J. Microbiol. Biotechnol. (2015), 25(8), 1234–1240
http://dx.doi.org/10.4014/jmb.1505.05014

jmb

Heat Shock RNA 1, Known as a Eukaryotic Temperature-Sensing Noncoding RNA, Is of Bacterial Origin S

Dongjin Choi, Hye Ji Oh, Chul Jun Goh, Kangseok Lee, and Yoonsoo Hahn*

Department of Life Science, Research Center for Biomolecules and Biosystems, Chung-Ang University, Seoul 156-756, Republic of Korea

Received: May 7, 2015
Revised: May 15, 2015
Accepted: May 19, 2015

First published online
May 26, 2015

**Corresponding author*
Phone: +82-2-820-5812;
Fax: +82-2-825-5206;
E-mail: hahny@cau.ac.kr

Supplementary data for this paper are available on-line only at http://jmb.or.kr.

pISSN 1017-7825, eISSN 1738-8872

Heat shock RNA 1 (HSR1) is described as a "eukaryotic heat-sensing noncoding RNA" that regulates heat shock response in human and other eukaryotic cells. Highly conserved HSR1 sequences have been identified from humans, hamsters, *Drosophila*, *Caenorhabditis elegans*, and *Arabidopsis*. In a previous study, however, it was suggested that HSR1 had originated from a bacterial genome. HSR1 showed no detectible nucleotide sequence similarity to any eukaryotic sequences but harbored a protein coding region that showed amino-acid sequence similarity to bacterial voltage-gated chloride channel proteins. The bacterial origin of HSR1 was not convincible because the nucleotide sequence similarity was marginal. In this study, we have found that a genomic contig sequence of *Comamonas testosteroni* strain JL14 contained a sequence virtually identical to that of HSR1, decisively confirming the bacterial origin of HSR1. Thus, HSR1 is an exogenous RNA, which can ectopically trigger heat shock response in eukaryotes. Therefore, it is no longer appropriate to cite HSR1 as a "eukaryotic functional noncoding RNA."

Keywords: Heat shock RNA-1, heat shock response, bacterial sequence, *Comamonas testosteroni*

Introduction

Heat shock RNA-1 (HSR1) is claimed to be a novel eukaryotic noncoding RNA that plays a pivotal role in inducing the expression of heat-shock protein genes by activating heat-shock transcription factor 1 [31, 32]. The HSR1 RNA sequence, which was first isolated from the hamster kidney cell line BHK-21, was reported to be ~2-kb-long with a poly(A) tail. The core segment was 604-bp-long without the poly(A) tail. Human HSR1 was isolated from a human cell line; it differed from the hamster HSR1 by only 4 bp [31]. Subsequently, highly conserved HSR1 sequences were found to be present in other animals and plants, including *Drosophila*, *Caenorhabditis elegans*, and *Arabidopsis* (US patents "US 8067558 B2" and "US 7919603 B2").

The novelty and importance of HSR1 have been highly recognized and it has been widely acknowledged as a potential RNA thermometer in eukaryotes, including humans [2, 3, 9, 18, 26, 27, 37]. However, a study suggested that HSR1 is derived from a bacterial species of the order Burkholderiales [14]. It showed no nucleotide sequence similarity to any eukaryotic sequences in the publicly available sequence database, although a large number of eukaryotic genome sequences were available. Instead, the hamster HSR1 sequence showed high similarity to those of bacterial proteins and marginal similarity to bacterial genomic sequences. The 3′-half of the HSR1 sequence was predicted to harbor a part of the sequence of the voltage-gated chloride channel protein gene, which is present in a wide variety of bacterial species. The 5′-half showed marginal nucleotide sequence similarity to the 5′-upstream region of the gene encoding the channel protein of *Burkholderia*, *Delftia*, and *Ralstonia* species. Based on this observation, it was proposed that HSR1 had originated from a bacterial genome, either by infection or by horizontal gene transfer [14].

In this study, we confirmed the bacterial origin of HSR1. A sequence virtually identical to that of HSR1 was found in a genomic contig sequence of *Comamonas testosteroni* (previously known as *Pseudomonas testosteroni*) strain JL14.

Based on this observation, it is argued that HSR1 is not a eukaryotic temperature-sensing noncoding RNA but a foreign RNA derived from a bacterial genome.

