

两种嗜盐古菌中 circRNA 的鉴定与特征分析

雒雯琦, 王小林, 杨彦, 单革, 陈亮

(中国科学技术大学生命科学与医学部, 安徽合肥 230027)

摘要: 环形 RNA(circRNA) 是一类特殊的闭合环状 RNA, 主要由 5' 末端和 3' 末端通过反向剪接的方式共价结合形成. 在多细胞动物中, circRNA 具有一定的进化保守性和组织特异性. 古菌作为生物三大领域之一, 大多生活在高温、高盐、低氧等极端环境中. 对古菌的研究是认知生命进化不可缺少的一环. 目前对于古菌 circRNA 鉴定及特征并没有系统性的研究. 选取了地中海富盐菌和西班牙盐盒菌两种嗜盐古菌为研究对象, 利用 RNA 测序(RNA-seq)技术和生物信息学手段, 鉴定出两种古菌中 circRNA 的存在, 并深入分析了 circRNA 的特点; 最后以分子生物学手段, 对两种嗜盐古菌中部分 circRNAs 进行验证. 在地中海富盐菌中, 检测到 67 个 circRNAs; 在西班牙盐盒菌中, 检测到 133 个 circRNAs. 在这两种嗜盐古菌中, circRNA 主要源于 16S rRNA、23S rRNA、16S rRNA-23S rRNA 和蛋白编码基因序列.

关键词: 地中海富盐菌; 西班牙盐盒菌; 环形 RNA; RNA 测序

中图分类号: Q522⁺.6 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2020.04.010

引用格式: 雒雯琦, 王小林, 杨彦, 等. 两种嗜盐古菌中 circRNA 的鉴定与特征分析[J]. 中国科学技术大学学报, 2020, 50(4): 457-466.

LUO Wenqi, WANG Xiaolin, YANG Yan, et al. Verification and characterization of circRNAs in two halophilic archaea[J]. Journal of University of Science and Technology of China, 2020, 50(4): 457-466.

Verification and characterization of circRNAs in two halophilic archaea

LUO Wenqi, WANG Xiaolin, YANG Yan, SHAN Ge, CHEN Liang

(Department of Life Science and Medicine, University of Science and Technology of China, Hefei 230027, China)

Abstract: Circular RNA (circRNA) is a special type of RNA, which forms covalent-closed continuous loop in most cases through 5' to 3' ends via back-splicing. CircRNAs demonstrate a certain degree of conservation and show some tissue specificity in multicellular animals. Archaea is one of the three domains of life, and archaeal organisms often live in high temperature, high salt, low oxygen, or other extreme environments. Study of archaea is an essential part of evolutionary research. So far, there have been no systematic studies on the identification and functional characterization of archaeal circRNAs. In this study, two halophilic archaea *Haloferox mediterranei* and *Haloarcula hispanica* were investigated. Through RNA-seq and bioinformatics, the existence of some circRNAs in two halophilic archaea were confirmed and the relevant characteristics of circRNAs were analyzed in depth. Furthermore, the existence of specific circRNAs in two halophilic archaea was confirmed by means of molecular biology. 67 circRNAs in the

收稿日期: 2020-01-16; 修回日期: 2020-04-22

基金项目: 国家杰出青年科学基金(31725016), 国家自然科学基金青年基金(31600657)资助.

作者简介: 雒雯琦, 女, 1994 年生, 硕士生. 研究方向: 非编码 RNA. E-mail: luowq18@mail.ustc.edu.cn

通讯作者: 单革, 博士/教授. E-mail: shange@ustc.edu.cn. 陈亮, 博士/副研究员. E-mail: anqingel@ustc.edu.cn

Haloferax mediterranei and 133 circRNAs in the *Haloarcula hispanica* were detected. CircRNAs in the two halophilic archaea were found to have derived mainly from sequences of 16S rRNA, 23S rRNA, 16S rRNA-23S rRNA, and protein-coding genes.

Key words: *Haloferax mediterranei*; *Haloarcula hispanica*; circRNA; RNA-seq

0 引言

环形 RNA (circRNA) 是一类共价闭环且无 polyA 尾巴的非编码 RNA, 最早在 1976 年发现于 RNA 病毒中^[1]. 1979 年首次在真核生物中发现内源性 RNA 剪接产物^[2]. 最开始仅仅认为它们是拼接错误的产物^[1], 随着 RNA 测序 (RNA-seq) 技术和生物信息学的发展, 研究人员在不同物种的细胞以及组织中鉴定出了大量的 circRNA, 也揭示了 circRNA 在生命体各个发育阶段和不同生理状态下的动态表达模式^[3]. 目前已经证实 circRNA 的功能主要有: miRNA 海绵, 调控转录, 细胞间信息传递和蛋白质翻译^[4,5], 选择性剪接^[6] 和与 RNA 结合蛋白相互作用等^[5,7]. 很多研究显示, circRNA 也参与了多种肿瘤的发生和发展过程, 暗示 circRNA 可能在多种人类疾病中起着重要的作用^[8].

古菌与细菌和真核生物并列为生命进化树的三大生命领域. 起初人们认为古菌是细菌的一种, 但后来研究人员发现, 两者存在明显的不同, 古菌的 tRNA、rRNA 与真核细胞有着相似的结构, 且古菌中的某些基因上也含有内含子, 在分子水平上更接近真核生物. 绝大多数古菌一直生活在极端 (高温、高盐等) 的环境中, 这种生活环境很像地球最初形成时的状态, 所以研究古细菌对于更好地了解生命起源以及物种进化具有重要的意义. 有研究表明, 古菌染色体与真核生物染色体具有相似的 DNA 簇, 这些 DNA 簇之间的互作形式, 对于基因的表达具有重要的作用, 因古菌细胞相对简单, 所以古菌可作为研究更为复杂的真核细胞 (特别是人类细胞) 生长过程中 DNA 互作的良好模型^[9].

