Genomics 91 (2008) 553–554

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Reply to Letter to the Editor

Genome sequences of *Halobacterium salinarum*: A reply

In a letter to the editor, Ng et al. [1] present their view of our publication describing the genome of *Halobacterium salinarum* strain R1 and its comparison to that of strain NRC-1 [2]. To make the key issues understandable to the readers of *Genomics*, we provide the following information.

One major point of controversy is whether "*Halobacterium* sp. strain NRC-1" belongs to *H. salinarum* or represents a different species. Ng et al. explain that a variety of designations exist, "*H. halobium, H. cutirubrum, H. salinarium,* and *H. salinarum,*" and that their designation for NRC-1 "is simply a way to show caution and objectivity to avoid confusion." However, they omitted some information that is useful in clarifying how these species have been described and named. To the readers not familiar with the haloarchaeal strain history, which is indeed complex, it should be noted that the above-mentioned four species were found to be so similar that they were combined into the single species *H. salinarum* by the International Committee on Systematic Bacteriology in 1996 [3]. Another bit of omitted information is that for most of its public life NRC-1 was published as *H. halobium* (now *H. salinarum*) strain NRC-1. Among the authors using this designation were Doolittle, who donated the strain to the DasSarma group, and DasSarma himself (all publications until 1996). Within the next 2 years the strain was "converted" into a different species, i.e., "*H.* sp. strain NRC-1." The strain was deposited by DasSarma in the American Type Culture Collection, where it is listed as *H. salinarum* (ATCC 700922). Concerning the relationship between the two sequenced strains, strain R1 has been described in numerous publications to be a spontaneous gas vesicle-negative mutant of NRC-1; the list of authors includes DasSarma himself (e.g., [4]). This is consistent with our finding that the chromosomes of the two strains are virtually identical and exhibit only 12 differences, mostly single base changes. The 12 differences can be compared to *Escherichia coli* for which two strains are recognized to belong to the same species despite differing by 75,168 single nucleotide polymorphisms and more than 1900 genes that occur only in one of the two strains [5]. However, there is even an additional proof that R1 and NRC-1 descended from the same natural isolate, i.e., the distribution of insertion elements in both genomes [2]. Sixteen of the plasmidal insertion elements are localized at identical positions, highly unlikely in two independent isolates from different species. In summary, NRC-1 was published for more than a decade to be a strain of *H. halobium* (now *H. salinarum*), it is listed today by type culture collections around the world as *H. salinarum*, and the genome comparison shows that NRC-1 and *H. salinarum* R1 descended from a single isolate. In spite of these facts, Ng et al. conclude in their last paragraph "one can place them both within the genus of *Halobacterium,*" at least inferring that they belong to different species. What is the motivation of trying to keep the "widely used strain" *H.* sp. strain NRC-1 (52 publications under this designation) apart from the species *H. salinarum* (nearly 1400 publications, including the old species names)?

A second major point of controversy is the late publication of the R1 genome sequence and annotation differences between the R1 and NRC-1 genomes. We feel very comfortable with acknowledging the extremely high quality of the sequence of the NRC-1 genome. Also the annotation was of high quality—in the year 2000. There was only a limited set of required corrections at that time and this was exactly the reason that we chose not to publish a very similar paper about the R1 genome soon after the NRC-1 genome publication. Instead, we started to sequence additional haloarchaeal genomes and to establish large-scale proteomics. Both types of projects helped to enhance the quality of gene annotation. However, the statement of Ng et al. that "the nucleotide sequence was not publicly available to the community" is simply not true. A publicly available website (www.halolex.mpg.de) was created and its existence was reported regularly at archaeal meetings starting in 2002; in addition, it was included in more than 20 publications that used the R1 genome sequence, also starting in 2002 [6]. Of course, the genome sequence was deposited in the EMBL/GenBank database upon publication [2] under Accession Nos. AM774415 (chromosome) and AM774416-AM774419 (4 plasmids).