Materials and Methods

The information on HSR1 sequences was obtained from the patents "US 8067558 B2" and "US 7919603 B2" (accessed online, The Lens; https://www.lens.org). The sequences were retrieved from the "Patent" division of the National Center for Biotechnology Information (NCBI) "Nucleotide" database by using the aforementioned patent numbers as queries.

Sequence similarity searches of sequence databases were performed using the NCBI BLAST server (http://blast.ncbi.nlm.nih.gov/Blast.cgi) [13]. BLASTN searches of the reference bacterial genomic and whole-genome shotgun assembly sequences (BLAST database "refseq_genomic") were performed using HSR1 sequences as queries. Matched genomic contigs were retrieved and aligned segments were extracted. Pairwise alignments were performed using the FASTA program ver. 36.3.6 (http://faculty.virginia.edu/wrpearson/fasta/fasta36) [25]. Multiple sequence alignments were generated using the MUSCLE program ver. 3.8.31 (http://www.drive5.com/muscle) [6].

Comamonas genome sequences were downloaded from the NCBI Genome database (http://www.ncbi.nlm.nih.gov/genome/genomes/859). The percent identity plot was generated using the MultiPipMaker Web server (http://pipmaker.bx.psu.edu/pipmaker) [30].

Results and Discussion

The HSR1 sequence was claimed to be elucidated initially from humans and hamsters, and subsequently from *Drosophila*, *C. elegans*, and *Arabidopsis* [31, 32]. These sequences were available in the "Patent" division of the NCBI "Nucleotide" database (Table 1). The NCBI accession numbers were GY530733 (human), GY530732 (hamster), GY530763 (*Drosophila*), GY530766 (*C. elegans*), and GY530764 (*Arabidopsis*); the length of these sequences was 562 bp (human), 604 bp (hamster), 605 bp (*Drosophila* and *C. elegans*), and 217 bp (*Arabidopsis*). The *Arabidopsis* sequence contained some ambiguous nucleotides (N), suggesting a low sequencing quality. The HSR1 sequences from these organisms showed strong similarity to one another, with differences in only some nucleotides. In fact, the *C. elegans* HSR1 sequence was identical to the *Drosophila* HSR1 sequence.

Sequence similarity searches of HSR1 sequences against the NCBI bacterial genome sequence database by using the BLASTN program showed that an almost identical sequence was present in a genomic contig ("contig66"; NCBI Accession No. NZ_AWTN01000134) of *C. testosteroni* strain JL14 (Fig. 1). The human HSR1 sequence showed 99.4% identity in the 544 bp overlap region with the "contig66" of *C. testosteroni* JL14. Other animal HSR1 sequences showed similar levels of identity: hamster, 98.5% identity in the 548 bp overlap region; and *Drosophila*, 98.4% identity in the 561 bp overlap region. The 217-bp-long *Arabidopsis* HSR1 sequence showed 94.7% identity in the 113 bp overlap region. The presence of a sequence virtually identical to that of HSR1 in a bacterial genome confirmed the previous suggestion that HSR1 RNAs are derived from a bacterial genome [14].

As previously identified, the HSR1 sequence matched the 5′-upstream region and 5′-part of a gene encoding a voltage-gated chloride channel (NCBI Accession No. WP_034383148). The channel protein, a member of the "ClC_sycA_like" chloride channel protein family (NCBI CDD Accession No. cd03682), confers acid resistance to the bacterial species [12, 28]. The inferred protein sequence from HSR1 was almost identical to that of a protein of *C. testosteroni* JL14, which showed strong similarity with other bacterial voltage-gated chloride channel proteins (Fig. S1 and Table S1).

Table 1. Sequence information cited in this study.

Accession No.	Length	Definition	Organism
GY530733.1	562	Sequence 2 from patent US 8067558	Human
GY530732.1	604	Sequence 1 from patent US 8067558	Hamster
GY530763.1	605	Sequence 33 from patent US 8067558	*Drosophila* (identical to GY530766)
GY530766.1	605	Sequence 36 from patent US 8067558	*Caenorhabditis elegans* (identical to GY530763)
GY530764.1	217	Sequence 34 from patent US 8067558	*Arabidopsis* (identical to a part of GY530761)
GY530761.1	1,080	Sequence 31 from patent US 8067558	Unknown (contains GY530764)
NZ_AWTN01000134.1	13,893	*Comamonas testosteroni* strain JL14 contig66, whole-genome shotgun sequence	*Comamonas testosteroni* JL14
NZ_KE384553.1	457,580	*Solimonas flava* DSM 18980 K343DRAFT_scaffold00004.4, whole-genome shotgun sequence	*Solimonas flava* DSM 18980

