



## 地中海富盐菌中非编码 RNA 的鉴定与分析

王磊<sup>1#</sup>, 程飞跃<sup>1,2#</sup>, 赵大贺<sup>1,2</sup>, 杨海波<sup>1,2</sup>, 向华<sup>1,2\*</sup>

<sup>1</sup> 中国科学院微生物研究所, 微生物资源前期开发国家重点实验室, 北京 100101

<sup>2</sup> 中国科学院大学, 北京 100049

**摘要:**【目的】通过转录组高通量测序技术(即 RNA-seq), 结合生物信息学分析和分子生物学方法, 在组学水平鉴定极端嗜盐菌中可能的非编码 RNA (ncRNA)。【方法】将培养至对数中期的地中海富盐菌在不同盐浓度下处理 30 分钟, 提取 RNA, 进行链特异的转录组测序和 5'端区分的转录组测序, 通过生物信息学分析在全基因组范围内鉴定 ncRNA, 预测其转录边界; 然后通过 Northern blot 和环化 RNA 反转录聚合酶链式反应(CR-RT-PCR)对部分预测的 ncRNA 进行实验验证。【结果】比较两种 RNA-seq 技术在不同培养条件下的 RNA 测序结果和对转录单元的精细分析, 共鉴定到 105 个高可信度的 ncRNA, 并发现 4 个在不同盐度下表达差异较大的 ncRNA, 通过 Northern blot 和 CR-RT-PCR 验证了 incRNA1436 和 incRNA1903 的表达情况、转录本、转录起始位点及终止位点等。【结论】首次在组学水平鉴定了地中海富盐菌中的 ncRNA, 不同盐浓度刺激下部分 ncRNA 的转录差异暗示其有可能参与地中海富盐菌对盐胁迫的适应, 高可信度 ncRNA 的组学发现为今后全面开展嗜盐古菌 ncRNA 的功能机制研究提供了基础数据及重要的切入点。

**关键词:** 地中海富盐菌, 非编码 RNA, 盐度刺激, 转录组测序

非编码 RNA (non-coding RNA, ncRNA)是指生物体内各种不参与蛋白质编码的 RNA。ncRNA 在细菌、古菌和真核生物中均广泛存在, 发挥着重要的生理功能<sup>[1-2]</sup>。目前在不同生物中进行 ncRNA 的系统发现仍然是极其重要的基础性工作。随着测序技术和生物信息学分析方法的进步, 鉴定 ncRNA 的技术也在不断发展。以第二代高通

量测序为基础的 RNA-seq 技术<sup>[3]</sup>大大促进了转录组复杂性研究, 并为 ncRNA 的鉴定提供了新方法。其中链特异性 RNA 测序(strand-specific RNA sequencing, ssRNA-seq)技术<sup>[4]</sup>可将 mRNA 链的方向信息保存到测序文库中, 测序后的数据分析可确定转录本是来自正义还是反义 DNA 链。与普通转录组测序相比, 它更能准确地统计转录本的数

基金项目: 国家自然科学基金青年项目(31301071)

\*通信作者。Tel/Fax: +86-10-64807472; E-mail: xiangh@im.ac.cn

#并列第一作者。

收稿日期: 2016-06-15; 修回日期: 2016-07-15; 网络出版日期: 2016-07-28

量和确定基因的结构,同时可以发现更多的反义转录本,成为当前发现基因组范围内 ncRNA 的重要方法。dRNA-seq (differential RNA sequencing)技术<sup>[5-6]</sup>是一种能够区分初始转录本和加工过的转录本的新的 RNA-seq 技术。通过 5'单磷酸依赖的外切酶 (Terminator 5'-phosphate-dependent exonuclease, TEX)能够特异性的降解加工过的转录本,通过比较经 TEX 处理的加工转录本库和不经 TEX 处理的初始转录本库,可以在全基因组水平上对转录起始位点(TSS)、启动子区域、5'-UTR 和操纵子的注释等方面进行更精确的研究。

