

DOI: 10.3724/SP.J.1118.2016.16037

嗜水气单胞菌 AH10 (CCTCC AB2014155)的全基因组测序及比较分析

张国亮¹, 王浩¹, 张也¹, 许丹¹, 吕利群^{1,2}

1. 国家水生动物病原库, 农业部淡水水产种质资源利用重点实验室, 上海海洋大学 水产与生命学院, 上海 201306;
2. 淡水水产健康养殖湖北省协同创新中心 湖北 武汉 430070

摘要: 为了解嗜水气单胞菌全基因组信息的结构和功能, 对嗜水气单胞菌(*Aeromonas hydrophila*) AH10(CCTCC 保藏号: AB2014155)进行了全基因组测序。将提取的 AH10 全基因组 DNA 进行片段化、构建文库后上机测序; 将基因组组装后进行注释和比较分析。结果显示: AH10 全基因组序列总长度为 4.91 Mb, GC 含量为 61.1%, 共预测到 4570 个基因, 3351 个基因具有明确的生物学功能。共找到 2592 个基因具有直系同源族(cluster of orthologous groups, COG)分类, 1281 个与代谢通路有关的基因, 基因组岛 31 个。比较基因组学研究发现: AH10 比对到 SNP 位点 102481 个, 与 *Aeromonas hydrophila* YL17 基因结构共线性最好, 与美国来源的嗜水气单胞菌标准菌株 *Aeromonas hydrophila* ATCC 7966 具有较近的亲缘关系, AH10 的致病性与溶血素基因 *hly* (hemolysin gene)、气溶素基因 *aer* (aerolysin gene)、黏附素基因 *ahal* (major adhesion gene)、丝氨酸蛋白酶基因 *ahp* (serine protease gene)、细胞兴奋性肠毒素基因 *alt* (cytotoxic enterotoxin gene)密切相关。这些基因在嗜水气单胞菌中同源性较高。存在自体诱导物信号分子-2 (autoinducer-2)与合成信号分子酶基因(signal molecule synthetase gene, *ahyI*)和转录调节蛋白基因(transcriptional regular gene, *AhyR*)群体感应(quorum sensing, QS)系统, 不存在 III 型分泌系统(type III secretion system, TTSS)。通过对 AH10 全基因组序列进行较为全面的注释和比较统计分析, 为嗜水气单胞菌的分子生物学研究提供理论支持。

关键词: 嗜水气单胞菌; 基因组测序; 基因; 比较分析

中图分类号: S941

文献标志码: A

文章编号: 1005-8737-(2016)05-0995-11

嗜水气单胞菌(*Aeromonas hydrophila*)广泛存在于水体环境中, 是淡水水产养殖中主要致病菌。该菌是嗜温性条件致病菌, 可引起青鱼、草鱼、鲢、鳙、鲫等鱼类患细菌性败血症、赤皮病、烂鳃病等。嗜水气单胞菌致病范围极广, 遍及世界各地^[1-2], 发病死亡率高^[3]。多年来, 不断有因嗜水气单胞菌暴发而带来重大经济损失的报道^[4]。嗜水气单胞菌同时也是典型的人-畜-鱼共患性病菌^[5], 引起各界关注, 被列为重点防治对象。2012 年徐丽娟^[6]在上海崇明某养殖场内, 从患典型细

菌性败血症草鱼体内分离到一株嗜水气单胞菌, 定名为 AH10(本实验菌株), 并对 AH10 做了初步研究。半数致死量(median lethal dose, LD50)结果显示, AH10 对健康草鱼的 LD50 为 1.18×10^6 CFU/mL; 回感实验表明, AH10 可导致草鱼和鲫鱼的严重败血症, 致病性较强^[6]。

随着二代测序技术的广泛应用, 三代测序技术的崭露头角^[7], 测序成本大幅度下降, 为从基因组学水平来分析嗜水气单胞菌提供了可能。通过基因组学研究, 不仅可以分析嗜水气单胞菌的

收稿日期: 2016-01-27; 修订日期: 2016-03-23.

基金项目: 现代农业产业技术体系建设专项资金项目(CARS-46-12).

作者简介: 张国亮(1990-), 男, 硕士研究生, 研究方向为水产病原学. E-mail: 314865661@qq.com

通信作者: 吕利群, 教授. Tel: 021-61900454; E-mail: lqlv@shou.edu.cn

基因结构、基因功能,还可以了解基因与基因之间的相互作用。基因组学研究潜力巨大,除进行基因水平研究外,还能进行 RNA 组学、蛋白质组学等相关研究^[8]。嗜水气单胞菌的基因组测序在过去的十几年中有很大的进展,迄今为止已有 38 株嗜水气单胞菌菌株的基因序列上传至美国国立信息技术生物中心(National Center for Biotechnology Information, NCBI),在这 38 株嗜水气单胞菌中仅有 10 株完成了全基因组测序(包括本研究中的 AH10, 序列登录号为: CP011100.1)。众所周知,供参考的基因组序列越多,用于基因和蛋白质分析的数据就越多,提供的结果就越可靠。在已经完成全基因组测序的 10 株嗜水气单胞菌中,有很多基因还未进行注释,其结构和功能还未知,需要进一步去探索。

本研究对嗜水气单胞菌 AH10 中国株进行了全基因组测序,并进行了基因组分析、基因功能注释和比较基因组学分析,为嗜水气单胞菌的基因结构与功能、核酸以及蛋白质的研究奠定了一定的基础,也为防治嗜水气单胞菌引起的疾病提供了理论依据。

1 材料与方法

1.1 菌株及其鉴定

嗜水气单胞菌 AH10 由本实验室从上海崇明某养殖场内濒死草鱼体内分离^[5],保存于武汉大学菌种保藏中心(CCTCC 保藏号: AB2014155)。选用北京陆桥科技有限公司的 Muller-Hinton(MH)培养基进行培养(每 1000 mL 含水解酪蛋白 17.5 g,牛肉粉 5 g,可溶性淀粉 0.15 g,琼脂 15~20 g, pH 为 7.2)。不加琼脂即为 MH 肉汤培养基。16S rDNA 基因序列鉴定:用 16S rDNA 的 PCR 扩增通用引物(F: 5'-AGAGTTTGTATCATGGCTCAG-3'; R: 5'-CTACGGTTACCTTGTACGAC-3',由徐丽娟^[6]设计)对其 16S rDNA 序列进行扩增测序并进行进化树分析。