Ng et al. write that "because NRC-1 was the first completed halophile genome, the advantage of extensive gene curation using a comparative genomic approach was not available." While this was certainly true in 2000, the opportunity now exists to use additional information, including comparative genomics and proteomics data, to improve the annotation of the *H. salinarum* genome. The recent publication of three additional haloarchaeal genomes (two were sequenced by the Oesterhelt group) and of numerous proteomic results indeed tremendously helped to discriminate between real genes and spurious ORFs and to identify the correct translational initiation points. We do not feel that we offended the NRC-1 annotation from 2000 by publishing a R1 annotation of a 2007 quality standard. In contrast, knowledge about the 111 genes that were missed in the early annotation and the 20% difference in start codon assignments will allow a more successful usage of the NRC-1 genome sequence. We agree with Ng et al. that correct start codon assignment is not absolutely necessary for a variety of approaches (including the identification of protein spots in 2D gels), but of course it is very helpful. The proposed use of a six-frame translation of the nucleotide sequence is possible but has disadvantages. In addition, we would like to emphasize that correct start codon assignment is of utmost importance for other experiments. Examples are the analyses of promoters or 5′-UTRs, the production of the correct protein for biochemical characterization, or bioinformatic genome analysis. Therefore, users of public databases like SwissProt should be aware that 20% of the NRC-1 proteome is currently not correct and that verification, especially of translational start points, is necessary before NRC-1 genes are included in experiments of the types mentioned above.

A third major point of controversy is the relationship between the plasmids from NRC-1 and R1. A careful analysis revealed that they are

0888-7543/$ – see front matter
doi:10.1016/j.ygeno.2008.04.006

composed of a patchwork of regions, most of which have nearly identical sequence, but which were assembled very differently into two (NRC-1) or four (R1) plasmids. This patchwork is due to a set of breakpoints, every single one associated with an ISH element. Verification of a correct assembly requires the experimental validation of connectivity across these ISH elements, which we have performed for the R1 but not for the NRC-1 plasmids [2]. ISH elements can be the reason for real genetic instability due to recombination *in vivo*, but also for assembly problems during genome sequencing projects. We have mentioned both possibilities in our paper with—admittedly—a personal preference for the latter. Ng et al. argue vigorously that their plasmid assembly is correct, but not all of their arguments are really convincing. They state that we ignored the "resequencing of the NRC-1 genome by the 454 company published over 2 years ago [7], which confirmed the original conclusion." The short read lengths from 454 sequencing are well below the size of even the shortest ISH element and thus cannot provide any evidence to confirm connectivities at any of the breakpoints. Thus, the statement that the Pinard et al. paper confirms the plasmid assembly is not true. They also state that "a decade of cloning, mapping, and sequencing had clearly established both the structures and rearrangements of these extrachromosomal elements." They cite a long list of 16 papers to confirm this conclusion, all of them self-citations. However, most of these papers do not deal with plasmid assembly but describe experiments with the gas vesicle gene cluster, which is located in a plasmid region for which the assembly is not controversial. Nevertheless, we have to admit that we overlooked one paper that presented evidence for the physical structure of pNRC100 [8], and we apologize for this mistake. Pulse gel electrophoresis and Southern blotting were used to generate a map of the plasmid, including an ordered HindIII library. Ng et al. declare in their letter that "a similar but less laborious approach was undertaken … to map" pNRC200 and cite their genome paper (which does not contain this information) and a university report (their reference [45]). As we have provided careful experimental evidence for the plasmid architecture of R1 [2], we accept that genetic recombination has led to the differences in plasmid structure in NRC-1 and R1 after the original natural isolate was separated in two lineages—evolution in the laboratory.

Ng et al. further elaborate on "insertion/deletion 3," which is one of the 12 biological differences between the chromosomes of the two strains listed in Table 1 of our paper [2]. The multiple sequence alignment of Ng et al. (Fig. 2 of their letter [1]) resolves the "insertion/deletion" uncertainty and reveals that the longer version is conserved and a deletion has occurred in R1. Otherwise, this biological difference is not controversial.