古菌是一类与细菌和真核生物并列的重要生命类群,古菌具有基因组简单和遗传机器为真核类型等特点。但是与真核生物和细菌相比,古菌 ncRNA 的研究才刚刚起步,迄今为止只在有限的几株古菌开展了 ncRNA 的组学分析,包括闪烁古生球菌(*Archaeoglobus fulgidus*)<sup>[7]</sup>、硫矿硫化叶菌(*Sulfolobus solfataricus*)<sup>[8]</sup>、詹氏甲烷球菌(*Methanocaldococcus jannaschii*)<sup>[9]</sup>、马氏甲烷八叠球菌(*Methanosarcina mazei*)<sup>[10]</sup>、强烈炽热球菌(*Pyrococcus furiosus*)<sup>[11]</sup>和沃氏富盐菌(*Haloferax volcanii*)<sup>[12-13]</sup>等。其中在硫矿硫化叶菌的 RNA-seq 研究中,研究人员找到了 310 个 ncRNA,包括 125 个反式 ncRNA 和 185 个顺式 ncRNA<sup>[8]</sup>;Heyer 等<sup>[14]</sup>对沃氏富盐菌的 RNA-seq 分析研究发现了 190 个 ncRNA,并鉴定到转录本中存在大量 tRNA 衍生片段,且这些 tRNA 衍生片段在不同生理条件下表达量存在较大差异。对古菌转录组的深度测序和相关初步研究表明,古菌 ncRNA 类型丰富,功能多样,既具有真核类型的 small nucleolar RNA (snoRNA)参与核酸修饰<sup>[15]</sup>,又具有原核生物特有的 CRISPR (clustered regularly interspaced short palindromic repeats) RNA 参与抗病毒及基因组稳

定性维持<sup>[16-17]</sup>,更多的 ncRNA 可能参与环境响应和基因表达调控。因此开展古菌 ncRNA 的系统研究,对于理解 ncRNA 的生理功能、分子机制以及相关机制的进化具有重要意义。

地中海富盐菌(*Haloferax mediterranei*)是极端嗜盐古菌的重要代表,相比其他古菌类群更容易培养和进行遗传操作,全基因组测序已经完成<sup>[18]</sup>,因此是研究古菌 ncRNA 的理想材料。ssRNA-seq 可以区分来自基因组正义链或反义链上的 reads,更加精确检测转录位置和方向;dRNA-seq 则能够精确分辨初始转录本,在全基因组水平上鉴定转录起始位点。本研究通过综合利用 ssRNA-seq 和 dRNA-seq 技术,结合生物信息学分析鉴定了地中海富盐菌中 ncRNA,并选取典型 ncRNA,通过 Northern blot 和 CR-RT-PCR(Circularized RNA reverse transcription-polymerase chain reaction)<sup>[19]</sup>对其验证。本工作为地中海富盐菌 ncRNA 功能机制的研究奠定了基础。

## 1 材料和方法

### 1.1 材料

**1.1.1 菌株及培养条件:**本研究所用的极端嗜盐古菌为地中海富盐菌 *H. mediterranei* ATCC33500,采用 AS-168 培养基于 37 °C、200 r/min 的摇床培养<sup>[20]</sup>。其中 NaCl 浓度根据需要分别调整至 10%、15%和 20%。分子克隆所用的菌株为 *E. coli* JM109,采用 LB 培养基常规培养<sup>[21]</sup>。

**1.1.2 主要试剂和仪器:**提取 RNA 的 TRIzol 试剂购自 Invitrogen 公司,T4 DNA 连接酶购自 New England Biolabs,酵母抽提物(Yeast extract)为 Oxoid 原装,酸水解酪素(Casamino acids)为 Difco 原装,PCR 反应中的 *Taq* DNA 聚合酶为上海生工

产品, M-MLV 反转录酶、pGEM-T Easy 载体购自 Promega, 其他常见生化试剂全部购自北京化学试剂公司。RNA-seq 相关试剂由相关测序公司提供。实验过程还用到低温高速离心机 (Eppendorf Centrifuge 5810 R), Nanodrop 1000 分光光度计 (Thermo Fisher Scientific) 等仪器。

## 1.2 极端嗜盐古菌的盐刺激和总 RNA 的提取

将 *H. mediterranei* 在 15% NaCl-AS-168 培养基中培养至稳定前期的种子液, 以 4 : 100 转接至 100 mL 相同培养基中扩大培养至对数后期 (约 2 d), 分别取 10 mL 菌液加到 40 mL 新鲜的不同 NaCl 浓度的培养基中, 使 NaCl 终浓度分别为 10%、15% 和 20%, 培养 30 min, 取 5 mL 菌体离心, 收集菌体, 用 TRIzol 试剂 (Invitrogen 公司) 提取总 RNA。取 1  $\mu$ L RNA 样品在 Nanodrop 1000 分光光度计上测定,  $OD_{260}/OD_{280}$  在 2.0 左右。