1.2 嗜水气单胞菌 AH10 的全基因组测序

将提取的 AH10 基因组 DNA 利用荧光标记原理,在上海瀚宇生物科技有限公司完成测序。取 5 μg 嗜

水气单胞菌 AH10 的 DNA,利用 Covaris® g-TUBE® (Covaris, 美国)将 DNA 打断至 10 kb。片段化后的 DNA 用 PacBio® SMRTbell™ Template Prep Kit (PacBio, 美国)构建文库,得到供测序的 SMRTbell templates,在 PacBio RS II 平台上进行测序反应。测序完成后经 SMRT® Portal 软件拼接组装,验证后得到 AH10 全基因组序列。

1.3 ORF 预测

将得到的基因组序列通过 Glimmer3.0 软件(Delcher 等^[9]方法)进行开放阅读框(open reading frame, ORF)预测,由于小于 30 个氨基酸的编码序列(coding sequence, CDS)没有生物学意义,所以首先除去小于编码 30 个氨基酸的 ORF。将得到的 ORF 整理,在 NCBI 上验证后,统计出全基因组的 ORF。

1.4 基因功能注释分析

将得到的 ORF 翻译成氨基酸后与 NCBI、京都基因和基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)数据库、保守结构域数据库(Conserved Domain Database, CDD)通过序列比对工具(basic local alignment search tool, BLAST)进行比对,逐个检索分析,比对到的蛋白质功能赋予其相对应的 ORF。没有比对到功能的 ORF,赋予其假设蛋白(hypothetical protein)。根据比对结果得到京都基因和基因组百科全书直系同源编号(Kyoto Encyclopedia of Genes and Genomes orthologynumber, KO number),根据此 KO number 获得预测蛋白参与的代谢通路信息。将所有 ORF 根据 CDD 比对结果进行直系同源族(cluster of orthologous groups, COG)分类。将数据库比对结果进行统计分析,预测其存在的基因组岛,并分析 AH10 致病性。

1.5 比较基因组学分析

根据 *Aeromonas hydrophila* ATCC 7966 (ATCC7966)、*Aeromonas hydrophila* ML09-119 (ML09-119)、*Aeromonas hydrophila* 4AK4(4AK4)、*Aeromonas hydrophila* YL17(YL17)(来自 NCBI, 登录号分别为: CP000462.1、CP005966.1、CP006579.1、CP007518.1)的序列和全基因组信息进行核心基因

统计分析, 寻找其同源物种。用 Mummer_3.23 软件进行单核苷酸多态性(single nucleotide polymorphism, SNP)、插入缺失标记(insertion-deletion, INDEL)和基因结构共线性分析。通过 Mega6.06 软件构建进化树。

2 结果与分析

2.1 AH10 菌株鉴定

AH10 在扫描电镜下观察呈短杆状, 两端钝圆, 直形, 菌体大小为(0.3~0.5 μm)×(0.8~1.6 μm); 应用全自动 ATB 细菌鉴定仪 ID32 E 细菌测定试剂条对 32 种菌株进行生理生化鉴定, 结果显示, 菌株 AH10 为嗜水气单胞菌, ID 值为 95%, T 值为 0.68; 利用通用引物扩增 AH10 的 16S rDNA 序列, 并与 GenBank 中相似序列进行同源性分析, 对比结果表明, 菌株 AH10 为嗜水气单胞菌(*Aeromonas hydrophila*, Ah), GeneBank 登录号为 CP011100.1。

2.2 AH10 基因组拼装

嗜水气单胞菌 AH10 全基因组测序共得到有效反应 2303270 个, 其中最短的测序片段(read)为 45 bp, 最长的 read 为 1013 bp, reads 平均长度为 261 bp。对原始的 reads 进行质量评估和序列组装, 共得到 25 个重叠群(contigs), 形成 25 个基因组骨架(scaffolds) (表 1), 基因组覆盖度为 98.5%。

表 1 嗜水气单胞菌 AH10 基因组拼装的数据统计
Tab. 1 Data statistics of *Aeromonas hydrophila* AH10 genome assembly

属性 property	重叠群 contig	基因组骨架 scaffold
总序列数量 total sequence number	25	25
序列长度大于 1 kb 以上的序列数量 number of sequences greater than 1 kb	25	25
GC 含量/% GC content	61.1	61.1
最短序列长度/bp shortest length	1713	1713
最长序列长度/bp longest length	917840	917840

2.3 基因的预测

AH10 序列总长度为 4.91 Mb, GC 含量为 61.1%, 共预测到 4570 个基因, 其中长度 0.09~0.5 kb 的基因有 1319 个; 长度 0.5~1.0 kb 的基因有 1616 个; 长度 1.0~1.5 kb 的基因有 957 个; 长度 1.5~2.0 kb

的基因有 395 个; 长度 2.0~2.5 kb 的基因有 156 个; 长度 2.5~3.0 kb 的基因有 60 个; 长度>3.0 kb 的基因有 67 个(图 1)。4327 个基因可以找到同源序列, 3351 个基因具有明确的基因功能, 共发现 128 个 tRNA 序列, 转运 22 种氨基酸。33 个 rRNA 序列, 其中 5S rRNA、23S rRNA 和 16S rRNA 各 11 个。

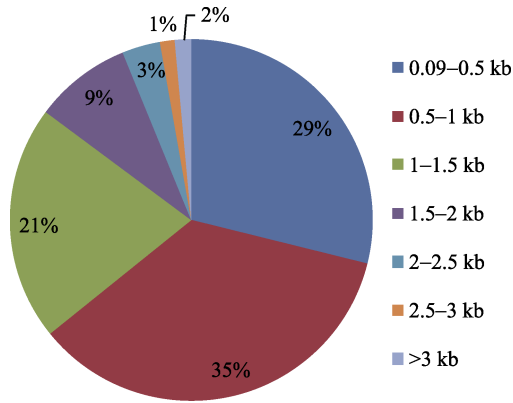


图 1 嗜水气单胞菌 AH10 基因片段长度统计
Fig. 1 Statistic result of the gene fragments of *Aeromonas hydrophila* AH10

2.4 基因功能分类

通过与数据库比对分析, 共有 2592 个基因具有 COG 功能分类, 分为 22 类(图 2)。这些基因大多与细菌的代谢与自身构成有关。其中参与自身构成的基因有 208 个, 占比 8%; 与代谢有关的基因有 676 个, 占比 26%。在功能已知的基因中, 氨基酸代谢与运输的最多, 为 261 个, 占比 10.1%; 与 RNA 修饰和染色体结构的相关的基因最少, 各发现 1 个。还有 270 个基因功能未知, 占比最高, 为 10.4%。