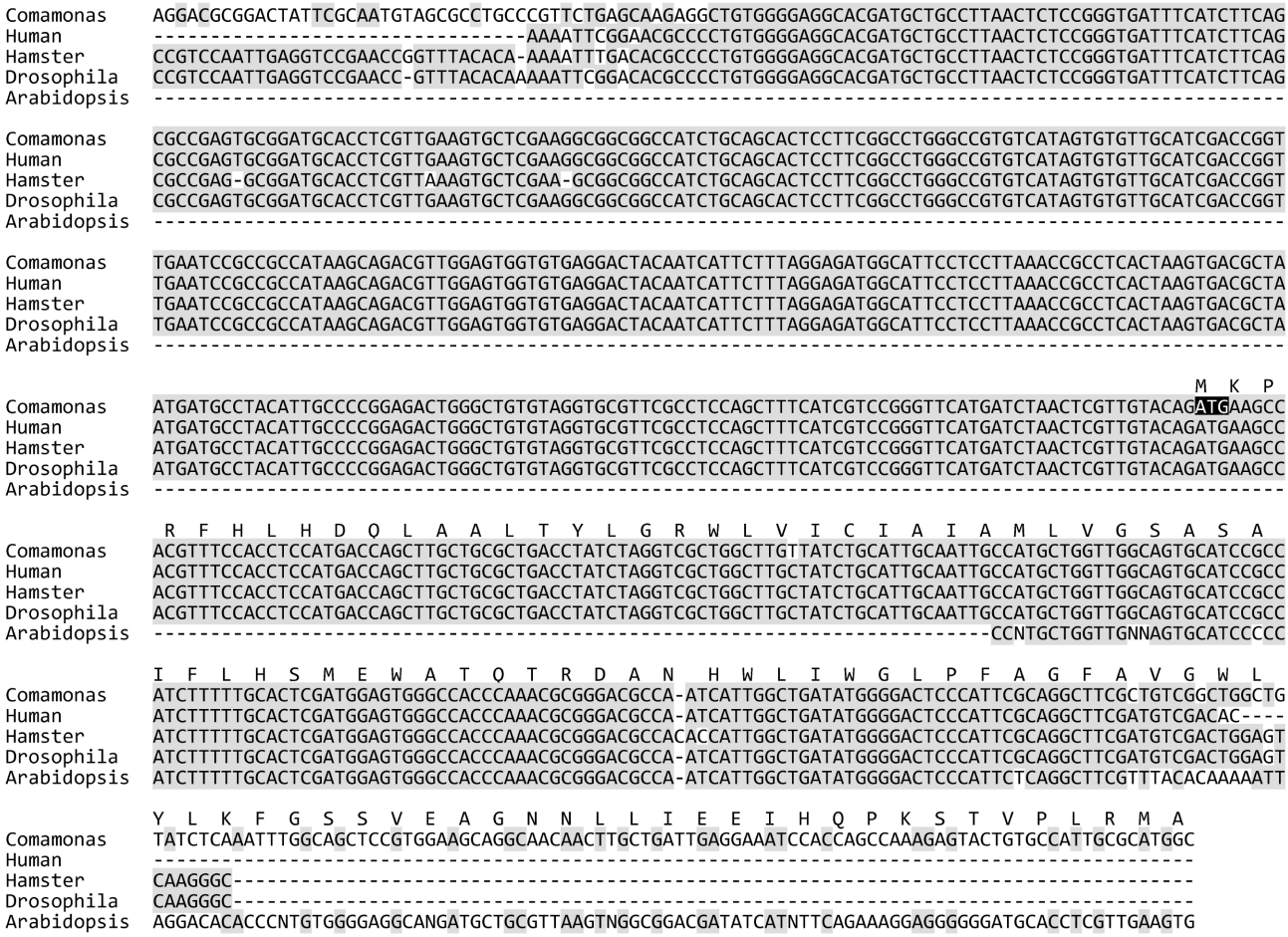

Fig. 1. Multiple alignment of HSR1 sequences and the *C. testosteroni* strain JL14 genomic "contig66."
A part of the *C. testosteroni* strain JL14 "contig66" (from residue 6387 to residue 5697) and the full-length sequences of human, hamster, *Drosophila*, and *Arabidopsis* HSR1 RNAs were multiply aligned. Sequences that are conserved in two or more species are highlighted against a gray background. The amino acid sequence of the voltage-gated chloride channel protein is shown at the top. The start codon is highlighted against a black background. Table 1 lists the NCBI accession numbers of these sequences.

