PAP2 family protein
Gene182	-	90 748	90 897	189	hypothetical protein

注:“-”表示此蛋白在数据库中搜索不到同源序列。

2.3 噬菌体 SRT6 的比较基因组学分析

表 2 与铜绿假单胞菌噬菌体 SRT6 基因组有较高相似的噬菌体列表

噬菌体	覆盖度(Coverage)/%	一致性(Identity)/%	发表文献
<i>Pseudomonas</i> phage C11	99	97	[13]
<i>Pseudomonas</i> phage Zigelbrucke	96	96	N
<i>Pseudomonas</i> phage JG004	97	95	[14]
<i>Pseudomonas</i> phage PAK_P2	94	94	[14]
<i>Pseudomonas</i> phage PaoP5	98	93	[14]
<i>Pseudomonas</i> phage phiMK	97	95	N
<i>Pseudomonas</i> phage PAK_P4	94	95	[14]
<i>Pseudomonas</i> phage vB_PaeM_C2-10_Ab15	95	94	[14]
<i>Pseudomonas</i> phage vB_PaeM_C2-10_Ab10	94	94	[14]
<i>Pseudomonas</i> phage vB_PaeM_C2-10_Ab1	94	94	[14]
<i>Pseudomonas</i> phage vB_PaeM_C2-10_Ab08	95	94	[14]
<i>Pseudomonas</i> phage vB_PaeM_C2-10_Ab02	95	94	[14]
<i>Pseudomonas</i> phage K5	95	98	[15]
<i>Pseudomonas</i> phage K8	95	98	[16]
<i>Pseudomonas</i> phage PaP1	94	98	[14]
<i>Pseudomonas</i> phage PA10	95	98	N
<i>Pseudomonas</i> phage YS35	94	98	[17]
<i>Pseudomonas</i> phage PAK_P1	94	93	[14]
<i>Pseudomonas</i> phage vB_PaeM_MAG1	94	93	[18]

注:N表示无相应参考文献。

我们利用 SRT6 的全基因组序列,在 NCBI 上用 blastn 进行了序列搜索,发现和 SRT6 具有序列相似性的噬菌体基因组有 19 个,它们都属于假单胞菌噬菌体(*Pseudomonas* phage)。它们的名称以及和 SRT6 的相似性和覆盖度都在表 2 中列出。直接利用比对结果在 NCBI 上用 Neighbor Joining 方法构建了进化树(图 3),结果显示和 SRT6 基因组最为相似的噬菌体是 *Pseudomonas* phage C11,它们位于同一个分枝上。进一步分析结果显示,噬菌体 Pap1、K5 和 K8 位于同一分枝;噬菌体 YS35、vB_PaeM_C2-10_Ab02、

vB_PaeM_C2-10_Ab15、vB_PaeM_C2-10_Ab10、vB_PaeM_C2-10_Ab1 和 vB_PaeM_C2-10_Ab08 位于同一分枝;噬菌体 vB_PaeM_MAG1、Zigelbrucke、JG004、PAK_P2、PAK_P4 和 PA10 位于同一分枝;而噬菌体 PAK_P1、PaoP5 和 phiMK 则位于进化树的远端。我们进一步利用 NCBI 上的 blastn 对噬菌体 SRT6 和噬菌体 C11 进行了全基因组的两两比对分析(图 4)。为了更清楚地展示,我们利用 C11 的反向互补序列进行的全基因组的比对分析。结果发现在基因组水平上,SRT6 和 C11 主要有四处基因组片段的替换、插入或缺失(图 4)。其中在 SRT6 基因组的 783-1010 bp 处和 C11 基因组的 1 608-2 155 bp 处有着序列替换现象;SRT6 在 C11 基因组的 71 453-71 553 bp 处有一个小片段的缺失;SRT6 在 C11 的 73 804-74 901 bp 处有一个大片段的缺失;SRT6 基因组在 73 545-74 202 bp 处有一个大片段的插入。这些片段的替换、插入和缺失说明铜绿假单胞菌噬菌体 SRT6 是不同于 C11 的全新噬菌体。

进一步分析这四处基因组片段的不同后发现:第一个片段的插入影响了 SRT6 基因组上的基因编码,C11 的 gene 171 和 gene 170 被替换为 SRT6 的 gene 2;第二个片段的缺失并没有影响 SRT6 基因组上的基因编码;第三个片段的缺失使得 SRT6 基因组上缺失了 C11 的 gene 44、gene 43 和 gene 42;第四个片段的插入使得 SRT6 基因组上存在了 C11 中所没有的 SRT6 的 gene 140 和 gene 141。