With our recent paper we had not intended to revive an old "race" or "controversy" but to publish an up-to-date annotation of the *H. salinarum* genome. As described above, comparison with the NRC-1 annotation highlights scientific progress since 2000. We agree with Ng et al. that *H. salinarum* has developed into an archaeal model species. Since they became available in 2000 and in 2002, the two genome sequences have induced numerous studies. Functional genomic approaches (transcriptomics, proteomics) have been established with both strains, NRC-1 and R1, and modeling of different aspects of gene regulation and metabolism has been reported recently. Instead of citing a long list of papers we would like to draw attention to four reviews. One is confined to results obtained with the strain NRC-1 [9] and the other three summarize different aspects of both strains and additional haloarchaeal species [10–12]. The progress obtained in recent years with both strains has been fascinating. We should close this—in our opinion rather unnecessary—discussion and all of us should instead concentrate on the exciting science that can be performed with *H. salinarum*.

References

[1] W.V. Ng, B.R. Berquist, J.A. Coker, M. Capes, T.H. Wu, P. DasSarma, S. DasSarma, Genome sequences of *Halobacterium* species, Genomics 91 (2008) 548–552.
[2] F. Pfeiffer, S.C. Schuster, A. Broicher, et al., Evolution in the laboratory: the genome of Halobacterium salinarum strain R1 compared to that of strain NRC-1, Genomics 91 (2008) 335–346.
[3] A. Ventosa, A. Oren, Halobacterium salinarum nom. Corrig., a name to replace Halobacterium salinarium (Elazari-Volcani) and to include Halobacterium halobium and Halobacterium cutirubrum, Int. J. Syst. Bacteriol. 46 (1996) 347.
[4] S. Dassarma, J.T. Halladay, J.G. Jones, J.W. Donovan, P.J. Giannasca, N.T. de Marsac, High-frequency mutations in a plasmid-encoded gas vesicle gene in Halobacterium halobium, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 6861–6865.
[5] N.T. Perna, G. Plunkett III, V. Burland, et al., Genome sequence of enterohaemorrhagic Escherichia coli O157:H7, Nature 409 (2001) 529–533.
[6] J. Soppa, K. Kobayashi, M.F. Noirot-Gros, D. Oesterhelt, S.D. Ehrlich, E. Dervyn, N. Ogasawara, S. Moriya, Discovery of two novel families of proteins that are proposed to interact with prokaryotic SMC proteins, and characterization of the Bacillus subtilis family members ScpA and ScpB, Mol. Microbiol. 45 (2002) 59–71.
[7] R. Pinard, A. de Winter, G.J. Sarkis, M.B. Gerstein, K.R. Tartaro, R.N. Plant, M. Egholm, J.M. Rothberg, J.H. Leamon, Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing, BMC Genomics 7 (2006) 216.
[8] W.L. Ng, S. Kothakota, S. DasSarma, Structure of the gas vesicle plasmid in Halobacterium halobium: inversion isomers, inverted repeats, and insertion sequences, J. Bacteriol. 173 (1991) 1958–1964.
[9] S. Dassarma, B.R. Berquist, J.A. Coker, P. Dassarma, J.A. Muller, Post-genomics of the model haloarchaeon Halobacterium sp. NRC-1, Saline Systems 2 (2006) 3.
[10] M. Falb, K. Muller, L. Konigsmaier, et al., Metabolism of halophilic archaea, Extremophiles 12 (2008) 177–196.
[11] J. Soppa, From genomes to function: haloarchaea as model organisms, Microbiology 152 (2006) 585–590.
[12] J. Soppa, A. Baumann, M. Brenneis, M. Dambeck, M. Hering, and C. Lange, Genomics and functional genomics with haloarchaea, Arch. Microbiol., doi:10.1007/s00203-008-0376-4.

F. Pfeiffer
S.C. Schuster
A. Broicher
M. Falb
P. Palm
K. Rodewald
J. Tittor
D. Oesterhelt*
Department of Membrane Biochemistry, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany
E-mail address: oesterhe@biochem.mpg.de (D. Oesterhelt).
*Corresponding author. Fax: +49 89 8578 3557.

A. Ruepp
Department of Molecular Structural Biology, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany

J. Soppa
Institute for Molecular Biosciences, Goethe University, Frankfurt, Germany

18 April 2008