2.5 代谢通路构建

根据 KEGG 数据库比对结果, 找到代谢通路基因 1281 个, 分为 20 类(图 3)。可归纳为 4 大类: 蓝色代表细胞过程(如细胞生长代谢、运动、分解代谢等), 共有基因 101 个; 紫色代表环境信息处理, 包括信号传导、膜的转运等, 共有 310 个基因; 粉色则代表基因信息传递(转录、翻译、蛋白质的折叠和加工等), 共有 169 个基因; 绿色代表 AH10 细胞机制, 这类基因最多, 行使的功能也最全面, 共有 703 个基因, 既有细胞本身合成氨基酸的基因, 也有氨基酸代谢、膜蛋白基因, 以及能量系统基因等。

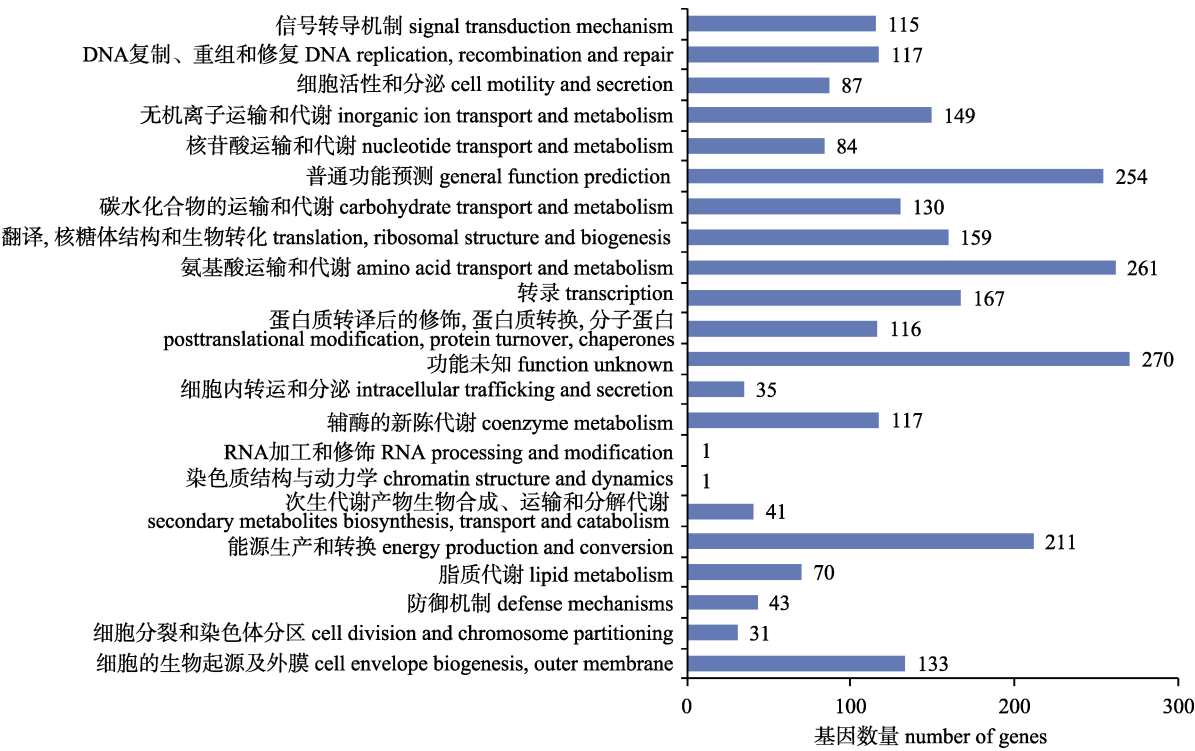


图 2 嗜水气单胞菌 AH10 的基因功能分类及解释

Fig. 2 The gene function classification and explanation for *Aeromonas hydrophila* AH10