C. testosteroni strain JL14 was isolated from antimony mine soil [19]. *Comamonas* species, which are members of the order Burkholderiales, are commonly found in soil, mud, and water as well as in animal tissues and blood, clinical samples, and the hospital environment [36]. Although *C. testosteroni* has rarely been implicated as a human pathogen, there have been some cases of bacteremia due to *C. testosteroni* infection, some of which were fatal [7, 23, 24, 34].

In addition, a 1080-bp-long HSR1 sequence was isolated from an unknown organism (NCBI Accession No. GY530761). This sequence contained many ambiguous nucleotides, especially at both ends, indicating that the sequence was not comprehensively determined. The sequence from 434 to 650 bp of the HSR1 sequence of the unknown organism was identical to the entire sequence of *Arabidopsis* HSR1, including seven ambiguous nucleotides (Fig. 2A). Therefore, it was assumed that the sequence GY530761 was also isolated from *Arabidopsis*. The 5′-part (433 bp) and 3′-part (430 bp) surrounding the *Arabidopsis* HSR1 sequence showed sequence similarity to each other in the reverse direction.

The 5′- and 3′-parts of the unknown sequence also showed strong sequence similarity to bacterial genome sequences. The most similar sequence in the current NCBI database is a genomic contig segment "K343DRAFT_scaffold00004.4" (NCBI Accession No. NZ_KE384553) of *Solimonas flava* (originally known as *Sinobacter flavus*) strain DSM 18980 [33, 40] (Fig. 2B). The *S. flava* genomic segment showed 92.5% identity in the 400 bp overlap region with the 5′-part and 80.1% identity in the 402 bp overlap region with the 3′-