在古菌中, circRNA 目前已知主要存在于 tRNA 和 rRNA 内含子, 以及作为 rRNA 加工中间体存在^[10], 其中的某些古菌环形内含子已被证实具有生物学功能. 例如, 在嗜盐古菌 *Haloferax volcanii* 中, tRNA^{Trp} 环形内含子上的 C/D box RNA, 可以指导成熟 tRNA^{Trp} 的第 34 和 39 位核苷酸的化学修饰^[11,13]. 在广古菌 (*Euryarchaeota*) 和

泉古菌 (*Crenarchaeota*) 两种古菌中, 已证实有以 rRNA 加工中间体形式存在的环形 23S rRNA 和 16S rRNA^[14-15]. 然而目前对于古菌 circRNA 的研究依然很少, 对其定位及潜在功能并不清楚.

地中海富盐菌 (*Haloferax mediterranei*) 和西班牙盐盒菌 (*Haloarcula hispanica*) 是两种极端嗜盐古菌的代表, 人们已经完成了对其全基因组的测序. 有研究人员首次在地中海富盐菌中证实了与真核生物类似的多复制起点、休眠复制起点及休眠点被激活的现象^[16]. 这一结果暗示古菌中 circRNA 的定位、特点和潜在的功能很可能与真核生物中 circRNA 相似, 可作为我们研究真核生物 circRNA 的新的突破点. 地中海富盐菌和西班牙盐盒菌相对易于培养, 且目前未有关于其 circRNA 的报道, 所以我们选择它们作为本文的研究对象. 我们利用 RNA-seq 技术, 对地中海富盐菌和西班牙盐盒菌中经过 RNase R 消化/未消化的 total RNA 进行测序. 结合生物信息学方法, 分析并归纳了两种古菌中 circRNA 特点, 并利用分子生物学手段, 验证部分 circRNAs 的存在. 本文在极端嗜盐古菌中对 circRNA 进行鉴定和特征分析, 扩大了我们对于 circRNA 在不同物种中的认知, 丰富了 circRNA 的研究领域.

1 材料和方法

1.1 材料

菌种及菌种的培养: 地中海富盐菌 (*Haloferax mediterranei* CGMCC1.3716)、西班牙盐盒菌 (*Haloarcula hispanica* CGMCC1.2049), 均购于中国微生物菌种保藏中心, 两者均采用 AS-168 培养基, 于 37 °C 摇床培养.

主要试剂: TRIzol (Invitrogen)、Yeast extract、Casamino acids、FeCl₂、NaCl、MgSO₄ · 7H₂O、KCl、Agar、RNase R (Lucigen).

1.2 RNA 的抽提

将冻存的地中海富盐菌、西班牙盐盒菌在 AS-168 固体斜面培养基中划线复苏, 待 2 d 后形成单克

隆菌体,挑取克隆至 20 mL AS-168 液体培养基中,37 °C,200 r/min 培养 2 d 后,以 1:20 转至 100 mL 液体培养基中扩大培养 2 d 至对数期,取 10 mL 菌液,分为两管,离心收集菌体,用 PBS 清洗 2 次,离心弃上清,加入适量 TRIzol (Invitrogen),抽提 RNA. 取 1 μ L RNA 样品,80 °C 变性 10 min,进行甲醛 RNA 变性胶电泳,检测 RNA 质量;取 1 μ L RNA 样品,检测其浓度及 OD260/OD280.

1.3 基因组 DNA(gDNA)的抽提

从固体培养基上挑取地中海富盐菌、西班牙盐盒菌的菌体至 20 mL AS-168 液体培养基中,37 °C 摇床培养至对数期,取 10 mL 菌液,离心收集菌体. PBS 清洗 2 次. 加入 Tris-EDTA 缓冲液,-20 °C 30 min 后加入玻璃珠,高速震荡破壁. 之后加入 1 μ L 蛋白酶 K,56 °C 消化 3 h,加入酚氯仿试剂,充分混匀,静置 2 min,12000 g,离心 10 min. 加入酚氯仿试剂抽提 2 次,向其中加入等体积异丙醇-20 °C 沉淀 1 h,12000 g,离心 15 min,去掉上清,加入 50 μ L ddH₂O,得到 gDNA. 取 1 μ L 检测其浓度及 OD260/OD280.

1.4 RNase R 消化

向 5 μ g total RNA 样品中加入 0.3 μ L RNase R,5 μ L RNase R buffer,加水补齐至 50 μ L,于 37 °C 消化 40 min,加入 500 mL TRIzol 抽提 RNA. 取 1 μ L RNA 样品进行 RNA 变性凝胶电泳,检测 RNA 质量;取 1 μ L RNA 样品,检测其浓度及 OD260/OD280.

1.5 RT-PCR

取等体积 RNase R 消化/未消化的 total RNA,用 ABScript II cDNA 合成试剂盒 (Abclonal) 进行反转录. 反转体系为:5 μ L RNA,2 μ L Random Primer,1 μ L 10 mmol/L dNTPs,80 °C 5min;后加入 10 μ L ABScript II Reaction Mix (2X),2 μ L ABScript II Enzyme Mix (10X) 混匀,反转程序为:25 °C,5 min,42 °C,1 h;80 °C,5 min. 以等体积 RNase R 消化/未消化的 cDNA,基因组 DNA 为模板,以 divergent/convergent 引物进行 PCR,验证 circRNA 的存在.