## 1.3 转录组测序

将上述不同盐浓度 (各设 3 个重复) 刺激下的 RNA 进行 RNA-seq。其中 ssRNA-seq 具体实验流程为: 使用 Ribo Zero kit 去除 rRNA, 将 RNA 打断为大约 350 nt 长度的片段, 使用 dUTP 法进行链特异性建库<sup>[4]</sup>, 然后采用 Hiseq 进行测序 (在北京基因组研究所进行测序)。而 dRNA-seq 测序实验流程为: 使用 Ribo Zero kit 来去除 rRNA, 然后进行 TEX 处理 (去除 5' 单磷酸 mRNA) 和 TAP 处理 (将 5' 三磷酸 mRNA 处理为 5' 单磷酸 mRNA), 连接 5' 末端接头, 反转录合成 cDNA 第一链, PCR 富集, 使用 Hiseq2000 测序仪进行测序 (由诺和致源生物信息科技有限公司完成)。

## 1.4 转录组数据处理

对于得到的 ssRNA-seq 原始数据, 首先去除 5' 和 3' 接头序列, 如果剩余序列长度小于原来长度

的 1/3, 则去除; 去除 N (N 表示无法确定碱基信息) 的比例大于 10% 的序列; 去除低质量序列 (质量值  $Q \leq 5$  的碱基数占整个读数的 50% 以上), 获得 clean reads。然后去掉重复序列, 获得唯一序列。最后用 Bowtie2<sup>[22]</sup> 将唯一序列映射至参考基因组。对于 dRNA-seq 原始读取序列, 处理方法同上, 用 Bowtie2 将 clean reads 映射至参考基因组。

## 1.5 转录边界预测<sup>[23]</sup>

基于 ssRNA-seq 数据, 首先通过 Rockhopper 软件<sup>[24]</sup> 来初步预测基因组中存在的操纵子 (同向表达基因如果距离 < 40 bp, 就默认为是一个操纵子)。然后根据 reads 分布, 来确定转录单元的转录起始位点 (transcription start site, TSS) (我们把起始密码子上游 100 bp 内 reads 数变化最大的位置预测为 TSS) 和转录终止位点 (transcription termination site, TTS) (终止密码子向下延伸, reads 数大幅下降处为转录终止位点; 表达量低于终止密码子处 reads 数 1/3 或者 reads 数 < 3, 也认为其为终止位点)。获得 TSS 和 TTS 位置信息后, 编写脚本寻找先前转录单元之间是否存在通读现象, 确定最后转录单元。

## 1.6 ncRNA 的预测

基于先前预测转录单元, 首先寻找反义 ncRNA: 寻找每个转录单元负链上 reads 数最高的点位, 确定其 TSS 和 TTS。然后以同样的方法寻找基因间区 ncRNA。

## 1.7 Northern blot 验证

选取部分预测的 ncRNA, 用 Northern blot 进行验证, 本实验中所用的探针见表 1。实验流程如下: 将 RNA 样品进行聚丙烯酰胺变性胶电泳, 移至半干胶转膜仪 (Trans-Blot SD Cell, Bio-Rad) 以 12 V/100 cm<sup>2</sup> 恒压转膜 45 min, 紫外照射使 RNA 与

表 1. 本文中使用的引物

Table 1. The primers used in this study

Primers	Sequences (5'→3')
Northern blot <sup>a</sup>	
incRNA1903-probe	ATCGGGCACGGGAAGTGGGC
incRNA1436-probe	GACGAGTGACCCGGGCAGGTC ACTG
CR-RT-PCR	
incRNA1903-F	GGAACGGCGAATCCTCAC
incRNA1903-R	CTCCCCTTACGGGTCGGT
incRNA1436-F	CGTGTCCCTGCCCAAACC
incRNA1436-R	GGGTTTGGGCAGGGACAC
7S RNA-F	CGGTTTCATCGGTGGTAC
16S RNA-R	GCACTACTCCCGTATGAA

<sup>a</sup>: 5'-end are labeled with biotin.

膜交联, 将尼龙膜装入杂交瓶, 按 20 mL/100 cm<sup>2</sup> 的比例加入预杂交液, 52 °C 预杂交 2 h。弃去预杂交液, 加入杂交液和 2 μL 标记的探针引物 (10 μmol/L)。52 °C 杂交 16 h, 之后进行洗膜。洗膜结束后, 进行化学发光法显色, 在暗室中压片、显影。

### 1.8 环化 RNA 反转录聚合酶链式反应(CR-RT-PCR)

选取 incRNA1903 和 incRNA1436 进行 CR-RT-PCR 实验以测定其精确的转录起始位点 (TSS) 和终止位点 (TTS)<sup>[25]</sup>。方法如下: 取 5–10 μg RNA 样品 65 °C 水浴 10 min 变性, 立即冰浴 2 min。然后进行环化, 环化体系如下: RNA 5–10 μg、10×T4 RNA 连接酶缓冲液 2.5 μL、T4 RNA 连接酶 2 μL、RNase 抑制剂 0.5 μL、DEPC 水 10 μL; 37 °C 水浴 1 h。用 DEPC 水调节反应体系至 500 μL, 加入 200 μL 中酚-氯仿, 充分混匀, 12000 r/min、4 °C 离心 10 min, 弃上清, 用 70% 乙醇洗涤 2 次, 加入 15 μL DEPC 水溶解。向已环化的 RNA 中加入 1 μL 反转录引物 (1 μmol/L), 65 °C 变性 10 min, 立即置于冰上 2 min。按以下