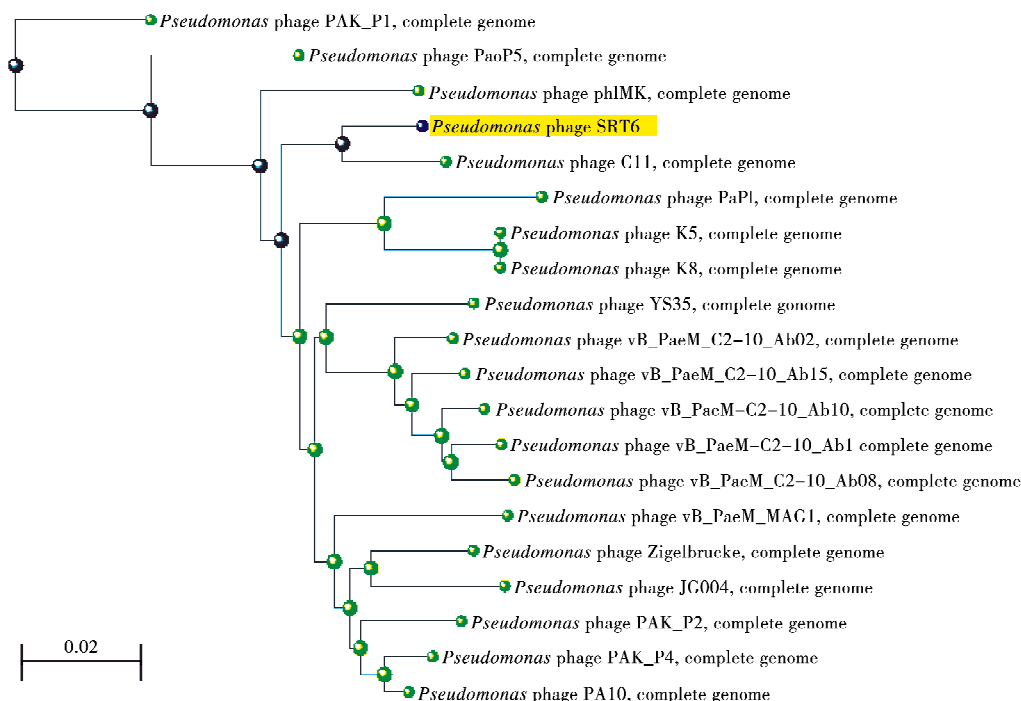


图 3 与铜绿假单胞菌噬菌体 SRT6 基因组有较高相似的噬菌进化树

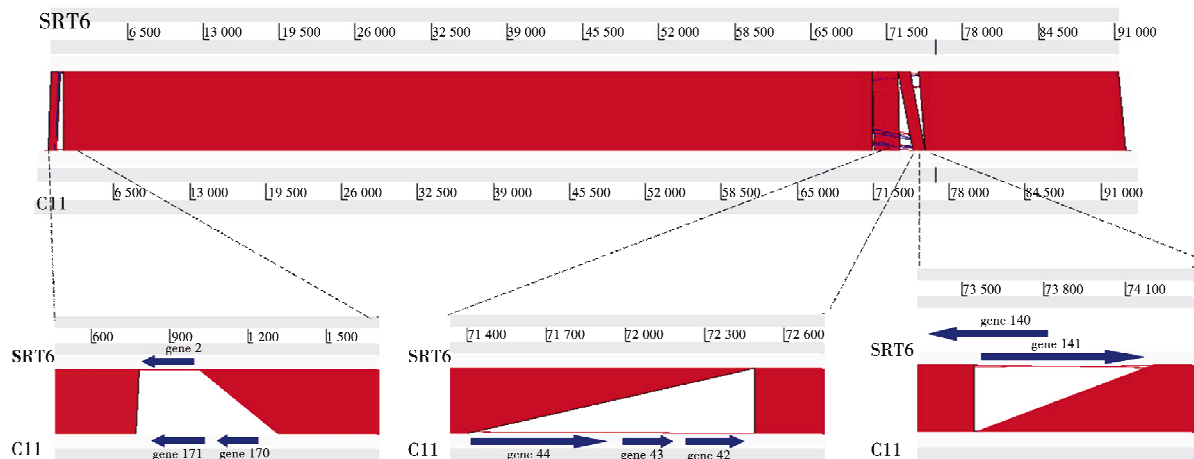


图 4 铜绿假单胞菌噬菌体 SRT6 和 C11 的全基因组两两比对分析
其中红色代表基因组之间有序列相似性;蓝色箭头代表基因

3 讨论与结论

在本研究中,我们以铜绿假单胞菌临床菌株作为宿主菌,从取自山东济南大学甲子湖畔的土壤中分离得到了一株特有的铜绿假单胞菌噬菌体,并将其命名为 SRT6。提取其基因组并测序之后,通过比较基因组学和生物信息学分析得知,该噬菌体的基因组大小为 91 364 bp,GC 含量为 49.3%,包含 182 个蛋白质编码基因,其中 96 个基因与已知功能的基因具有同源性。同时还预测到了 11 个启动子序列位点。我们又将 96 个同源性基因细分为结构基因模块、DNA 代谢模块、裂解酶模块等不同的模块。其中 Gene150 作为一种细胞壁水解酶基因,可能在裂解细菌过程中起到重要的作用,可以对其进一步的改造用作新型抗菌药物。最后,通过比较基因组学分析和系统进化树的构建,找到了铜绿假单胞菌噬菌体 SRT6 的最相似菌株 C11。对它们进行基因组的两两比较后发现,SRT6 是一株新的噬菌体。

噬菌体治疗中,噬菌体会利用宿主资源进行繁殖将子代病毒传播到细菌群体中导致宿主损害或死亡,但不同的研究还表明细菌与噬菌体相互作用过程中细菌发展出多种抗噬菌体感染的能力,目前细菌主要采用被动免疫和主动防御来抵抗噬菌体感染。细菌可利用其细胞外结构如生物膜抵御恶劣环境及抵抗噬菌体侵染,且细菌可利用自身限制修饰系统、流产感染系统及适应性免疫等来对噬菌体感染进行防御^[19]。除此外,噬菌体治疗目前也存在一些局限性:噬菌体宿主特异性强,治疗的宿主谱相对窄,临床应用需对致病菌做出快速诊断方能应用相应的噬菌体治疗;噬菌体作为异源性物质,在体内存活时间短,可刺激机体的免疫系统产生免疫应答,其产生的抗体可能会和宿主菌结合抑制其抗菌作用;噬菌体可能携带有某些毒素基因;噬菌体治疗的最佳时间和剂量不易掌握等。面对细菌的防御抵抗,噬菌体也进化出一系列对应变化如噬菌体可通过产生多糖裂解酶等来裂解胞外多糖进而消除生物膜的阻碍作用,噬菌体还可通过多种方式来逃避限制修饰系统及其他防御系统如靶序列突变、抑制限制性内切酶等。此外,研究通过不同的噬菌体开发相应的溶壁酶抗菌药物来进行治疗;噬菌体的联合应用或与抗生素联合应用、筛选培育能在体内长期循环的噬菌体、通过人工自然选择筛选广谱噬菌体、利用分子生物学技术对噬菌体进行改造等方法来解决噬菌体治疗目前存在的问题,有研究表明美国已研制出“系列病原体培育法”可以分离长期循环的噬菌体^[20]。