1.6 RNA 测序

从样品中提取 total RNA 后,将其分为两组,一组为 RNase R 未消化处理 total RNA 样品,一组为

RNase R 消化处理的 total RNA 样品. 之后两组样品通过试剂盒从总 RNA 中去除 rRNA 来富集 mRNA 和非编码 RNA. 利用片段缓冲液,将 mRNA 和非编码 RNA 片段化成短片段(约 200~700 bp),然后以短片段为模板,通过随机六聚体引物合成第一链 cDNA. 加入缓冲液,dNTP,RNase H 和 DNA 聚合酶 I 合成第二条链 cDNA. 用 QIAquick PCR 提取试剂盒纯化双链 cDNA,然后用于末端修饰. 将测序衔接子连接至片段,然后使用 UNG(尿嘧啶-N-糖基化酶)最终降解第二条链. 片段通过琼脂糖凝胶电泳纯化并通过 PCR 富集,文库产品可用于通过 Illumina HiSeqTM 4000 仪进行测序分析. 对于高通量测序,根据测序的要求制备地中海富盐菌和西班牙盐盒菌的 RNaseR 消化和未消化的 RNA 样品,并将其应用于 Illumina GAIIX 系统进行 150 nt 双端测序. 该过程在华大基因完成测序.

1.7 生物信息学分析

地中海富盐菌和西班牙盐盒菌的参考基因组从 NCBI 基因组网站 (<https://www.ncbi.nlm.nih.gov>) 下载. 滤除短于 50 nt 的读段(reads),其余的 reads 根据算法 find_circ^[17] 预测 circRNA. 首先,我们将 reads 与参考基因组比对,将与基因组连续且完整地比对上的 reads 舍弃;然后,从其余的 reads 中,我们从两端分别提取 20 nt,作为一对 anchor,将其分别与参考基因组进行比对,确定它们在基因组上的位置,以相反方向比对上的 anchor 预示着环形 RNA 剪切;最后,扩展 anchor 的比对以使整个读段与参考基因组比对,根据 GU/AG 确定剪切位点.

1.8 CircRNA 的验证

我们利用生物信息学手段对测序结果进行分析,将所得 circRNA 序列比对到相应古菌的基因组上,确定 circRNA 序列的起始位点和终止位点,得到成环的序列片段. 随机选择在测序数据中丰度较大的一些 circRNA 进行验证. 依据其成环连接区域(junction)设计 divergent/convergent 引物^[18],以 RNase R 消化和未消化的 total RNA 反转录出的 cDNA 和基因组 DNA(gDNA)为模板进行 PCR,验证这两个样本中 circRNA 的存在. 本文所用引物如表 1 所示.

表 1 本文使用的引物
Tab. 1 The primers used in this study

primers	sequences(5'-3')	purpose	archaea
circ26-di-f	GGAGTTGTTCTGTGATCGTC	Divergent primers for RT-PCR detection of circ28 incDNA or genomic DNA(gDNA)	<i>H. mediterranei</i>
circ26-di-r	GCGATATCCCGATCGTCACG		
circ48-di-f	CACTTAATCCACACTTTGGG	Divergent primers for RT-PCR detection of circ48 incDNA or genomic DNA(gDNA)	
circ48-di-r	TACTGCATGGACGGAAGTAC		
circ53-di-f	AGGCATAGGGAAACCGATA	Divergent primers for RT-PCR detection of circ53 incDNA or genomic DNA(gDNA)	
circ53-di-r	TCGAGCCTTCCGGTCCCAA		
circ157-di-f	ACGGACCACCGATGGCGAAA	Divergent primers for RT-PCR detection of circ157 incDNA or genomic DNA(gDNA)	
circ157-di-r	CTTATCGCACTTGCGTGCAG		
circ183-di-f	ATCACTTGAGCTGCCGGTAT	Divergent primers for RT-PCR detection of circ157 incDNA or genomic DNA(gDNA)	
circ183-di-r	AAATCCGCCAGCTCAACTG		
circ9-di-f	TACCTTCGGAAGAAGGGATG	Divergent primers for RT-PCR detection of circ9 incDNA or genomic DNA(gDNA)	<i>H. hispanica</i>
circ9-di-r	AGCATCGTTTACAGCTAGGAC		
circ56-di-f	CTCTTAAGGTAGCGTAGTAC	Divergent primers for RT-PCR detection of circ56 incDNA or genomic DNA(gDNA)	
circ56-di-r	CTATCGCACTGTGCTGCAGT		
circ135-di-f	GAAGAAGGGATGCCTGCTCC	Divergent primers for RT-PCR detection of circ135 incDNA or genomic DNA(gDNA)	
circ135-di-r	CCTGCGGCCATGTCGGTTT		
circ156-di-f	TGAGGGACGAAAGCTAGGGT	Divergent primers for RT-PCR detection of circ156 incDNA or genomic DNA(gDNA)	
circ156-di-r	TTGTCTCAGATACCGTCTCC		
circ161-di-f	CGAGACGTTCTGCTGCTCGG	Divergent primers for RT-PCR detection of circ161 incDNA or genomic DNA(gDNA)	
circ161-di-r	AAGGTGAAGCGGCTCTCGG		
circ26-con-f	AGGAGGTCGCCGAGGGTCG	Convergent primers for RT-PCR detection of circ48 incDNA or gDNA	<i>H. mediterranei</i>
circ26-con-r	AACGCACAGTTCATCAAGCAA		
circ48-con-f	CATGATTTTCGTTGCTGCAC	Convergent primers for RT-PCR detection of circ48 incDNA or gDNA	
circ48-con-r	CTTGTGCACCGTGAGGGA		
circ53-con-f	ACGAGCGAGACCCGCACTTC	Convergent primers for RT-PCR detection of circ53 incDNA or gDNA	
circ53-con-r	CGATCAAGTTTAGGAGATTA		
circ157-con-f	AATAAGAGCTGGGCAAGACC	Convergent primers for RT-PCR detection of circ157 incDNA or gDNA	
circ157-con-r	GGTTCGAGACCCTAGCTTTC		
circ183-con-f	GCCTTCCGGTCCCAAGCCAC	Convergent primers for RT-PCR detection of circ183incDNA or gDNA	
circ183-con-r	TCGAGGAATAAGAGCTGGGC		
circ9-con-f	GATGAAGGACGTGCCAAGCT	Convergent primers for RT-PCR detection of circ9 incDNA or gDNA	<i>H. hispanica</i>
circ9-con-r	GCTTCGGTTTGGGCTGTATC		
circ56-con-f	TTCTAGTTGCCAGCAATACC	Convergent primers for RT-PCR detection of circ56 incDNA or gDNA	
circ56-con-r	ACTACCGAATCCAGCTTCA		
circ135-con-f	ACACCTGTCAAACCTCAAAC	Convergent primers for RT-PCR detection of circ135 incDNA or gDNA	
circ135-con-r	TCTTAGAGGGTAGCTGCTTC		
circ156-con-f	TTACTACTGCACGACAGTGCG	Convergent primers for RT-PCR detection of circ156 incDNA or gDNA	
circ156-con-r	ACAGTTTCCGCTGGACGCCG		
circ161-con-f	GGAACAGACGGACTACACCG	Convergent primers for RT-PCR detection of circ161 incDNA or gDNA	
circ161-con-r	GTCGATGCCCGGTCAGCGG		