体系添加: RNase 抑制剂 0.5 μL、5×RT 缓冲液 8 μL、dNTPs (10 mmol/L) 1 μL、DEPC 水 13.5 μL, 在 42 °C 水浴预热 2 min 后, 加入 1 μL M-MLV 反转录酶。42 °C 继续水浴 1 h, 然后 75 °C 水浴 15 min。取 2 μL 上述反转录产物为模板, 用位于基因编码区域内部且靠近两端的一对反向引物进行 PCR 扩增。

本实验中所用的 PCR 扩增引物见表 1。常规的 PCR 反应体系 (25 μL) 如下: 模板 1 μL、10×*Taq* 缓冲液 2.5 μL、dNTPs (2 mmol/L) 2.5 μL、*Taq* 聚合酶 0.3 μL、正向引物 (10 μmol/L) 1 μL、反向引物 (10 μmol/L) 1 μL、ddH<sub>2</sub>O 16.7 μL; 反应程序如下: 95 °C 5 min; 95 °C 30 s, 54 °C 30 s, 72 °C 45 s, 40 个循环; 72 °C 10 min。

对于 incRNA1436, 由于预测转录本大小只有 124 nt, 并不利于环化自连, 故将其分别和 7S RNA 与 16S RNA 连接然后进行 PCR 扩增。将得到的 PCR 产物与 pGEM-T Easy 载体连接, 转化 *E. coli* JM109, 挑取单克隆测序, 鉴定发生衔接的位点, 从而确定 5' 和 3' 末端。

## 2 结果和分析

### 2.1 高可信度 ncRNA 的鉴定

通过对 ssRNA-seq 结果的分析, 在 10%、15% 和 20% NaCl 盐浓度刺激条件下, 映射至基因组上的读数分别为 1149872、1041661 和 1088969。通过对 dRNA-seq 结果的分析, 映射至基因组上的读数为 11100543。

基于 ssRNA-seq 数据, 我们首先找到反义链上表达量最高的点, 然后寻找其转录起始位点 TSS 和转录终止位点 TTS, 最终发现 350 个反义 ncRNA (asRNA), 其中 TSS 位置与 dRNA-seq 数

据完全吻合的 ncRNA 个数为 44 个(作为高可信度 ncRNA)。采用类似鉴定反义 ncRNA 的方法, 我们最终发现 216 个基因间区 ncRNA (incRNA), 其中 TSS 位置与 dRNA-seq 数据完全吻合的基因间区 ncRNA 个数为 61 个(为高可信度 ncRNA)。因此共获得高可信度 ncRNA 共计 105 条。

## 2.2 嗜盐古菌 ncRNA 的特点分析

统计发现, 我们鉴定到的这 105 个高可信度的 ncRNA (ssRNA-seq 数据与 dRNA-seq 数据同时支持), 其中 81 个 ncRNA 长度位于 100–299 nt 这个范围内(图 1), 占 77.14%。100–299 nt 范围内, asRNA 有 32 个, 占所预测 asRNA 的 72.73%; incRNA 有 49 个, 占所预测 incRNA 的 80.33%。

计算 RPKM(Reads Per kilobase of transcript per million mapped reads)值, 比较样品间的基因表达差异<sup>[26]</sup>(表 2)。对 105 个 ncRNA 在不同盐浓度条件下的表达差异(以 RPKM 值为参考)分析发现, RPKM 值差异超过 2 倍的 ncRNA 有 45 个, 其中 RPKM 值差异超过 5 倍的有 4 个, 全部都是 incRNA (表 2)。

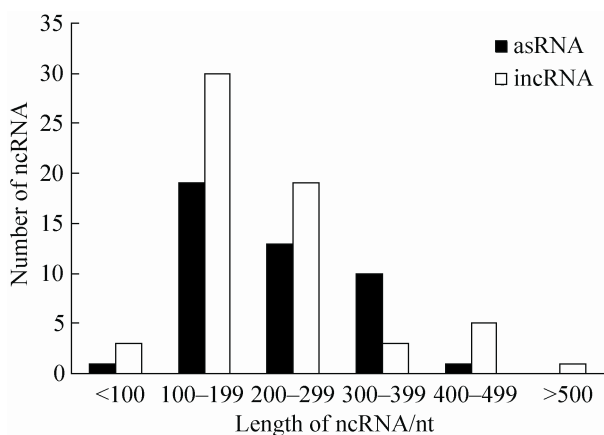


图 1. ncRNA 长度分布

Figure 1. Length distribution of ncRNA in the *Haloferax mediterranei*.