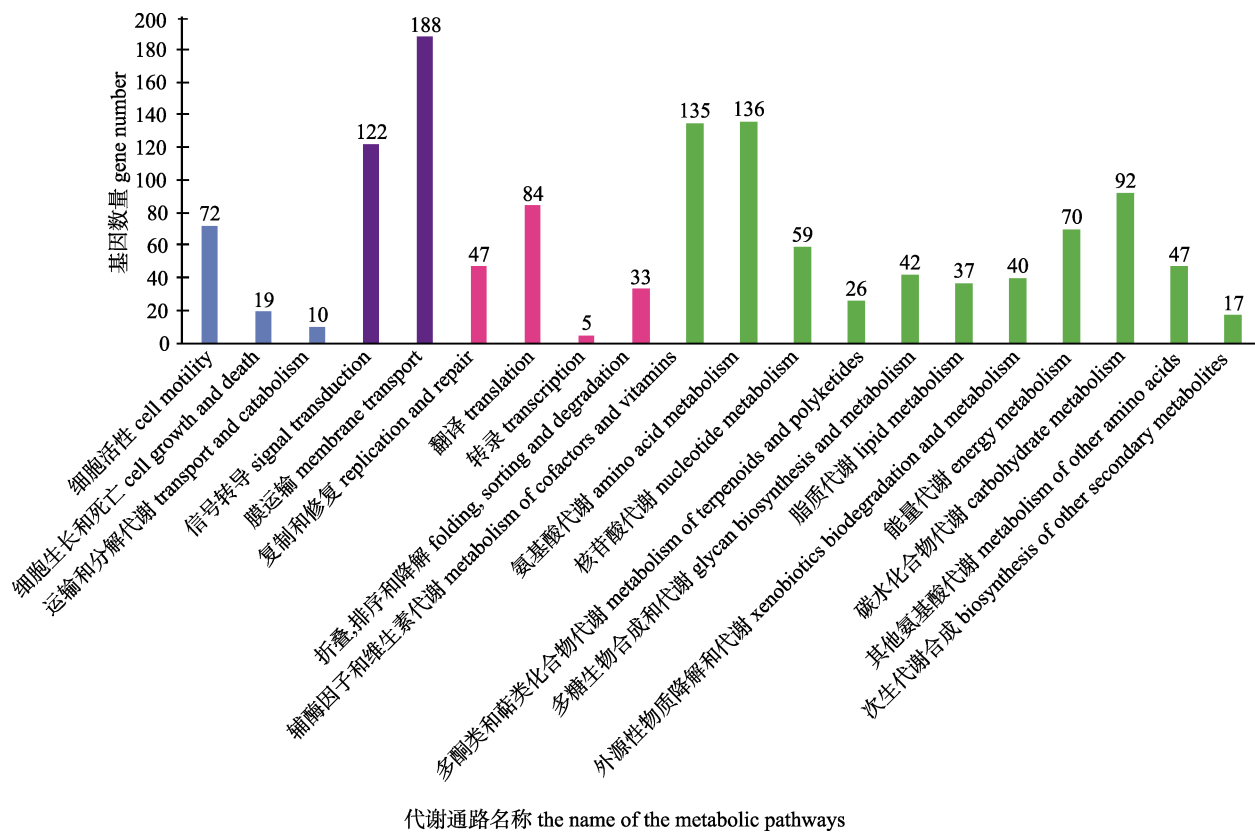


图 3 嗜水气单胞菌 AH10 的基因代谢通路

Fig. 3 The genetic metabolic pathways of *Aeromonas hydrophila* AH10

2.6 基因组岛

通过数据库多重比较方法比对, 共发现基因组岛(质粒、溶源性噬菌体、转座子和染色体上一些可移动的基因组区域)31 个。由表 2 可知, 基因组岛长度在 5~10 kb 的最多, 占比 41.93%。这些基因组岛中仅有 12 个具有生物功能, 虽然具有生物功能的基因组岛的数量较少, 但功能多样, 与蛋白质转运、酶合成、抗毒素、脂多糖合成、蛋白质修复、能量系统均有密切的关系。这些基因组岛中有 7 个基因组岛与多种生物功能相关, 其中有 1 个基因组岛含有的生物功能最多, 能够编码与蛋白质转运、脂多糖合成、蛋白质修补、酶合成、能量系统等相关的蛋白。

表 2 嗜水气单胞菌 AH10 的基因组岛长度分布
Tab. 2 The genomic island length distribution of *Aeromonas hydrophila* AH10

长度/kb length	数量 count
0-5	5
5-10	13
10-15	2
15-20	6
20-25	1
25-30	4

2.7 比较基因组学

2.7.1 突变位点分析 AH10 基因组通过与 ATCC7966、4AK4、ML09-119 和 YL17 基因组进行序列比对, 共找到 SNP 位点 102481 个。其中同义 SNP(三联密码改变但仍编码同一种氨基酸)85831 个, 占比为 83.8%; 非同义 SNP(该三联密码改变, 从编码一种氨基酸变为了编码另一种氨基酸)16650 个, 占比 16.2%。这些突变共来自 3771 个基因, 占 AH10 基因总数的 82.5%。互补碱基替换较为频繁(如 A 与 G 替换、T 与 C 替换), 而非互补碱基的替换则相对较少(如 A 与 T、T 与 G 替换)。A 碱基突变为 T 碱基的数量与 T 碱基突变为 A 碱基的数量差别较小, 其他碱基突变亦是如此(图 4)。

2.7.2 核心基因分析 将 AH10 与 ATCC7966、4AK4、ML09-119、YL17 的基因组进行比对, 发

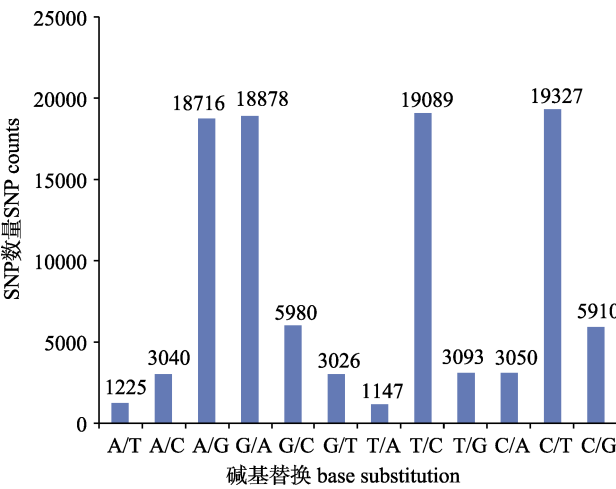


图 4 嗜水气单胞菌 AH10 基因组的 SNP 数量统计
Fig. 4 SNP number statistic result of *Aeromonas hydrophila* AH10 genome

现 AH10 中特有的基因最多, 占 AH10 基因组全部基因的 3.34%, 而 ML09-119 的特有基因最少, 占基因组全部基因的比例仅为 0.51%。在这些特有的基因中发现有蛋白质运输、金属离子转运通道相关基因(表 3)。

表 3 各嗜水气单胞菌菌株的核心基因分析结果
Tab. 3 The core gene analysis result of each *Aeromonas hydrophila* strain

菌株 strain	总数 total	非核心 non-core	非核心百分比% non-core percentage	特有 unique	特有百分比% unique percentage
AH10	4570	1377	30.13	154	3.34
ATCC 7966	4284	1091	25.47	50	1.17
4AK4	4074	881	21.62	56	1.37
ML09-119	4501	1308	29.06	23	0.51
YL17	4281	1088	25.41	25	0.58

2.7.3 基因共线性分析 基因共线性比较分析发现,AH10 与 ATCC7966、4AK4、ML09-119、YL17 共线性均较好, 存在倒位、重排、插入和缺失等现象(图 5)。AH10 与 ATCC7966、4AK4 的局部共线性(locally colinear blocks, LCBs)较多, 其中与 ATCC 7966 有 3 个较大的 LCBs, 对应的 AH10 基因组位置为: 176470-2161763、2202342-2663261、2794117-4908265; 与 4AK4 有 7 个 LCBs, 对应的 AH10 基因组位置为: 4391176-4812764、3188235-3592314、1532352-2061764、417647-814705、817956-1356092、2027960-2981256、3630866-4312421。AH10 与