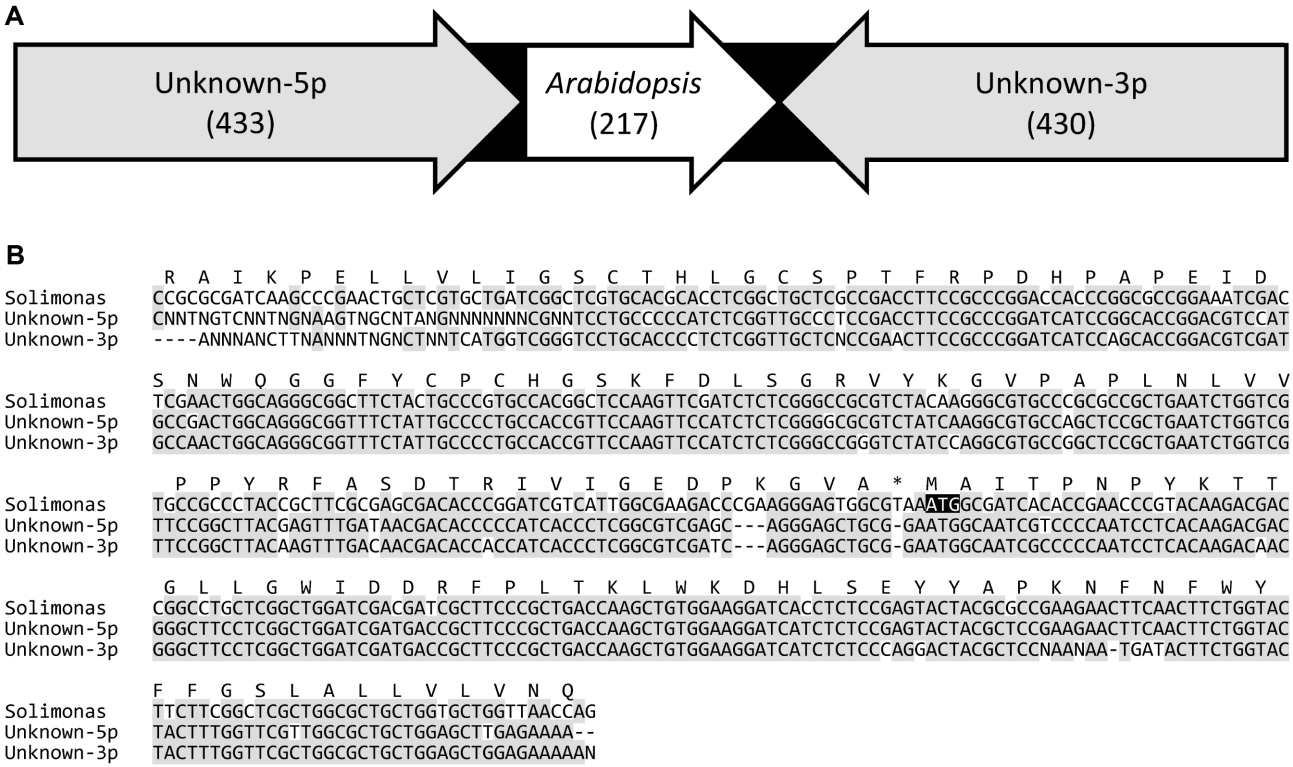

Fig. 2. Structure of the HSR1 sequence of an unknown organism and comparison with the genomic sequence of *Solimonas* species. (**A**) The HSR1 sequence of an unknown organism is composed of three parts. The central part is identical to the *Arabidopsis* HSR1 sequence. The two flanking regions are palindromic to each other. (**B**) The flanking sequences "Unknown-5p" and "Unknown-3p" show strong similarities with a genomic segment of *Solimonas* species. Positions that are conserved in two or more sequences are highlighted against a gray background. The amino acid sequences of *S. flava* DSM 18980 ubiquinol-cytochrome C reductase and cytochrome B are shown at the top. An asterisk (*) indicates the stop codon of the ubiquinol-cytochrome C reductase. The start codon of the cytochrome B is highlighted against a black background.

part of the unknown sequence. The matched genomic segment included parts of two protein-coding genes: 3′-part of ubiquinol-cytochrome C reductase (NCBI Accession No. WP_028009503) and 5′-part of cytochrome B (WP_043113184). These two proteins were arranged in a tail-to-head manner; the start codon of the cytochrome B gene immediately followed the stop codon of the ubiquinol-cytochrome C reductase gene.

The unknown HSR1 sequence, which was probably isolated from an *Arabidopsis* sample, was a hybrid molecule with the central part from *C. testosteroni* and two terminal parts from *Solimonas*-related species. It is likely that the unknown sequence originated from the fusion of cDNA or genomic DNA molecules derived from two different bacterial species during cDNA preparation or PCR amplification. The *C. testosteroni* fragment could have originated from contaminated "animal" HSR1 molecules. The *S. flava* fragment could be derived from soil contaminants during the preparation of the *Arabidopsis* sample, because *Solimonas* bacteria have mainly been isolated from soil [15, 16].

Currently, there are 19 genome sequence assemblies of *C. testosteroni* strains in the NCBI genome sequence database [8, 10, 19, 20, 29, 39]. However, the HSR1 sequence was identified only in *C. testosteroni* strain JL14. When the JL14 "contig66" was aligned with the other 18 *C. testosteroni* genome assemblies, only strain JL14 was shown to harbor the HSR1 sequence as well as the voltage-gated chloride channel protein gene sequence (Fig. 3). Interestingly, an integrase gene (NCBI Accession No. WP_034383169) was found adjacent to the channel protein gene, indicating that the segment containing the HSR1 sequence was part of a mobile genetic element, probably an integrative conjugative element [4, 38]. It is likely that this segment has been mobilized into *C. testosteroni* strain JL14 from a closely related Burkholderiales species, because the channel protein showed a strong sequence similarity to orthologous proteins found in other Burkholderiales species [14]. This could explain why only strain JL14 carried the HSR1 sequence,