2 结果与分析

2.1 嗜盐古菌中高可信度 circRNA

之前有研究报道,在 *Sulfolobus solfataricus* P2 (硫磺矿硫化叶菌)和 *Sulfolobus acidocaldarius* (嗜酸嗜热古菌)中 circRNA 存在于 rRNA 和蛋白编码区域^[12]. 同时作为富集 circRNA 的有效手段之一, RNase R 可以在 3'—5' 的方向上降解线性 RNA 分子,但完整保留 circRNA 分子和某些结构化 RNA^[19]. 为了获得这两种极端嗜盐古菌更高可信度的 circRNA 数据,我们采用 RNase R 对 total RNA 进行处理. 对经过 RNase R 消化和未消化处理的 RNA 样品测序,比较两组数据获得最终的 circRNA 数据. 其中地中海富盐菌中有 58% 的 circRNA 来源于 rRNA 区域,剩下 42% 的 circRNA 来源于非 rRNA 区域(见图 1(a)),西班牙盐盒菌中有 81% 的 circRNA 来源于 rRNA 区域,剩下 19% 的 circRNA 来源于非 rRNA 区域(见图 1(b)). 这一结果与在 *Sulfolobus solfataricus* P2 (硫磺矿硫化叶菌)和 *Sulfolobus acidocaldarius* (嗜酸嗜热古菌)中发现的 circRNA 的来源相似^[12],均来源于 rRNA 和蛋白编码基因^[12]. 比较 RNase R 消化和未消化处理的 RNA-seq 中的 circRNA,我们寻找在两个测序数据中均存在的 circRNA,最终确定可信度较高的 67 个地中海富盐菌的 circRNAs 和 133 个西班牙盐盒菌的 circRNAs.

2.2 嗜盐古菌 circRNA 来源组成

在真核细胞中, circRNA 大部分来源于基因组中蛋白编码区^[4]. 为了探索嗜盐古菌 circRNA 在基因组中的来源组成,我们对测到的在地中海富盐菌、西班牙盐盒菌中具有较高可信度的 circRNA 在基因组中的分布情况进行归纳. 结果显示,地中海富盐菌中的 67 个 circRNAs 主要分布在 16S rRNA, 23S rRNA, 跨 16S rRNA-23S rRNA 区以及蛋白编码区. 其中有 40% 分布于 16S rRNA 上,有 25% 分布于蛋白编码区(见图 1(a)). 与前者来源相似,西班牙盐盒菌中的 133 个高可信度 circRNAs 主要分布在 16S rRNA, 23S rRNA, 跨 16S rRNA-23S rRNA 区以及蛋白编码区,但较多分布在跨 16S rRNA-23S rRNA 区域,占总数的 49% (见图 1(b)). 说明在古菌中 circRNA 主要来源于基因组 rRNA 区域,这也与之前报道过的在古菌中 circRNA 的来源组成研究结果一致^[10].

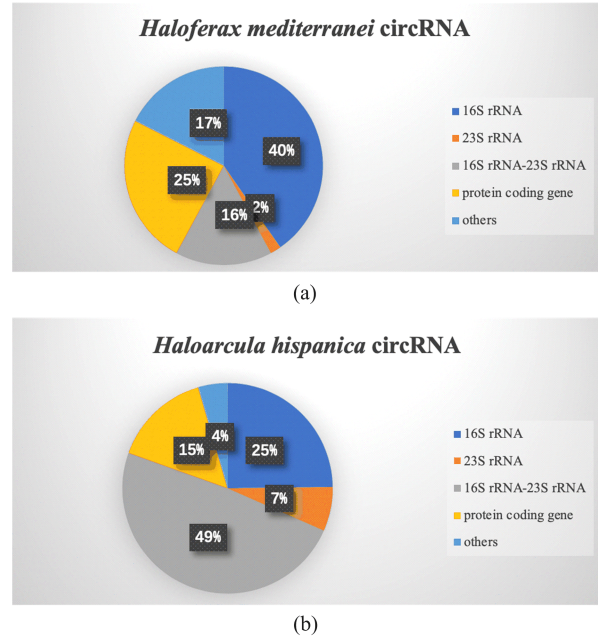


图 1 地中海富盐菌、西班牙盐盒菌中高可信度 circRNA 的来源组成

Fig. 1 The composition of circRNA of high-reliability circRNA in *Haloferax mediterranei* and *Haloarcula hispanica*

2.3 嗜盐古菌中 circRNA 的染色体定位

真核生物中,大量的 circRNA 定位于蛋白编码区,甚至有的 circRNA 可以调控其母基因的转录^[20-21]. 我们同样也探索了这两种嗜盐古菌中存在的 circRNA 的功能性定位. 在地中海富盐菌中,有 14 个 circRNAs 定位在蛋白编码区,占总数的 25%;在西班牙盐盒菌中,有 20 个 circRNAs 定位在蛋白编码区,占总数的 15% (如图 1 所示). 这说明 circRNAs 在以上两种古菌中不仅是在 rRNA 和 tRNA 上成环,在编码区上也存在成环现象.