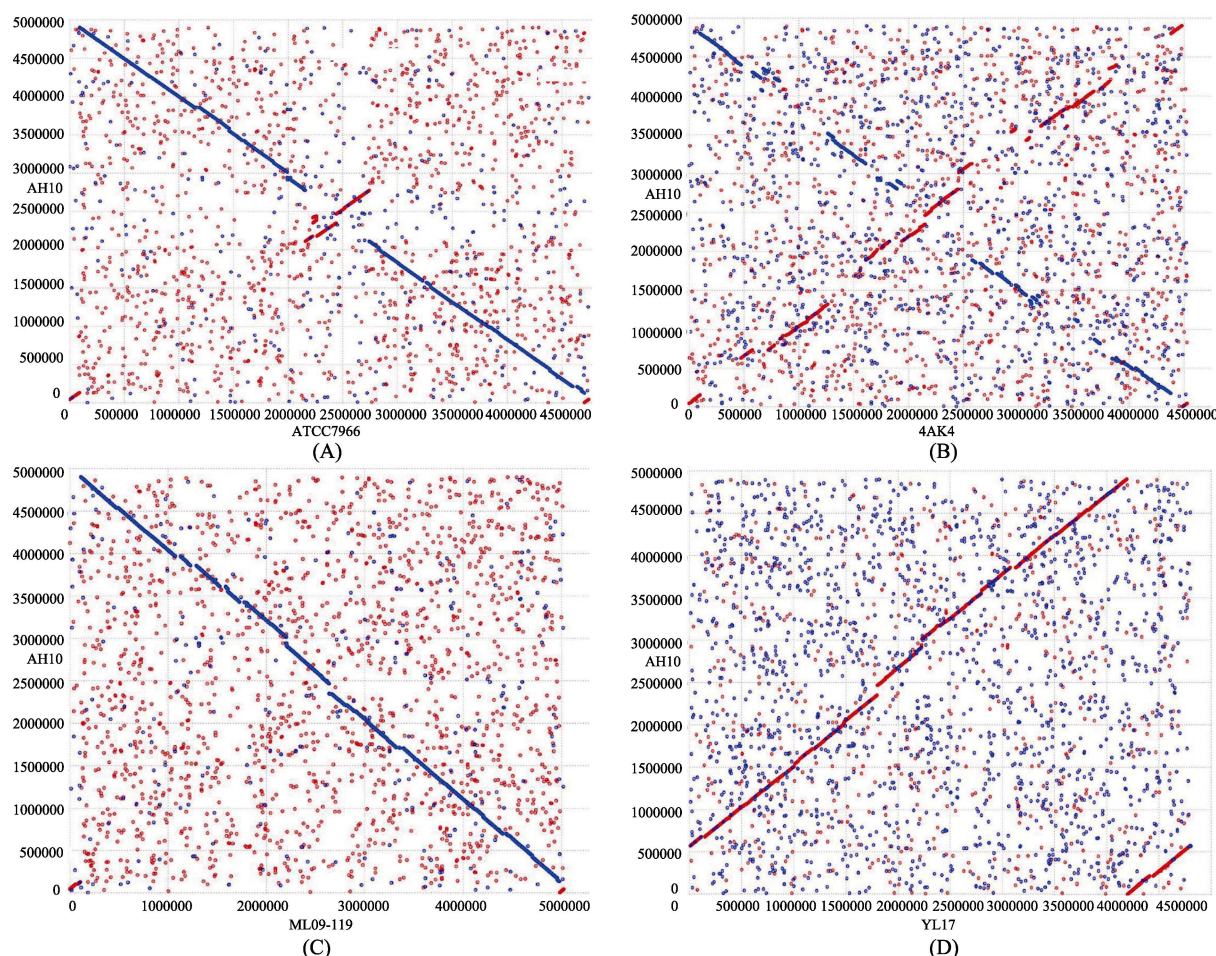


图 5 嗜水气单胞菌 AH10 与 ATCC7966(A)、4AK4(B)、ML09-119(C)、YL17(D)菌株基因组序列线性结构比较
红色代表两序列在此区域正向匹配, 蓝色代表两序列在此区域反向互补匹配。

Fig. 5 Global alignment between *Aeromonas hydrophila* AH10 and ATCC7966(A), 4AK4(B), ML09-119(C), YL17(D) strains, respectively

Red means the forward sequences of the two genome are the same in this region; blue means the reverse sequences are the same in this region.

ATCC7966 序列互补比对照较为一致, 但中间存在一个较大的倒位序列, 对应的 AH10 基因组位置为 2535294–2676326。AH10 与 4AK4 基因组序列比对中有较多缺失与插入现象, 正向与反向序列比对均有较多的 LCBs。AH10 与 ML09-119、YL17 无明显缺失与插入现象, 其中与 ML09-119 反向比对照较为一致。AH10 与 YL17 正向比对照较为一致, 但存在一个较大的重排现象, 在 AH10 基因组中的位置为 1–601310。

2.7.4 进化树构建 通过 Mega6.06 软件基于 16S rDNA 序列构建进化树(图 6)。AH10 与 ATCC7966、4AK4 处于一个分支上, 亲缘关系较近, 而 YL17 则单独处在一个分支上, 与其他菌株亲缘关系较

远。AH10 与美国株 ATCC7966 的亲缘关系比国内已经报道的嗜水气单胞菌菌株 J-1 近。

2.8 嗜水气单胞菌致病性分析

2.8.1 群体感应系统 群体感应(quorum sensing, QS)可以调控细菌生物被膜的形成, 进而增强细菌对抗生素、环境压力及宿主免疫系统攻击的耐受能力, 提高其毒性。已知在嗜水气单胞菌中, QS 系统中存在合成信号分子酶基因(signal molecule synthetase gene, *yI*)和转录调节蛋白基因(transcriptional regular gene, *AhyR*)。AH10 同时包含这两个基因, AH10 中存在 *ahyI*/QS 和 *ahyR*/QS 系统。据报道, 自体诱导物信号分子-2(autoinducer-2, I-2)QS 系统能够参与调控嗜水气单胞菌毒力基因表达^[10]。能够

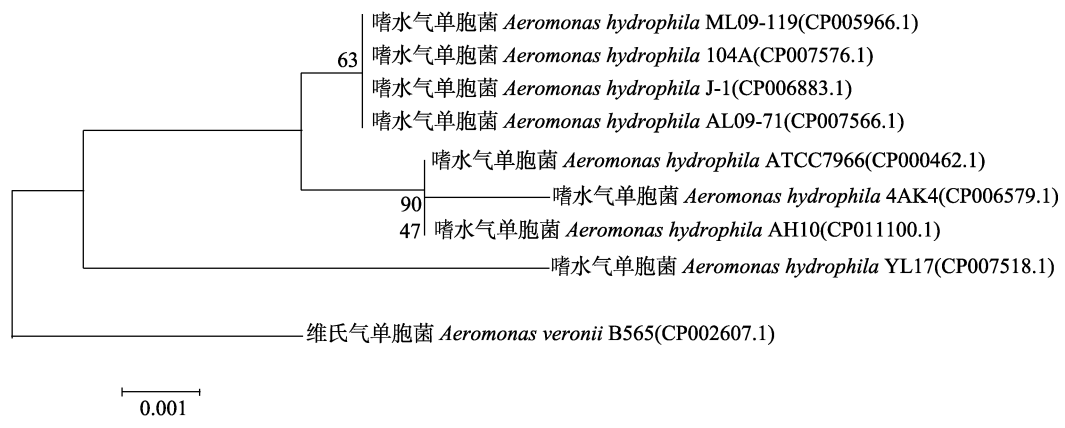


图 6 基于 16S rDNA 序列构建的系统发育邻近树
Fig. 6 Neighbor-joining algorithm based on 16S rDNA sequences

编码荧光酶 S(luciferase S)的基因 *LuxS* 是 AI-2QS 系统中的重要基因, AH10 和 ATCC7966 均发现 *LuxS* 的序列 (AH10 对应的基因组位置为: 4248595–4249104; ATCC7966 对应的基因组位置为: 748411–748920)。AH10 的致病性受 *ahyI* 和 *ahyR*/QS 系统与 *LuxS*/AI-2 QS 系统同时调控。具有典型 III 型分泌系统(type III secretion system, TTSS)的嗜水气单胞菌菌株 AH-1(来自 NCBI, 登录号: AY394563.2), 其 TTSS 基因簇在基因组中的位置为 109–26123, 共编码 35 个基因, 将其 TTSS 基因簇在 NCBI 中与 AH10 和 ATCC7966 进行同源序列比对, 并未找到同源序列, AH10 与 ATCC7966 均不存在 TTSS, 而国内株 J-1(登录号: CP006883.1)存在 TTSS^[11]。