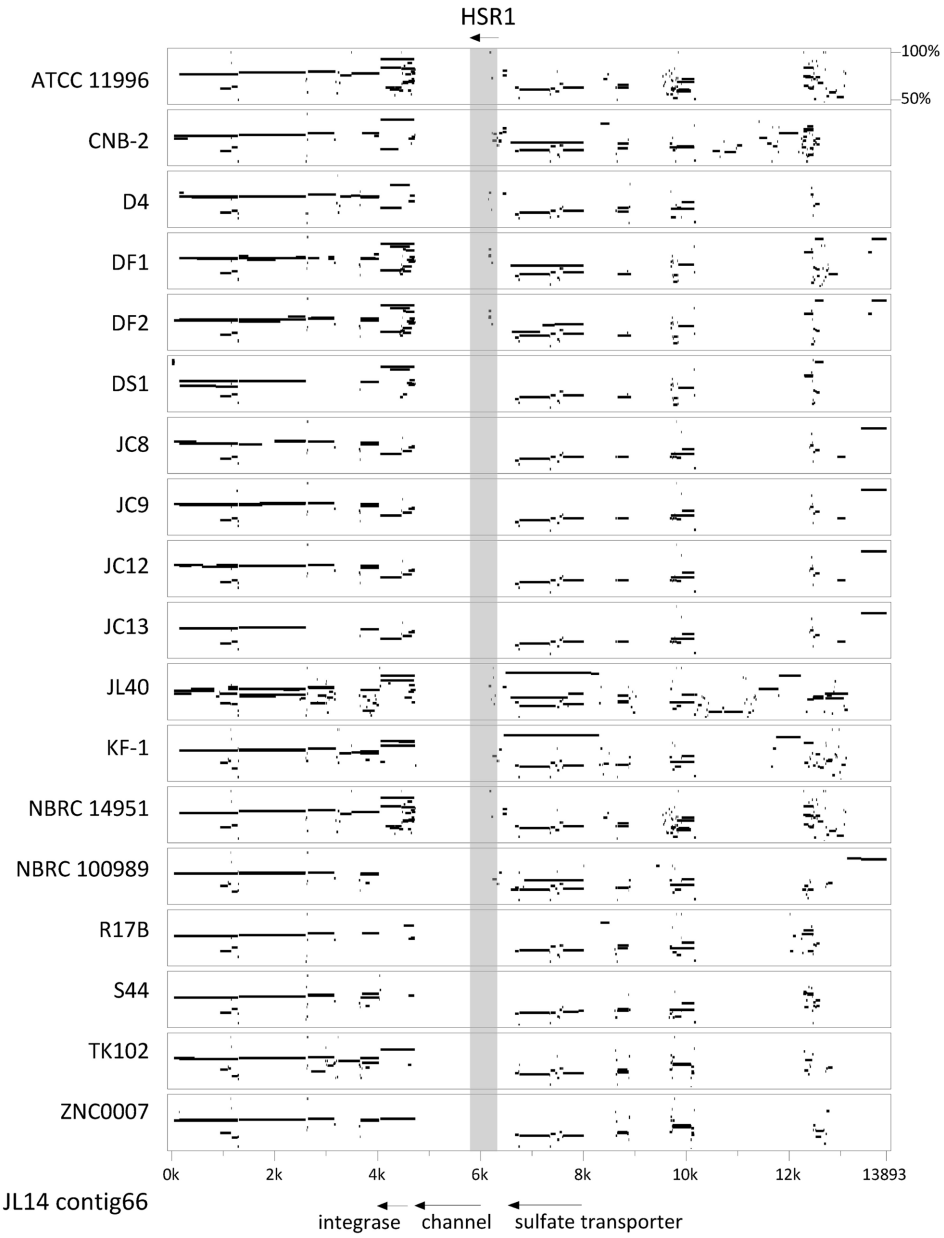

Fig. 3. Comparison of the genomic sequences of *C. testosteroni*.
The *C. testosteroni* JL14 "contig66" sequence was compared with 18 other genome assemblies of *C. testosteroni* strains by using the MultiPipMaker Web server. The segment identical to the HSR1 sequence is marked with a vertical gray bar. The sequences encoding the integrase, voltage-gated chloride channel, and sulfate transporter (NCBI accession numbers: WP_034383169, WP_012347091, and WP_003052400, respectively) are indicated at the bottom.

unlike the other strains. The mobile element containing HSR1 harbors a voltage-gated channel protein, which provides acid resistance so that it may be beneficial to the bacteria in an acidic soil environment [12, 28].

It is interesting how RNA molecules with a bacterial origin could be isolated from various eukaryotic cells, with a role in the modulation of eukaryotic heat shock response [31]. Two explanations are possible: (i) eukaryotic cells may have identified bacterial RNA molecules to regulate protein activity and gene expression as a defense mechanism [1, 21] or (ii) eukaryotic cells may ectopically respond to foreign RNA molecules that are transferred from bacterial to eukaryotic cells possibly *via* outer membrane vesicles [5, 17]. It is well known that eukaryotic cells respond to

exogenous RNAs; environmental RNA interference is an example of this [11, 22, 35]. Therefore, it is possible that RNA molecules derived from bacteria affect on the eukaryotic heat shock response, although it may not be their original function.