表 2. 表达差异 5 倍以上的 ncRNA

Table 2. The ncRNA with expression level more than five times (at least between 2 treatments)

ncRNA	RPKM (10% NaCl)	RPKM (15% NaCl)	RPKM (20% NaCl)
incRNA0723	8.735249	11.86299	74.1222
incRNA1436	34.85254	231.4789	86.45368
incRNA2271	27.16288	157.687	32.7739
incRNA6338	1.145029	2.073311	8.9866

## 2.3 Northern blot 分析

选择表达量最高的 incRNA1903 和表达差异 5 倍以上的 incRNA1436, 对基于 RNA-seq 预测的结果, 进行 Northern blot 杂交实验验证。结果显示(图 2), incRNA1903 和 incRNA1436 均有表达。incRNA1903 测序预测的大小是 417 nt, 图中显示的约是 450 nt, 与预测的结果基本相符。有趣的是, 结果显示 incRNA1436 有两个信号带, 表明可能有两个转录本。其中, 较大的转录本的大小约是 120 nt, 与预测的 124 nt 大小相似, 小转录本约是 90 nt。总体说明 RNA-seq 发现的 ncRNA 真实存在。

进一步, 我们选取了在不同盐浓度刺激下的 incRNA1436 进行 Northern blot 验证(图 3)。在 15% NaCl 培养基中培养的地中海富盐菌经低盐(10%

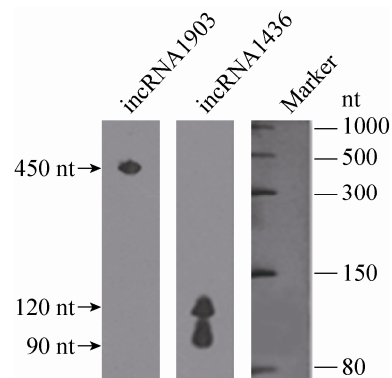


图 2. Northern blot 验证 incRNA1436 和 incRNA1903 的表达

Figure 2. Northern analysis of incRNA1436 and incRNA1903.

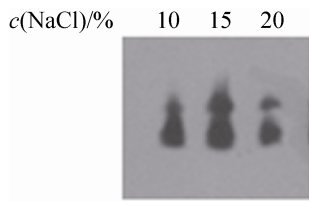


图 3. 不同盐浓度刺激下 *incRNA1436* 的 Northern 实验结果

Figure 3. Northern blot analysis of *incRNA1436* under the stimulation of different salt concentration. Loading equal RNA, duplicating 3 times.

NaCl)和高盐(20% NaCl)环境刺激 30 min 后提取 RNA。结果显示,在 15% NaCl 培养基中生长的菌体, *incRNA1436* 的大、小 2 个转录本表达量较高,当在低盐(10% NaCl)和高盐(20% NaCl)刺激后,大、小转录本的表达量均下降,而大转录本表达量的下降尤为明显,这与 RNA-seq 观察到该 ncRNA 在 15% NaCl 条件下表达量最高的结果一

致,说明该 ncRNA 表达确实受盐度变化的诱导(图 3 为 3 次重复的代表性结果),初步验证了基于 RNA-seq 的结果鉴定 ncRNA 和分析其表达量的可信性。而其大、小 2 个转录本的加工机制及如何参与调控地中海富盐菌对盐胁迫的应答过程则有待进一步的研究。

### 2.4 采用 CR-RT-PCR 分析进一步测定 ncRNA 的起始与终止位点

将 CR-RT-PCR 实验单克隆测序结果与基因组序列比对,可得到 *incRNA1903* 转录本的 TSS 为图中+1 处的 A 碱基(图 4-A),比 RNA-seq 预测的转录本在 5'端长 13 nt;3'末端为图中 449 位处的 T,比 RNA-seq 预测的转录本在 3'端长 19 nt。而对 *incRNA1436* 转录本进行序列比对发现有 2 个转录本存在,大转录本的 TSS 为图中+1 处的 G (图 4-B),