古菌的基因组由染色体和质粒组成,且含有多个复制起点,古菌的染色体上富含古菌所必需的基因,其编码的产物是古菌生存和进行一系列的活动所必需的物质^[16]. 古菌中的质粒为游离的独立环状 DNA 片段,相对于染色体,质粒携带较少的遗传信息,并且通过古菌相互物理接触时进行细胞间传递^[22].

地中海富盐菌有 1 条染色体,3 个大质粒 (pHM100、pHM300、pHM500),其中 pHM300 质粒上有 287 个开放阅读框及 1 个 tRNA^{Ser} 基因. 涉及 DNA 复制、重组、转录、翻译等过程,以及信号转导、大分子运输、代谢等生命过程^[16]. 我们测到的 67 个 circRNAs 中,有 98.5% 的定位在染色体上,在 pHM300 质粒上只有 1 个,而 pHM100、pHM500

两个质粒上没有检测到 circRNA(如表 2 所示).

西班牙盐盒菌有 2 条染色体,1 条主染色体,1 条小染色体,1 个质粒(pHH406). 其中定位到主染色体上的 circRNAs 有 130 个,占 97.7%,小染色体

上的 circRNAs 有 3 个,而在 pHH406 质粒上并没有测到 circRNA(如表 2 所示).

这说明极端嗜盐古菌 circRNA 产生的主要位置在其染色体上.

表 2 地中海富盐菌、西班牙盐盒菌中 circRNA 在染色体上的分布

Tab. 2 CircRNA distribution on chromosomes in *H. mediterranei* and *H. hispanica*

archaea	location	circRNA	percentage (%)	reads	percentage (%)
<i>H. mediterranei</i>	chromosome	66	98.5	93	98.9
	plasmid: pHM100	0	0	0	0
	plasmid: pHM300	1	1.5	1	1.1
	plasmid: pHM500	0	0	0	0
	total	67	100	94	100
<i>H. hispanica</i>	Chromosome I	130	97.7	162	98.2
	Chromosome II	3	2.3	3	1.8
	plasmid: pHH406	0	0	0	0
	total	133	100	165	100

2.4 嗜盐古菌 circRNA 特点

有研究表明,在真核细胞中 circRNA 的出核效率与其长度相关^[23-24]. 此外,circRNA 外显子的长度也与其成环效率相关^[25]. 于是我们将两种古菌中的 circRNA 的长度进行了探索和统计归纳. 在地中海富盐菌测到的 67 个 circRNAs 中,有 16 个 circRNAs 长度小于 200 nt,有 17 个 circRNAs 长度在 200~400 nt 之间,占所测到 circRNA 总数的一半(如图 2(a)所示). 在西班牙盐盒菌中测到的 133 个 circRNAs 中,有 59 个 circRNAs 长度在 2000~5000 nt 之间,占总数的 44%(如图 2(b)所示).

为了加深对 circRNA 结构的分析,我们使用一

种新的分析方法——FUCHS,对同一基因上产生的 circRNA 进行了亚型分类. 这一方法通过考虑大于 150 bp 的所有 RNA 序列信息来扩展对环形 RNA 的特征鉴定,以表征同种宿主基因产生的不同 circRNA^[26]. 如图 3 所示,我们以 overlap, same end, same start, within, single 五个维度来描述,并发现同一基因的 circRNAs 主要采用 same start 的方式成环,而以 same end 的方式成环的较少. 编码蛋白基因的成环方式大多为 single 成环,同时我们未发现同一个编码基因上存在 2 个及以上的 circRNAs.