In summary, HSR1 noncoding RNAs, claimed to be isolated from eukaryotic cells, were confirmed to have originated from a bacterial genome. The eukaryotic heat shock response may be ectopically triggered by the exogenous HSR1 RNA molecules. Therefore, it is no longer appropriate to cite HSR1 as a "eukaryotic functional noncoding RNA."

Acknowledgments

This study was supported by the Chung-Ang University Research Scholarship Grants in 2014, and the National Research Foundation of Korea (NRF) grant (NRF-2014R1A2A2A09052791), funded by the Ministry of Education, Science and Technology, Republic of Korea.

References

1. Abdullah Z, Knolle PA. 2014. Scaling of immune responses against intracellular bacterial infection. *EMBO J.* **33:** 2283-2294.
2. Akerfelt M, Morimoto RI, Sistonen L. 2010. Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* **11:** 545-555.
3. Anckar J, Sistonen L. 2011. Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu. Rev. Biochem.* **80:** 1089-1115.
4. Bellanger X, Payot S, Leblond-Bourget N, Guedon G. 2014. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* **38:** 720-760.
5. Berleman J, Auer M. 2013. The role of bacterial outer membrane vesicles for intra- and interspecies delivery. *Environ Microbiol.* **15:** 347-354.
6. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32:** 1792-1797.
7. Farshad S, Norouzi F, Aminshahidi M, Heidari B, Alborzi A. 2012. Two cases of bacteremia due to an unusual pathogen, *Comamonas testosteroni* in Iran and a review literature. *J. Infect. Dev. Ctries.* **6:** 521-525.
8. Fukuda K, Hosoyama A, Tsuchikane K, Ohji S, Yamazoe A, Fujita N, *et al.* 2014. Complete genome sequence of polychlorinated biphenyl degrader *Comamonas testosteroni* TK102 (NBRC 109938). *Genome Announc.* **2:** e00865-14.
9. Geisler S, Coller J. 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* **14:** 699-712.
10. Gong W, Kisiela M, Schilhabel MB, Xiong G, Maser E. 2012. Genome sequence of *Comamonas testosteroni* ATCC 11996, a representative strain involved in steroid degradation. *J. Bacteriol.* **194:** 1633-1634.
11. Ivashuta S, Zhang Y, Wiggins BE, Ramaseshadri P, Segers GC, Johnson S, *et al.* 2015. Environmental RNAi in herbivorous insects. *RNA* **21:** 840-850.
12. Iyer R, Iverson TM, Accardi A, Miller C. 2002. A biological role for prokaryotic ClC chloride channels. *Nature* **419:** 715-718.
13. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36:** W5-W9.
14. Kim DS, Lee Y, Hahn Y. 2010. Evidence for bacterial origin of heat shock RNA-1. *RNA* **16:** 274-279.
15. Kim MK, Kim YJ, Cho DH, Yi TH, Soung NK, Yang DC. 2007. *Solimonas soli* gen. nov., sp. nov., isolated from soil of a ginseng field. *Int. J. Syst. Evol. Microbiol.* **57:** 2591-2594.
16. Kim SJ, Moon JY, Weon HY, Ahn JH, Chen WM, Kwon SW. 2014. *Solimonas terrae* sp. nov., isolated from soil. *Int. J. Syst. Evol. Microbiol.* **64:** 1218-1222.
17. Kulp A, Kuehn MJ. 2010. Biological functions and biogenesis of secreted bacterial outer membrane vesicles. *Annu. Rev. Microbiol.* **64:** 163-184.
18. Lakhotia SC. 2012. Long non-coding RNAs coordinate cellular responses to stress. *Wiley Interdiscip. Rev. RNA* **3:** 779-796.
19. Liu L, Zhu W, Cao Z, Xu B, Wang G, Luo M. 2015. High correlation between genotypes and phenotypes of environmental bacteria *Comamonas testosteroni* strains. *BMC Genomics* **16:** 110.
20. Ma YF, Zhang Y, Zhang JY, Chen DW, Zhu Y, Zheng H, *et al.* 2009. The complete genome of *Comamonas testosteroni* reveals its genetic adaptations to changing environments. *Appl. Environ. Microbiol.* **75:** 6812-6819.
21. Maldonado-Bonilla LD, Eschen-Lippold L, Gago-Zachert S, Tabassum N, Bauer N, Scheel D, Lee J. 2014. The *Arabidopsis* tandem zinc finger 9 protein binds RNA and mediates pathogen-associated molecular pattern-triggered immune responses. *Plant Cell Physiol.* **55:** 412-425.
22. McEwan DL, Weisman AS, Hunter CP. 2012. Uptake of extracellular double-stranded RNA by SID-2. *Mol. Cell* **47:** 746-754.
23. Nseir W, Khateeb J, Awawdeh M, Ghali M. 2011. Catheter-related bacteremia caused by *Comamonas testosteroni* in a hemodialysis patient. *Hemodial. Int.* **15:** 293-296.
24. Orsini J, Tam E, Hauser N, Rajayer S. 2014. Polymicrobial bacteremia involving *Comamonas testosteroni*. *Case Rep. Med.* **2014:** 578127.
25. Pearson WR, Lipman DJ. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85:** 2444-2448.
26. Place RF, Noonan EJ. 2014. Non-coding RNAs turn up the heat: an emerging layer of novel regulators in the mammalian heat shock response. *Cell Stress Chaperones* **19:** 159-172.
27. Ponting CP, Oliver PL, Reik W. 2009. Evolution and