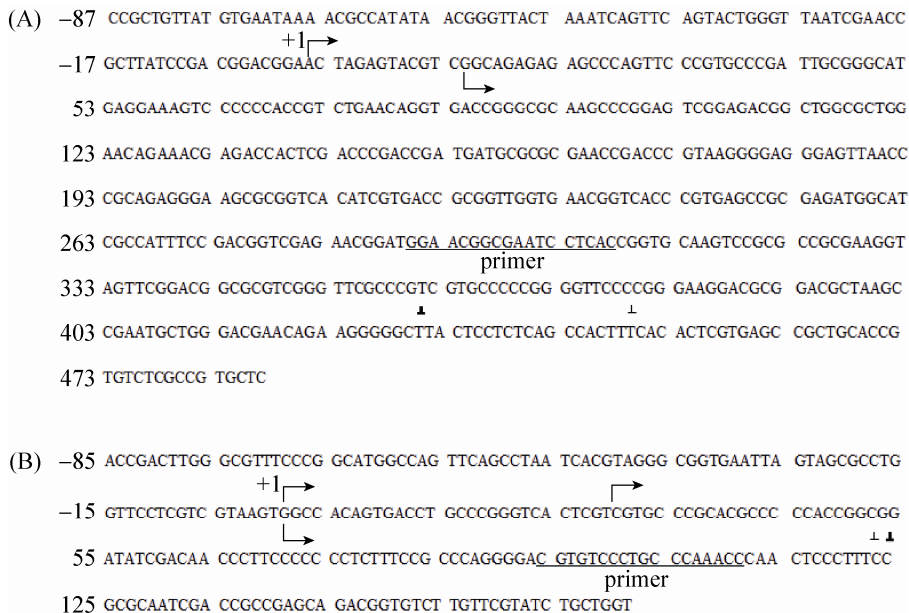


图 4. CR-RT-PCR 测定转录起始位点与终止位点

Figure 4. Sequence of the transcripts of *incRNA1903* and *incRNA1436*. A: identification of 5'- and 3'-ends of *incRNA1903* transcripts. B: identification of 5'- and 3'-ends of *incRNA1436* transcripts. The two transcripts of *incRNA1436* were identified by CR-RT-PCR. Transcription start sites (TSS) are marked with arrows. Transcription termination sites (TTS) are marked with inverted T. The bold ones are identified by RNA-seq and the other ones are identified by CR-RT-PCR. The primers used for RT-PCR are underlined.

与 RNA-seq 预测的转录本的 TSS 是一致的；另一个小转录本 TSS 为图中 30 位处的 C，这两个转录本的 3'末端均为图中 123 位处的 C，比 RNA-seq 预测的转录本在 3'端处少 1 nt。CR-RT-PCR 与 Northern blot 实验关于 *incRNA1436* 有 2 个转录本的结果是一致的。本实验再次表明，RNA-seq 系统发现的 ncRNA 真实存在，但其精确的 5'端和 3'端、以及可能的加工信息等，可能还需通过更加精确的遗传与分子生物学方法进行校准。

### 3 讨论

相对于细菌和真核生物，古菌 ncRNA 的研究比较薄弱，而嗜盐古菌 ncRNA 的鉴定及研究更是甚少。本研究首次通过 RNA-seq 技术在地中海富盐菌中进行高通量转录组测序分析，鉴定其基因组范围内的 ncRNA。本文中综合采用了 ssRNA-seq 和 dRNA-seq 的测序结果进行分析，并使用去重复后的 reads 进行后续的分析，充分保证了分析结果的可信度。

通常 ncRNA 相比 mRNA 长度较短，因而较难被鉴定，dRNA-seq 可在全基因组水平上鉴定转录本，极大地促进了 ncRNA 的研究。dRNA-seq 在古菌 ncRNA 的研究中也是一个重要的技术手段，Jager 等<sup>[10]</sup>首次用 dRNA-seq 对在不同含氮条件下马氏甲烷八叠球菌进行氮素响应的转录组测序研究分析，在全基因组水平上鉴定了 876 个 TSS，不仅发现马氏甲烷八叠球菌含有多个可能的 ncRNA，而且证实了多个 ncRNA 可在氮胁迫环境下特异表达，调控马氏甲烷八叠球菌快速适应底物转换。在本研究中，我们基于两种不同的 RNA-seq 技术测序分析，在地中海富盐菌中鉴定到高度可信的 ncRNA 转录本 105 个。值得我们注

意的是，在这 105 个高可信的 ncRNA 中，仅在盐度刺激下表达差异(以 RPKM 值为参考)超过 2 倍的 ncRNA 则有 45 个，推测 ncRNA 可能参与了极端嗜盐古菌对盐刺激的响应、信号传递或适应。而这些高可信度 ncRNA 的发现，还将为研究 ncRNA 在其它生理条件下的功能机制奠定基础。