噬菌体作为自然界中丰度最高的生物类群,具有物种丰富、天然有效、绿色安全、研发成本低等优势,可以被用作抗生素的替代品,具有广阔的研究前景。我国在噬菌体领域的研究目前还处于起步阶段,虽然已经取得了一些成就,但在噬菌体临床应用方面跟美国、东欧等国家相比还有很大的差距。这株新的铜绿假单胞菌噬菌体 SRT6 作为一株具有自主知识产权的烈性噬菌体,本课题组将要对其进行动物学实验和临床实验,为进一步的噬菌体治疗绿脓杆菌感染打下研究基础。

参 考 文 献

- [1] 吴伟清,李国明.铜绿假单胞菌耐药机制的研究进展[J].医学综述,2012,18(22):3812-3815.
- [2] Bonomo R A, Szabo D. Mechanisms of multidrug resistance in *Acinetobacter* species and *Pseudomonas aeruginosa*[J]. *Clinical Infectious Diseases An Official Publication of the Infectious Diseases Society of America*, 2006, 43 (2): S49-S56.
- [3] 王海霞,黄勇.烟台海滨自由生活线虫一新记录种(英文)[J].聊城大学学报(自然科学版),2015,28(3):38-41.
- [4] 高群,黄勇.东海自由生活线虫三个新记录(英文)[J].聊城大学学报(自然科学版),2015,28(3):42-50.
- [5] Fan X, Teng T, Wang H, et al. Biology of a novel mycobacteriophage, SWU1, isolated from Chinese soil as revealed by genomic characteristics[J]. *Journal of Virology*, 2012, 86(18): 10230-10231.
- [6] 邢少贞,张湘莉,裴广信,等.粪肠球菌噬菌体 vB_EfaP_IME195 的生物学特性及其全基因组分析[J].微生物学通报,2017,44(2):305-313.
- [7] 滕国栋,陈敏亮.全基因组测序技术的发展和应[J].中国美容医学,2013,22(4):503-506.
- [8] 樊祥宇.分枝杆菌噬菌体的系统生物学研究[D].重庆:西南大学,2014.
- [9] 宁淑香,陆敏.VP10——一株海洋噬菌体的分离及特性研究[J].水产科学,2000,19(2):14-16.
- [10] Delcher AL, Bratke KA, Powers EC, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer[J]. *Bioinformatics*, 2007, 23(6):673-679.