functions of long noncoding RNAs. *Cell* **136:** 629-641.
28. Rojas-Jimenez K, Sohlenkamp C, Geiger O, Martinez-Romero E, Werner D, Vinuesa P. 2005. A ClC chloride channel homolog and ornithine-containing membrane lipids of *Rhizobium tropici* CIAT899 are involved in symbiotic efficiency and acid tolerance. *Mol. Plant Microbe Interact.* **18:** 1175-1185.
29. Schleheck D, Knepper TP, Fischer K, Cook AM. 2004. Mineralization of individual congeners of linear alkylbenzenesulfonate by defined pairs of heterotrophic bacteria. *Appl. Environ. Microbiol.* **70:** 4053-4063.
30. Schwartz S, Elnitski L, Li M, Weirauch M, Riemer C, Smit A, *et al.* 2003. MultiPipMaker and supporting tools: alignments and analysis of multiple genomic DNA sequences. *Nucleic Acids Res.* **31:** 3518-3524.
31. Shamovsky I, Ivannikov M, Kandel ES, Gershon D, Nudler E. 2006. RNA-mediated response to heat shock in mammalian cells. *Nature* **440:** 556-560.
32. Shamovsky I, Nudler E. 2009. Isolation and characterization of the heat shock RNA 1. *Methods Mol. Biol.* **540:** 265-279.
33. Sheu SY, Cho NT, Arun AB, Chen WM. 2011. Proposal of *Solimonas aquatica* sp. nov., reclassification of *Sinobacter flavus* Zhou *et al*. 2008 as *Solimonas flava* comb. nov. and *Singularimonas variicoloris* Friedrich and Lipski 2008 as *Solimonas variicoloris* comb. nov. and emended descriptions of the genus *Solimonas* and its type species *Solimonas soli*. *Int. J. Syst. Evol. Microbiol.* **61:** 2284-2291.
34. Tsui TL, Tsao SM, Liu KS, Chen TY, Wang YL, Teng YH, Lee YT. 2011. *Comamonas testosteroni* infection in Taiwan: reported two cases and literature review. *J. Microbiol. Immunol. Infect.* **44:** 67-71.
35. Whangbo JS, Hunter CP. 2008. Environmental RNA interference. *Trends Genet.* **24:** 297-305.
36. Willems A, De Vos P. 2006. *Comamonas*, p. 723-736. *The Prokaryotes*. Springer, New York.
37. Wilusz JE, Sunwoo H, Spector DL. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23:** 1494-1504.
38. Wozniak RA, Waldor MK. 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* **8:** 552-563.
39. Xiong J, Li D, Li H, He M, Miller SJ, Yu L, *et al.* 2011. Genome analysis and characterization of zinc efflux systems of a highly zinc-resistant bacterium, *Comamonas testosteroni* S44. *Res. Microbiol.* **162:** 671-679.
40. Zhou Y, Zhang YQ, Zhi XY, Wang X, Dong J, Chen Y, *et al.* 2008. Description of *Sinobacter flavus* gen. nov., sp. nov., and proposal of Sinobacteraceae fam. nov. *Int. J. Syst. Evol. Microbiol.* **58:** 184-189.