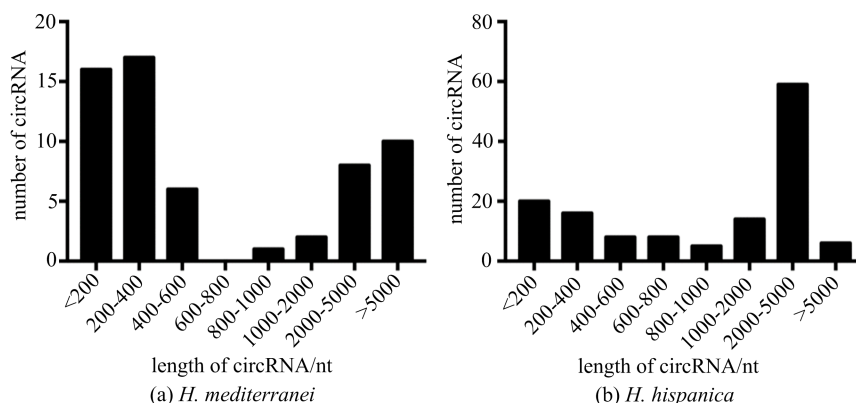


图 2 circRNAs 长度分布

Fig. 2 The length of circRNAs

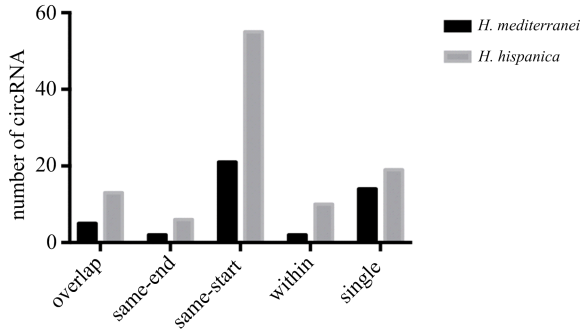


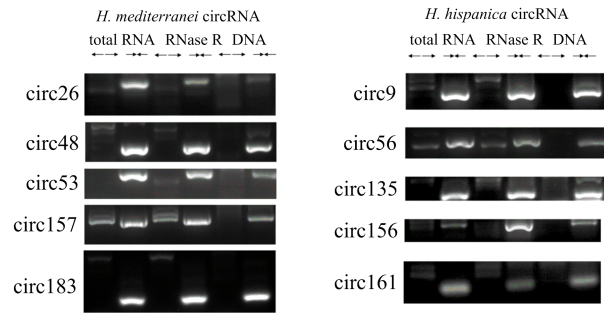
图 3 同一基因 circRNA 不同亚型

Fig. 3 Categories of isoforms from one gene

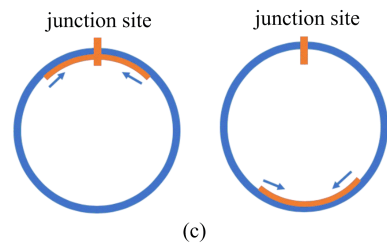
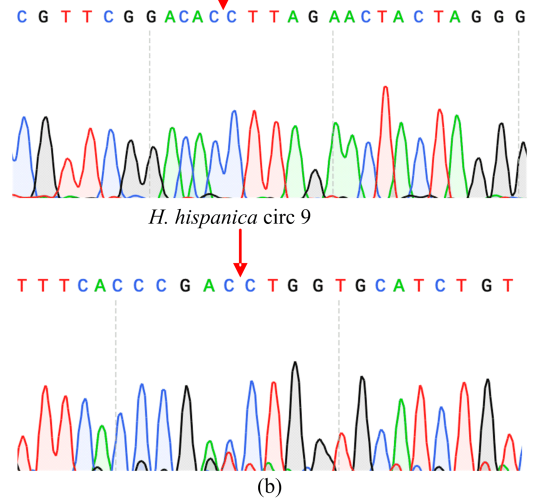
2.5 嗜盐古菌 circRNA 的验证

因为 find-circ^[17] 算法检测 circRNA 是通过 RNA-seq 数据中的 junction reads 来判断的, 所以, 我们只能判断 circRNA 的 junction 位点在基因组的位置, 也就是 circRNA 的起始和终止位置. 而对于 circRNA 的内部具体的序列, 算法并不完全可靠. 此外, 在前面的步骤中, 将在 reads 中的顺序与比对到基因组上的顺序相反的 reads 作为下游的候选, 这可能会导致通过反式剪切而来的线性 RNA 也被判定为 circRNA.

因此, 为了更加精确地表明 circRNA 的存在, 我们在地中海富盐菌和西班牙盐盒菌中各随机选取了来源于 16S rRNA, 23S rRNA, 跨 16S rRNA-23S rRNA 以及蛋白编码基因上的 5 个 circRNAs 进行验证. 根据其在基因组上的序列以及 junction 的位点, 设计了 divergent 和 convergent 引物, 以地中海富盐菌和西班牙盐盒菌的 RNase R 消化/未消化的 total RNA 反转录产物为模板进行验证, 以 gDNA 为模板进行 PCR 作为负参照. 结果显示, 在地中海富盐菌和西班牙盐盒菌中均成功验证上述选择的 circRNAs (如图 4(a) 所示), 分别分布于 16S rRNA, 23S rRNA, 跨 16S rRNA-23S rRNA 区以及蛋白编码基因 (如表 3 所示), 进一步证明了上述测序数据以及分析结果的可靠性. 我们列举出了地中海富盐菌和西班牙盐盒菌中丰度较高的 circRNA 的测序序列信息, 便于之后对其功能的研究 (如图 4(b) 所示).



(a) *H. mediterranei* circ 53



(a) 地中海富盐菌和西班牙盐盒菌中验证 circRNA; Total RNA、RNase R 消化的 RNA、gDNA 为负参照模板进行 RT-PCR, RT-PCR 扩增结果, 以 divergent (背对背箭头) 为引物, 可扩增 circRNA; 以 convergent (头对头箭头为引物, 可扩增线性和 circRNA. (b) 地中海富盐菌 circ53 和西班牙盐盒菌 circ9 测序序列信息, 箭头位置为 junction 位置. (c) 背对背箭头为 divergent 引物, 头对头箭头为 convergent 引物, 蓝色圆圈表示 RNA 模板, 黄色线表示预期的 RT-PCR 产物.

图 4 RT-PCR 验证地中海富盐菌、西班牙盐盒菌中 circRNA
Fig. 4 Verification of circRNAs by RT-PCR in *H. mediterranei* and *H. hispanica*

表 3 RT-PCR 验证地中海富盐菌、西班牙盐盒菌中 circRNAs
 Tab. 3 Verification of circRNAs by RT-PCR in *H. mediterranei* and *H. hispanica*

archaea	location	circRNA	class	circle start	circle end
<i>H. mediterranei</i>	chromosome	circ26	HFX-RS14005	2764334	2764905
		circ48	23S rRNA	2838409	2838708
		circ53	16S rRNA-23S rRNA	1746408	1748503
		circ157	16S rRNA	2841000	2841330
		circ183	16S rRNA	2841019	2841299
<i>H. hispanica</i>	Chromosome I	circ9	16S rRNA-23S rRNA	213847	216847
		circ56	16S rRNA-23S rRNA	213477	217043
		circ135	23S rRNA	215685	216847
		circ156	16S rRNA	213406	213923
		circ161	HAH-RS09825	1962080	1962206