2.8.2 嗜水气单胞菌主要致病相关基因 国内外学者已进行了大量有关嗜水气单胞菌的毒力基因及其与菌株致病力关系的研究, 众多的毒力相关基因中, 目前已鉴定的嗜水气单胞菌的毒力基因有溶血素基因 *hly* (hemolysin gene)、气溶素基因 *aer* (aerolysin gene)、细胞兴奋性肠毒素基因 *alt* (cytotoxic enterotoxin gene)、细胞毒性肠毒素基因 *act* (cytotoxic enterotoxin gene)、黏附素基因 *ahal* (major adhesion gene)、丝氨酸蛋白酶基因 *ahp* (serine protease gene)等。分析发现, AH10、ATCC7966、4AK4、YL17、ML09-119 以及国内已经报道的嗜水气单胞菌株 J-1 和 NJ-35(登录号: CP006870.1)全基因组中均包含 *hly*、*aer*、*ahal*、*ahp*、

alt、*ahyI*、*ahyR* 等毒力相关基因, AH10 编码区分别对应的基因组位置为: 3358273–3359580、4525089–4527598、1698424–1699526、183025–1831189、4876906–4878006、4402613–4403116、591813–592102。AH10 不编码 *act* 基因。因此, AH10 致病性可能与毒力基因 *hly*、*aer*、*ahal*、*ahp*、*alt* 相关。

2.8.3 致病相关基因序列比对分析 以 AH10 作为参考菌株, 比对分析发现, AH10 的 7 个致病相关基因 *hly*、*aer*、*ahal*、*ahp*、*alt*、*ahyI*、*ahyR* 与 ATCC7966 和国内株 J-1 和 NJ-35 在核苷酸水平上均高度保守(表 4)。各基因序列同源性在 82%~99.5%, 存在插入、缺失和替换现象。

表 4 各菌株主效毒力基因序列与参考菌株 AH10 的同源性

Tab. 4 Nucleotide homologies comparison of the virulence-associated genes of each strain with AH10 strain %							
菌株 strain	基因 gene						
	<i>alt</i>	<i>aer</i>	<i>ahp</i>	<i>ahal</i>	<i>hly</i>	<i>ahyI</i>	<i>ahyR</i>
ATCC 7966	98.0	96.9	98.3	99.3	98.9	99.4	99.5
J-1	98.1	97.8	98.0	98.9	98.2	98.6	99.2
NJ-35	98.0	97.8	98.0	99.2	98.5	98.6	99.8
ML09-119	98.1	98.0	98.0	99.3	98.2	98.6	99.8
4AK4	85.9	93.9	82.6	83.6	88.6	85.5	87.0
YL-17	94.5	—	97.3	87.0	97.4	94.6	97.8

注: “—”表示该基因序列不存在。
Note: “—” indicates that the gene sequence does not exist.

2.8.4 致病相关基因编码氨基酸序列比对分析
以 AH10 为参考, 各毒力基因编码的氨基酸

序列保守性差别较大。*hly* 编码的氨基酸序列保守性在 94%~100%。国内株 NJ-35 除基因 *ahal* 外, 其余各基因编码的氨基酸序列与 AH10 各基因的氨基酸均在 99%以上。ATCC7966 与 AH10 中 *ahyI* 基因虽然序列同源性较高(表 4), 但编码的氨基酸序列同源性较低, 仅为 87.4%。4AK4 各毒力基因与 AH10 各毒力基因编码的氨基酸序列同源性均较低(表 4, 表 5)。

表 5 各菌株主效毒力基因编码氨基酸序列与参考菌株 AH10 的比对结果

Tab. 5 Amino acid sequence homologies comparison of the virulence-associated genes of each strain with AH10 %

菌株 strain	基因 gene						
	<i>alt</i>	<i>aer</i>	<i>ahp</i>	<i>ahal</i>	<i>hly</i>	<i>ahyI</i>	<i>ahyR</i>
ATCC 7966	99.8	98.9	99.3	100	100	87.4	99.5
J-1	99.4	99.4	99.3	*	100	99.5	*
NJ-35	99.4	99.4	99.3	92.4	100	100	100
ML09-119	99.4	99.2	99.1	99.4	100	100	89.2
4AK4	86.7	83.4	82.8	80.3	94.1	87.4	92.1
YL-17	96.3	—	99.0	87.0	99.0	96.6	98.5

注: “*”表示未比对到该氨基酸序列; “—”表示该氨基酸序列不存在。
Note: “*” indicates that the amino acid sequence could not be found; “—” indicates that the amino acid sequence does not exist.

3 讨论

在人类基因组计划的带动下, 微生物的研究越来越集中于全基因组层面。全基因组测序正在彻底改变细菌学的研究领域, 1995 年美国 TIGR 研究所完成流感嗜血杆菌(*Haemophilus influenzae*)的全基因组测序, 标志着微生物基因组学时代的到来^[12]。在基因组学快速发展的同时, 大量的细菌模式生物开始建立, 为细菌性疾病的治疗与防治提供了重要依据^[13]。嗜水气单胞菌致病性强, 其致病性与毒力因子密切相关^[14]。1987 年, Chakraborty 等^[15]对嗜水气单胞菌气溶素(*aerA*)结构基因做了较为全面的研究, 证实该基因对嗜水气单胞菌的致病性有较大影响。2010 年 Singh 等^[16]对溶血素 *hlyA* 基因进行了克隆表达, 表明该基因对宿主的破坏力较强。本研究发现, AH10 存在着气溶素基因和溶血素基因, 这说明这两个基因簇在 AH10 致病性中起着重要的作用。对比各菌株

hly、*aer*、*ahal*、*ahp*、*alt*、*ahyI*、*ahyR* 基因的序列及氨基酸, 保守性均较高, 说明这些基因在遗传过程中发生变异的概率较小。各毒力相关基因在各菌株基因组中存在的位置不同, 可能与菌株在遗传过程中发生了基因重组有关。2010 年 Carvalho-castro 等^[17]对嗜水气单胞菌 III 型分泌系统(the type III secretion system, TTSS)进行了系统深入的研究, 发现存在该系统的嗜水气单胞菌菌株致病力较强。但 AH10 却不存在 TTSS, 而是存在 *ahyI*/QS 和 *ahyR*/QS 系统与 LuxS/AI-2 QS 系统。这表明嗜水气单胞菌存在着多种 QS 系统, QS 系统在嗜水气单胞菌致病性中起着不可忽视的作用。本研究中对 AH10 全基因组进行了较为全面的注释, 并对基因组信息进行了 COG 功能分类和代谢通路的构建, 不但能为快速找到其毒力因子基因提供帮助, 还能对毒力因子相关基因有一个较为全面的认识, 对于研究嗜水气单胞菌致病性及其致病的分子机制意义重大, 也为嗜水气单胞菌疫苗的研制, 免疫学研究打下了基础。