在鉴定到的 105 个高可信度的 ncRNA 中，选取了表达差异超过 5 倍的 *incRNA1436* 和表达量高的 *incRNA1903* 进行 Northern Blot 实验，结果显示了这两个 ncRNA 基因的转录本的真实存在；且在不同盐浓度刺激下 *incRNA1436* 的表达量存在较大差异，都与我们的 RNA-seq 分析的结果一致，表明基于 RNA-seq 的结果鉴定的 ncRNA 是可信的，通过生物信息学分析可以有效地发现 ncRNA 存在与否。但是，CR-RT-PCR 结果表明 RNA-seq 预测的 ncRNA 转录本边界与 CR-RT-PCR 实验得到的转录本边界可能会出现小的差异。比如，Northern Blot 和 CR-RT-PCR 实验证明 *incRNA1436* 存在两个转录本，其中较大转录本的 TSS 与 RNA-seq 结果相同，而小转录本并没有通过 RNA-seq 鉴定出来。其主要原因可能是程序设计时默认 5'远端的转录起点为唯一转录起点，因此通过 Northern blot 和 CR-RT-PCR 实验对相关 ncRNA 进行进一步分析，对于发现其加工与适应机制是非常必要的。

ncRNA 虽然不编码蛋白质产物，但是在生物体内同样发挥着重要的生理功能，如 *incRNA1436* 在不同盐浓度刺激下存在较大的差异表达，可能参与地中海富盐菌对盐刺激的响应、信号传递和适应。而 ncRNA 在极端嗜盐古菌的遗传和生理代谢方面的功能将是我们下一阶段研究的重点部分。本研究通过高通量测序为基础的 RNA-seq 技

术, 结合生物信息学方法对地中海富盐菌 ncRNA 的系统鉴定, 为今后嗜盐古菌 ncRNA 对功能机制的研究奠定了坚实的基础。

## 参 考 文 献

- [1] Brantl S. Bacterial chromosome-encoded small regulatory RNAs. *Future Microbiology*, 2009, 4(1): 85–103.
- [2] Liu CN, Bai BY, Skogerbø G, Cai L, Deng W, Zhang Y, Bu DB, Zhao Y, Chen RS. NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Research*, 2005, 33: D112–D115.
- [3] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 2009, 10(1): 57–63.
- [4] Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, 2010, 7(9): 709–715.
- [5] Bischler T, Tan HS, Nieselt K, Sharma CM. Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods*, 2015, 86: 89–101.
- [6] Fan B, Li L, Chao Y, Förstner K, Vogel J, Borriss R, Wu XQ. dRNA-Seq reveals genomewide TSSs and noncoding RNAs of plant beneficial rhizobacterium *Bacillus amyloliquefaciens* FZB42. *PLoS One*, 2015, 10(11): e0142002.
- [7] Tang TH, Bachellerie JP, Rozhdestvensky T, Bortolin ML, Huber H, Drungowski M, Elge T, Brosius J, Hüttenhofer A. Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(11): 7536–7541.
- [8] Wurtzel O, Sapra R, Chen F, Zhu YW, Simmons BA, Sorek R. A single-base resolution map of an archaeal transcriptome. *Genome Research*, 2010, 20(1): 133–141.
- [9] Schattner P. Searching for RNA genes using base-composition statistics. *Nucleic Acids Research*, 2002, 30(9): 2076–2082.
- [10] Jager D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA. Deep sequencing analysis of the *Methanosarcina mazei* Go1 transcriptome in response to nitrogen availability. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(51): 21878–21882.
- [11] Klein RJ, Misulovin Z, Eddy SR. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proceedings of the National Academy of Sciences of the United States of America*, 2002, 99(11): 7542–7547.
- [12] Fischer S, Benz J, Späth B, Jellen-Ritter A, Heyer R, Dörr M, Maier LK, Menzel-Hobeck C, Lehr M, Jantzer K, Babski J, Soppa J, Marchfelder A. Regulatory RNAs in *Haloferax volcanii*. *Biochemical Society Transactions*, 2011, 39(1): 159–162.
- [13] Soppa J, Straub J, Brenneis M, Jellen-Ritter A, Heyer R, Fischer S, Granzow M, Voss B, Hess WR, Tjaden B, Marchfelder A. Small RNAs of the halophilic archaeon *Haloferax volcanii*. *Biochemical Society Transactions*, 2009, 37(1): 133–136.
- [14] Heyer R, Dörr M, Jellen-Ritter A, Späth B, Babski J, Jaschinski K, Soppa J, Marchfelder A. High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in *Haloferax volcanii*. *RNA Biology*, 2012, 9(7): 1011–1018.
- [15] Lui L, Lowe T. Small nucleolar RNAs and RNA-guided post-transcriptional modification. *Essays in Biochemistry*, 2013, 54: 53–77.
- [16] Babski J, Maier LK, Heyer R, Jaschinski K, Prasse D, Jäger D, Randau L, Schmitz RA, Marchfelder A, Soppa J. Small regulatory RNAs in Archaea. *RNA Biology*, 2014, 11(5): 484–493.
- [17] Li M, Liu HL, Han J, Liu JF, Wang R, Zhao DH, Zhou J, Xiang H. Characterization of CRISPR RNA biogenesis and Cas6 cleavage-mediated inhibition of a provirus in the haloarchaeon *Haloferax mediterranei*. *Journal of Bacteriology*, 2013, 195(4): 867–875.
- [18] Han J, Zhang F, Hou J, Liu XQ, Li M, Liu HL, Cai L, Zhang B, Chen YP, Zhou J, Hu SN, Xiang H. Complete genome sequence of the metabolically versatile halophilic archaeon *Haloferax mediterranei*, a poly(3-hydroxybutyrate-co-3-hydroxyvalerate) producer. *Journal of Bacteriology*, 2012, 194(16): 4463–4464.
- [19] Brenneis M, Hering O, Lange C, Soppa J. Experimental characterization of Cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genetics*, 2007, 3(12): e229.
- [20] Han J, Lu QH, Zhou LG, Zhou J, Xiang H. Molecular characterization of the *phaEC<sub>Hm</sub>* genes, required for biosynthesis of poly(3-hydroxybutyrate) in the extremely halophilic archaeon *Haloarcula marismortui*. *Applied and Environmental Microbiology*, 2007, 73(19): 6058–6065.
- [21] Sambrook J, Russell D W. *Molecular Cloning: A Laboratory Manual*. 3rd ed. New York: Cold Spring Harbor Laboratory