3 讨论

circRNA 目前已经在多个物种中被发现和研 究,在真核生物中,已经有 2 万多种不同的 circRNA 被鉴定^[27],涉及人类^[28-30]、小鼠^[28,30]、线虫^[28,31]、斑马鱼^[32]、果蝇^[33]和植物^[34]等多个物种,此外, circRNA 还具有进化保守序列,其中有 457 个人类 circRNAs 可以映射到小鼠的基因上,其中 69 个包含同源的反向剪接位点^[35]. circRNA 还具有组织特异性,例如小鼠中的 Rmsh 和 Klhl2 的 circRNA 在大脑中高表达,而在肝脏和肺中不表达^[36];果蝇中 circ-Mbl 表达在卵巢中远低于头部^[37]. 本课题组之前发现了一种新型 circRNA-EIciRNA^[18],其通过与 U1 snRNP 相互作用,顺式调控亲本基因的转录. 目前对 circRNA 的研究主要集中在真核生物领域,而对古菌中 circRNA 的研究相对较少. 古菌在遗传背景上与真核生物更为接近,所以研究古菌中的 circRNA,既扩大了对 circRNA 在不同物种领域的认知,也对真核生物中的 circRNA 的存在及功能的研究有了更广的知识扩充. 在之前针对古菌 circRNA 的研究中,在嗜超高温古菌 (*Sulfolobus solfataricus*) 和嗜热碱甲烷杆菌 (*Pyrodicticum abyssi*) 两种古菌中发现了 circRNA,其主要产生于 RNA^{Tri} 的内含子的环化,并且已经确定了精确的 junction 位点^[38].

我们在地中海富盐菌中检测到 67 个 circRNAs,在西班牙盐盒菌中检测到 133 个 circRNAs. 我们在这两种古菌中选择了丰度较高的 5 个 circRNA 进

行 RT-PCR,验证了这些 circRNAs 的存在. 相较于之前在其他古菌中发现的 circRNA 主要来源于 rRNA 或 tRNA^[38],我们在两种嗜盐古菌中发现了来源于蛋白编码基因上的 circRNA. 这一结果,引起对这类 circRNA 是否如同真核细胞一样,也参与相应的顺式/反式转录调控过程的猜想. 与之前的研究^[10]不同的是,我们对两种古菌中 circRNA 在染色体和质粒上的分布以及同一个基因上产生的 circRNA 的类型进行了统计分析. 发现在地中海富盐菌中 circRNA 主要定位于其染色体上,在其质粒 (pHM300) 上也存在少量 circRNA. 在西班牙盐盒菌中, circRNA 主要定位于其主染色体上,在其小染色体上也有但仅占 2.3%,而在其质粒上未检测到 circRNA. 这说明 circRNA 产生的主要位置在其染色体上. 而两种古菌中同一基因产生的 circRNA 主要以 same start 的方式成环,而同一个蛋白编码基因上未发现存在 2 个及以上的 circRNAs.

随着测序技术的发展,越来越多的古菌的基因组被鉴定. 最近有研究表明,古细菌中 DNA 簇存在“离散区室化”排列方式,其表现形式与真核生物,特别是人类细胞非常相似^[12],若发生错误的折叠或捆绑,会导致错误的基因,尤其是致癌/抑癌基因的打开/关闭,最终可能引起癌症的发生. 此外,古菌相对于真核细胞更加简单,以古菌为模型,有利于探索人类疾病的发生发展^[12]. 本文中我们首次在极端嗜盐古菌中研究 circRNA,不仅对于研究古菌的遗传进化具有重要的意义,而且对于探究其生物学功能,尤其是对真核生物相关研究有着很大的启示作用. 本

研究通过 RNA-Seq、生物信息学和分子生物学的手段,首次在极端嗜盐古菌地中海富盐菌和西班牙盐盒菌中验证 circRNA,并对其来源组成,染色体定位以及长度、成环特征进行分析,对古菌中 circRNA 的存在、分子表征及功能的研究奠定了重要的基础,分子表征和功能将是后续研究的内容,并且丰富了我们对物种间 circRNA 特点的认知,有利于扩展从 circRNA 角度思考生命进化及探索真核细胞的进化产生。

参考文献(References)