- [11] 孙啸, 陆祖宏, 谢建明. 生物信息学基础[M]. 北京: 清华大学出版社, 2006.
- [12] Carver T J, Rutherford K M, Berriman M, et al. ACT: the Artemis Comparison Tool[J]. *Bioinformatics*, 2005, 21(16): 3422-3423.
- [13] Cui X, You J, Sun L, et al. Characterization of *Pseudomonas aeruginosa* phage C11 and identification of host genes required for virion maturation[J]. *Sci Rep*, 2016, 6: 39130.
- [14] Shen M, Le S, Jin X, et al. Characterization and comparative genomic analyses of *Pseudomonas aeruginosa* phage paoP5: new members assigned to PAK_P1-like viruses[J]. *Sci Rep*, 2016, 6: 34067.
- [15] Li L, Pan X, Cui X, et al. Characterization of *Pseudomonas aeruginosa* phage K5 genome and identification of its receptor related genes[J]. *J Basic Microbiol*, 2016, 56(12): 1344-1353.
- [16] Pan X, Cui X, Zhang F, et al. Genetic evidence for o-specific antigen as receptor of *Pseudomonas aeruginosa* phage K8 and its genomic analysis[J]. *Frontiers in Microbiology*, 2016, 7(e744): 252.
- [17] Yu S, Huang H, Hao Y, et al. Complete genome sequence of the myoviral bacteriophage YS35, which causes the lysis of a multi-drug-resistant *Pseudomonas aeruginosa* strain[J]. *Genome Announc*, 2018, 6(6): e01395-17.
- [18] Kwiatek M, Parasion S, Rutyna P, et al. Isolation of bacteriophages and their application to control *Pseudomonas aeruginosa* in planktonic and biofilm models[J]. *Research in Microbiology*, 2016, 168(3): 194-207.
- [19] 蒋羽, 李伟, 范蒙光. 细菌抗噬菌体感染分子机制[J]. *疾病监测*, 2015, 30(8): 679-682.
- [20] 钱震雯, 岳启安, 田凤丽. 噬菌体治疗的研究概况[J]. *医学综述*, 2007, 13(16): 1256-1258.

Isolation and Genome Sequence Analysis of a Novel *Pseudomonas Aeruginosa* Phage SRT6

LIU Zi-chen LI Qi¹ ZHANG Fu-kang ZHANG Chun-long MA Hong-rui¹

YI Ting-xu GAO Xiao-meng FAN Xiang-yu

(School of Biological Science and Technology, University of Jinan, Jinan 250022, China)

Abstract We want to isolate a novel *Pseudomonas aeruginosa* phage, study its biology characteristics and carry on genome sequencing and comparative genomics analysis so as to understand the evolutionary process of *Pseudomonas* phage and potential application of the *Pseudomonas* phage. Double-layer plate method was carried on to isolate and purify *Pseudomonas* phage. Phenol chloroform was used to extract phage genome. The Illumina Hiseq genome sequencing platform was carried on to sequence. Some bioinformatics methods were used to analyze the phage genome and comparative genomics study. A novel virulent phage SRT6 which can infect *Pseudomonas aeruginosa* was isolated. Its genome was 91364 bp dsDNA and had a G+C content of 49.3%. There were 182 protein-coding genes and no tRNA and tmRNA coding genes. 96 proteins of them were homologous with known function proteins. Comparative genomic analysis revealed that *Pseudomonas* phage SRT6 was a new species. A novel virulent *Pseudomonas* phage SRT6 was isolated and it was a new species. The study lays a foundation for the future development of phage therapy and treatment of drug resistant *Pseudomonas aeruginosa*.

Key words *Pseudomonas aeruginosa*; *Pseudomonas* phage; genome sequencing; comparative genomics analysis