- Press, 2001.
- [22] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012, 9(4): 357–360.
- [23] Jorjani H, Zavolan M. TSSer: an automated method to identify transcription start sites in prokaryotic genomes from differential RNA sequencing data. *Bioinformatics*, 2014, 30(7): 971–974.
- [24] McClure R, Balasubramanian D, Sun Y, Bobrovskyy M, Sumby P, Genco CA, Vanderpool CK, Tjaden B. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research*, 2013, 41(14): e140.
- [25] Kuhn J, Binder S. RT-PCR analysis of 5' to 3'-end-ligated mRNAs identifies the extremities of *cox2* transcripts in pea mitochondria. *Nucleic Acids Research*, 2002, 30(2): 439–446.
- [26] Liu CH, Fan C. *De novo* transcriptome assembly of floral buds of pineapple and identification of differentially expressed genes in response to ethephon induction. *Frontiers in Plant Science*, 2016, 7: 203.

## Identification of non-coding RNA in *Haloferax mediterranei*

Lei Wang<sup>1#</sup>, Feiyue Cheng<sup>1,2#</sup>, Dahe Zhao<sup>1,2</sup>, Haibo Yang<sup>1,2</sup>, Hua Xiang<sup>1,2\*</sup>

<sup>1</sup> State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** [Objective] To identify non-coding RNAs in *Haloferax mediterranei* through high-throughput RNA sequencing, bioinformatics analysis and molecular techniques. [Methods] After *H. mediterranei* cells under log phase of growth were treated with different salt concentrations for 30 minutes, total RNA was extracted for the following strand-specific RNA sequencing and differential RNA sequencing. These RNA-seq data were used to identify the genome-wide ncRNAs and to predict the 5' and 3'-ends of the transcripts by bioinformatics analysis. A few selected ncRNAs were further confirmed by Northern blotting and Circularized RNA reverse transcription-PCR analysis. [Results] We identified 105 highly credible ncRNAs. Expression of four ncRNAs showed difference in different salt concentrations. We confirmed the expression, length of transcripts, transcription start and termination sites of incRNA1436 and incRNA1903 by Northern blotting and CR-RT-PCR. [Conclusion] We identified the ncRNAs of *H. mediterranei* in a genome-wide scale, including identification of a few ncRNAs involved in the responses of *H. mediterranei* to different salt concentrations. Our results have provided fundamental data and novel insights for future study of the function of ncRNA in haloarchaea.

**Keywords:** *Haloferax mediterranei*, non-coding RNA, salinity stimulation, RNA-sequencing

(本文责编: 张晓丽)

Supported by the National Natural Science Foundation of China (31301071)

\*Corresponding author. Tel/Fax: +86-10-64807472; E-mail: xiangh@im.ac.cn

#These authors contributed equally to this paper.

Received: 15 June 2016; Revised: 15 July 2016; Published online: 28 July 2016