- [1] SANGER H L, KLOTZ G, RIESNER D, et al. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rodlike structures[J]. Proc Natl Acad Sci USA, 1976, 73: 3852-3856.
- [2] GUO J U, AGARWAL V, GUO H, et al. Expanded identification and characterization of mammalian circular RNAs[J]. Genome Biology, 2014, 15(7): Article number 409.
- [3] LASDA E, PARKER R. Circular RNAs; Diversity of form and function[J]. RNA, 2014, 20(12): 1829-1842.
- [4] DU W W, ZHANG C, YANG W, et al. Identifying and characterizing circRNA-protein interaction [J]. Theranostics, 2017, 7(17): 4183-4191.
- [5] HOLDT L M, KOHLMAIER A, TEUPSER D. Molecular roles and function of circular RNAs in eukaryotic cells [J]. Cellular and Molecular Life Sciences, 2018, 75(6): 1071-1098.
- [6] LI F, ZHANG L, LI W, et al. Circular RNA ITCH has inhibitory effect on ESCC by suppressing the Wnt/ β -catenin pathway[J]. Oncotarget, 2015, 6: 6001-13.
- [7] ASHWAL-FLUSS R, MEYER M, PAMUDURTI N R, et al. circRNA biogenesis competes with pre-mRNA splicing[J]. Mol Cell, 2014, 56: 55-66.
- [8] CHEN Y, LI C, TAN C, et al. Circular RNAs: A new frontier in the study of human diseases[J]. Journal of Medical Genetics, 2016, 53(6): 359-365.
- [9] TAKEMATA N, SAMSON R Y, BELL S D. Physical and functional compartmentalization of archaeal chromosomes[J]. Cell, 2019, 179(1): 165-179.
- [10] DANAN M, SCHWARTZ S, EDELHEIT S, et al. Transcriptome-wide discovery of circular RNAs in Archaea[J]. Nucleic Acids Res, 2011, 40: 3131-3142.
- [11] LYKKE-ANDERSEN J, AAGAARD C, SEMIONENKOV M, et al. Archaeal introns: Splicing, intercellular mobility and evolution[J]. Trends Biochem Sci, 1997, 22: 326-331.
- [12] SINGH S K, GURHA P, TRAN E J, et al. Sequential 2'-O-methylation of archaeal pre-tRNA^{Trp} nucleotides is guided by the intron-encoded but *trans*-acting box C/D ribonucleoprotein of pre-tRNA[J]. J Biol Chem, 2004, 279: 47661-47671.
- [13] CLOUET D'ORVAL B, BORTOLIN M L, GASPIN C, et al. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp} [J]. Nucleic Acids Res, 2001, 29: 4518-4529.
- [14] SALGIA S R, SINGH S K, GURHA P, et al. Two reactions of Haloferax volcanii RNA splicing enzymes: Joining of exons and circularization of introns [J]. RNA, 2003, 9: 319-330.
- [15] TANG T H, ROZHDESTVENSKY T S, D'ORVAL B C, et al. RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing[J]. Nucleic Acids Res, 2002, 30: 921-930.
- [16] MIAO Di. Construction of a novel shuttle expression vector based on haloarchaeal plasmid pSCM201 and studies on replicon of minichromosome pHM300[D]. Beijing: Institute of Microbiology, Chinese Academy of Sciences, 2009.
- [17] MEMCZAK S, JENS M, ELEFSINIOTI A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. Nature, 2013, 495(7441): 333-338.
- [18] LI Z, HUANG C, BAO C, et al. Exon-intron circular RNAs regulate transcription in the nucleus[J]. Nat Struct Mol Biol, 2015, 22: 256-264.
- [19] SUZUKI H, ZUO Y, WANG J, et al. Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing[J]. Nucleic Acids Res, 2006, 34: e63.
- [20] MEMCZAK S, JENS M, ELEFSINIOTI A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. Nature, 2013, 495: 333-338.
- [21] CONN S J, PILLMAN K A, TOUBIA J, et al. The RNA binding protein quaking regulates formation of circRNAs[J]. Cell, 2015, 160: 1125-1134.
- [22] SOTA M, TOP E M. Horizontal gene transfer mediated by plasmids [C]// Plasmids; Current Research and Future Trends. Poole, UK: Caister Academic Press, 2008.
- [23] HUANG C, LIANG D, TATOMER D C, et al. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs[J]. Genes

- Dev, 2018, 32: 639-644.
- [24] WANY, HOPPER A K. Size matters: Conserved proteins function in length-dependent nuclear export of circular RNAs[J]. *Genes Dev*, 2018, 32: 600-601.
- [25] BARRETT S P, WANG P L, SALZMAN J. Circular RNA biogenesis can proceed through an exon-containing lariat precursor [J]. *eLife*, 2015, 4: e07540.
- [26] METGE F, CZAJA-HASSE L F, REINHARDT R, et al. FUCHS-towards full circular RNA characterization using RNAsEq[J]. *PeerJ*, 2017, 5: e2934.
- [27] GLAZAR P, PAPAVALSILEIOU P, RAJEWSKY N. circBase: A database for circular RNAs [J]. *RNA*, 2014, 20:1666-1670.
- [28] MEMCZAK S, JENS M, ELEFSINIOTI A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency[J]. *Nature*, 2013, 495: 333-338.
- [29] RYBAK-WOLF A, STOTTMEISTER C, GLAZAR P, et al. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed[J]. *Mol Cell*, 2015, 58: 870-885.
- [30] SALZMAN J, CHEN R E, OLSEN M N, et al. Cell-type specific features of circular RNA expression[J]. *PLoS Genet*, 2013, 9: e1003777.
- [31] IVANOV A, MEMCZAK S, WYLER E, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals [J]. *Cell Rep*, 2015, 10:170-177.
- [32] SHEN Y, GUO X, WANG W. Identification and characterization of circular RNAs in zebrafish [J]. *FEBS Lett*, 2017, 591: 213-220.
- [33] HOUSELEY J M, GARCIA-CASADO Z, PASCUAL M, et al. Noncanonical RNAs from transcripts of the *Drosophila* muscleblind gene[J]. *J Hered*, 2006, 97: 253-260.
- [34] WANG P L, BAO Y, YEE M C, et al. Circular RNA is expressed across the eukaryotic tree of life[J]. *PLoS One*, 2014, 9: e90859.
- [35] JECK W R, SORRENTINO J A, WANG K, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats[J]. *RNA*, 2013, 19: 141-157.
- [36] WESTHOLM J O, MIURA P, OLSON S, et al. Genome-wide analysis of *Drosophila* circular RNAs reveals their structural and sequence properties and age-dependent neural accumulation [J]. *Cell Rep*, 2014, 9: 1966-1980.
- [37] YOU X, VLATKOVIC I, BABIC A, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity [J]. *Nat Neurosci*, 2015, 18: 603-610.
- [38] CLOUET D'ORVAL B, BORTOLIN M L, GASPIN C, et al. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA^{Trp} intron guides the formation of two ribose-methylated nucleosides in the mature tRNA^{Trp} [J]. *Nucleic Acids Res*, 2001, 29(22): 4518-4529.