随着各种基因组信息的积累, 比较基因组学已经成为各种生物学发现强有力的工具^[18-20]。不同细菌在长时间的进化中, 其基因的结构和功能存在着明显的不同^[21]。本研究, AH10 与 ATCC7966、4AK4、ML09-119、YL17 虽为同一个属, 由一个分支进化而来, 但在长期进化过程中其基因组信息差别较大: SNP 位点较多, 这些突变遍布于基因组的每一个结构的基因, 为研究基因变化和蛋白表达的不同提供了原材料。基因共线性基因研究发现, AH10 与 ATCC7966、4AK4、ML09-119、YL17 的基因组中存在大量插入缺失和重排现象, 基因组信息有明显不同。这些基因共线性结构的差异, 为研究不同菌株基因组框架的组成和种属进化关系提供了很好的向导。胡攀^[22]研究发现, 基因组岛在基因组中广泛存在。在 AH10 基因组中共发现基因组岛 31 个, 这些基因组岛广泛存在于 AH10 基因组中, 这与胡攀^[22]的研究结果较为一致。胡力文等^[23]研究发现: 基因组岛常呈现出多个功能模块相嵌合的结构特征。AH10 中基因组岛也呈现出多个功能模块相嵌合的现象。在核

心基因统计中,每株嗜水气单胞菌均有其特有的基因,这为区别同属不同株的研究指明了方向。

基因组学的发展使生物特性、种群进化、基因分析、致病机制、新药开发等方面的研究具有更加广阔的前景^[24]。近年来,多株嗜水气单胞菌基因组序列的获得,使人们对嗜水气单胞菌的基因组信息有了较为深入的了解,这些信息有助于下一步阐明嗜水气单胞菌的致病机制及开发新的抗病策略。

参考文献:

- [1] Lu C P. Pathogenic *Aeromonas hydrophila* and the fish diseases caused by it[J]. Journal of Fisheries of China, 1992, 16(3): 282–288.[陆承平. 致病性嗜水气单胞菌及其所致鱼病综述[J]. 水产学报, 1992, 16(3): 282–288.]
- [2] Yang S M, Wang M S. *Aeromonas hydrophila* and its pathogenicity to human[J]. Chinese Journal of Disease Control and Prevention, 2006, 10(5): 511–514.[杨守明, 王民生. 嗜水气单胞菌及其对人的致病性[J]. 疾病控制杂志, 2006, 10(5): 511–514.]
- [3] Lv A J, Li J N, Yu W Y. Progress of research on *Aeromonas Hydrophila*[J]. Chinese Journal of Animal Quarantine, 2000, 10(11): 42–43.[吕爱军, 李瑾年, 余为一. 嗜温气单胞菌研究进展[J]. 中国动物检疫, 2000, 10(11): 42–43.]
- [4] Zhang X J. Construction and characterization of *luxS* deleted mutant strain in *Aeromonas hydrophila* J-1[D]. Nanjing: Nanjing Agricultural University, 2009.[张小军. 嗜水气单胞菌 J-1 株 *luxS* 基因缺失株的构建及特性分析[D]. 南京: 南京农业大学, 2009.]
- [5] Yu X H, Wang Y W, Tang C, et al. Progress in studies on *Aeromonas hydrophila*[J]. Journal of Southwest University for Nationalities: Natural Science Edition, 2007, 33(3): 507–514.[于学辉, 王远微, 汤承, 等. 嗜水气单胞菌的研究进展[J]. 西南民族大学学报: 自然科学版, 2007, 33(3): 507–514.]
- [6] Xu L J. Establishing the medication regimen for prevention of drug resistance using enrofloxacin against *Aeromonas hydrophila* freshwater[D]. Shanghai: Shanghai Ocean University, 2014.[徐丽娟. 恩诺沙星控制淡水鱼类嗜水气单胞菌性败血症的防耐药用药方案研究[D]. 上海: 上海海洋大学, 2014.]
- [7] Liu W J. Whole genome sequencing and comparative genomics analysis of *Pasteurella multocida*[D]. Wuhan: Huazhong Agricultural University, 2012.[刘文静. 多杀性巴氏杆菌全基因组测序与比较基因组学分析[D]. 武汉: 华中农业大学, 2012.]
- [8] Li G Y. Comparative genomic and transcriptomic analysis of two strains of *Vibrio anguillarum*[D]. Qingdao: Graduate University of Chinese Academy of Sciences (Institute of oceanography), 2011.[李贵阳. 两株鳃弧菌全基因组序列测定及转录组比较分析[D]. 青岛: 中国科学院研究生院(海洋研究所), 2011.]
- [9] Delcher A L, Bratke K A, Powers E C, et al. Identifying bacterial genes and endosymbiont DNA with Glimmer[J]. Bioinformatics, 2007, 23(6): 673–679.
- [10] Zhao J, Zhang F R, Cui Y, et al. In vitro biosynthesis and activity detection of quorum sensing signal molecule AI-2 of *Aeromonas hydrophila*[J]. Microbiology China, 2015, 42(10): 1858–1865.[赵晶, 张福蓉, 崔一, 等. 嗜水气单胞菌群体感应信号分子 AI-2 的细胞外生物合成及活性检测[J]. 微生物学通报, 2015, 42(10): 1858–1865.]
- [11] Xie Z X. Construction of a *AscV* isogenic knockout *Aeromonas hydrophila* J-1 mutant and study on omp of *Aeromonas hydrophila*[D]. Nanjing: Nanjing Agricultural University, 2005.[谢振宣. 嗜水气单胞菌 III 型分泌系统 *AscV* 基因突变株的构建及外膜蛋白的研究[D]. 南京: 南京农业大学, 2005.]
- [12] Fleischmann R D, Adams M D, White O, et al. Whole genome random sequencing and assembly of *Haemophilus influenza* Rd[J]. Science, 1995, 269(5223): 496–512.
- [13] Liu X Y. Microbial genome and development of antimicrobial[J]. Epidemiology Lemology Foreign Medical Sciences, 2002, 29(3): 138–142.[刘晓颖. 微生物基因组与抗微生物药物的发展[J]. 国外医学(流行病学传染病学分册), 2002, 29(3): 138–142.]
- [14] Qiu J Q, Yang X L, Cheng X J. Advances in the characteristics and pathogenic mechanisms of virulence factors in *Aeromonas hydrophila*[J]. Journal of Pathogen Biology, 2009, 4(8): 616–619.[邱军强, 杨先乐, 程训佳. 嗜水气单胞菌毒力因子特性及作用机理研究进展[J]. 中国病原生物学杂志, 2009, 4(8): 616–619.]
- [15] Chakraborty T, Huhle B, Hof H, et al. Marker exchange mutagenesis of the aerolysin determinant in *Aeromonas hydrophila* demonstrates the role of aerolysin in *A. hydrophila*-associated systemic infections[J]. Infect Immun, 1987, 55(9): 2274–2280.
- [16] Singh V, Somvanshi P, Rathore G, et al. Gene cloning, expression, and characterization of recombinant aerolysin from *Aeromonas hydrophila*[J]. Appl Biochem Biotechnol, 2010, 160(7): 1985–1991.
- [17] Carvalho-castro G A, Lopes C O, Leal C A G, et al. Detection of type III secretion system genes in *Aeromonas hydrophila* and their relationship with virulence in Nile tilapia[J]. Vet Microbiol, 2010, 144(3/4): 371–376.
- [18] King D C, Taylor J, Zhang Y, et al. Finding cis-regulatory

- elements using comparative genomics:some lessons from ENCODE data[J]. *Genom Res*, 2007, 17(6): 775–786.
- [19] Miller W, Makova K D, Nekrutenko A, et al. Comparative genomics[J]. *Annu Rev Genom Hum Genet*, 2004, 5: 15–56.
- [20] Nobrega M A, Pennacchio L A. Comparative genomic analysis as a tool for biological discovery[J]. *Physiol Soc*, 2004, 554 (Pt 1): 31–39.
- [21] Zhou D S, Yang R F. The development device and present studies of biological nitrogen fixation in Chinese agriculture production[J]. *Journal of Microbiology*, 2003, 23(5): 31–34, 43.[周冬生, 杨瑞馥. 细菌比较基因组学和进化基因组学[J]. *微生物学杂志*, 2003, 23(5): 31–34, 43.]
- [22] Hu P. Whole genome sequencing and comparative genomics analysis of *Streptococcus suis*[D]. Wuhan: Huazhong Agricultural University, 2011.[胡攀. 猪链球菌全基因组测序与比较基因组学研究[D]. 武汉: 华中农业大学, 2011.]
- [23] Hu L W, Yao X Y, Hu F Q. The transfer mechanism, evolutionary process and functional feature of genomic islands[J]. *Chemistry of Life*, 2013, 33(2): 75–82.[胡力文, 姚新月, 胡福泉. 细菌基因组岛转移机制、进化历程及功能特点[J]. *生命的化学*, 2013, 33(2): 75–82.]
- [24] Saccone N L, Neuman R J, Saccone S F, et al. Genetic analysis of maximum cigarette-use phenotypes[J]. *BMC Genet*, 2003, 4(Suppl 1): S105.

Whole-genome sequencing and comparative analysis of *Aeromonas hydrophila* AH10 (CCTCC AB2014155)

ZHANG Guoliang¹, WANG Hao¹, ZHANG Ye¹, XU Dan¹, LÜ Liqun^{1,2}

1. National Pathogen Collection Center for Aquatic Animals, Key Laboratory of Freshwater and Germplasm Resources Utilization, Ministry of Agriculture; College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China;
2. Aquaculture Collaborative Innovation Center of Hubei Province, Wuhan 430070, China

Abstract: *Aeromonas hydrophila* exists widely in aquatic environments throughout the world. It is highly pathogenic and infection with this bacterium entails extremely high mortality. To facilitate the prevention and treatment of the diseases caused by *A. hydrophila*, we sequenced the whole genome of *A. hydrophila* AH10 (CCTCC AB2014155), to fully understand its information structure and functions and to provide a basis for future molecular pathology research. Genomic DNA was extracted from strain AH10, and a library was constructed and sequenced. The raw data were filtered and their quality was assessed, the sequence was assembled, and the genome was annotated and analyzed comparatively. The full-length genome of strain AH10 is 4.91 Mb, with a G+C content of 61.1%, and encodes 4570 predicted genes, including 3351 genes with a clear biological function, 2592 genes found in the Clusters of Orthologous Groups (COG) database, 1281 genes related to metabolic pathways, and 31 genomic islands. The study of comparative genomics showed that AH10 has 102481 single-nucleotide polymorphisms (SNPs), 402 insertions/deletions, and a linear genomic organization most similar to *A. hydrophila* YL17. AH10 is closely related to *A. hydrophila* ATCC7966, the standard strain in the USA. The pathogenicity of AH10 is closely associated with the hemolysin gene (*hly*), aerolysin gene (*aer*), major adhesion gene (*ahal*), serine protease gene (*ahp*), and cytotoxic enterotoxin gene (*alt*). These genes are highly homologous in different strains. Autoinducer-2 (AI-2) quorum sensing (QS) and AhyI/AhyR QS are found in AH10, but there is no type III secretion system (TTSS). With a comparative analysis, we demonstrate the specificity of the AH10 genomic structure, its SNPs, and the pathogenic system differences in different strain. Our analysis not only provides a guide to distinguishing the specificities of different species, but also a reference for the analysis of the genomic information of *A. hydrophila*. Research into the *A. hydrophila* genome has accelerated related studies of its pathogenicity. Identifying the causes of disease requires that the accuracy of their clinical diagnosis be effectively improved. This, in turn, will offer clues to their pathogenic mechanisms in humans. The study of *Aeromonas hydrophila* genome also provides the foundation to the research of proteomics. According to the specific functional areas of protein conservative sequences, setting up a corresponding protein database can help to quickly find the protein function area. In this way, we can not only study the cell wall composition and the functions of the membrane proteins and secreted proteins of *A. hydrophila*, but can also predict the functions of related proteins based on genomic information. This research provides unlimited opportunities for the development of new drugs that not only effectively avoid the generation of drug resistance, but also circumvent the abuse of antibiotics, thus providing people with a healthy environment. Studies of the *A. hydrophila* genome should also promote vaccine research. With the comparison of homologous sequences, effective antigen can be rapidly identified and vaccines prepared, efficiently reducing the incidence of disease. Genomic information has potentially wide applications, and its use in the prevention and treatment of the diseases caused by *A. hydrophila* will continue to increase.

Key words: *Aeromonas hydrophila*; genome sequencing; gene; comparative analysis

Corresponding author: LÜ Liqun. E-mail: lqlv@shou.edu.cn