Contents lists available at ScienceDirect

Genomics

GENOMICS

Letter to the Editor

Genome sequences of *Halobacterium* species

The last decade of the 20th century was an exciting period of genomic research, with the appearance of genome sequences for several dozen diverse microorganisms. Not surprisingly, *Science* magazine chose sequenced genomes as the breakthrough of the year in 2000 [1] and the American Society for Microbiology published a timeline for completed microbial genomes on the cover of *ASM News* in the June edition of 2001 [2]. Among these notable achievements was the genome sequence of a representative halophilic archaeon, a member of the third branch of life that could grow at saturated concentrations of sodium chloride. In fact, this haloarchaeon, *Halobacterium* sp. NRC-1, required brine several times more concentrated than seawater for viability, survived desiccated in salt crystals, and tolerated ultraviolet and ionizing irradiation at intensities thousands of times higher than most cells can withstand [3–5]. The genome sequence of *Halobacterium* sp. strain NRC-1 was published on October 24, 2000, in a paper co-authored by 44 scientists belonging to an international consortium of 12 research groups [6]. This genome has served as the prototype haloarchaeal genome ever since.

The publication of the NRC-1 genome, together with development of a facile knockout system using the selectable and counter-selectable *ura3* gene, provided ample research opportunities for the community of microbiologists, geneticists, and evolutionary biologists interested in this class of archaea [7]. Subsequent development of whole genome DNA microarrays showed the value of the completed sequence and led to analyses of the response of this organism to a wide variety of environmental changes [8]. These resources coupled with a bioinformatic database and tool, HaloWeb [9], publicly available since the publication of the genome sequence in 2000, recognized by Thomson and indexed in ISI Web of Knowledge Current Web, has catapulted this archaeal strain to a model system used in laboratories world-wide and employed for teaching in schools and colleges [10]. Not surprisingly, the original genome paper, published in the *Proceedings of the National Academy of Sciences USA*, has been cited over 300 times to date in the literature and two related publications in *Genome Research*, "Sequencing of the NRC-1 small megaplasmid" [11] and "Computational analysis of the complete genome sequence" [12], have been cited over 100 times.

One of the most anticipated pieces of data following up on the *Halobacterium* sp. NRC-1 genome sequence was the complete genome sequence of R-1. Although this genome was reported as being complete in many meetings over the years and touted as being essentially identical to NRC-1, as far back as 2002 on a private website (HaloLex [13]), the nucleotide sequence was not made publicly available to the community. In fact, the purported R-1 genome sequence was not deposited in any public database and was not available for download from NCBI until July 27, 2007. By comparison, the NRC-1 genome had been shared with all 12 collaborating laboratories during the draft phase and the final sequence subsequently deposited for the entire international research community on July 14, 2000. During the interim period between public release of the initial *Halobacterium* sp. genome sequence and the second *Halobacterium* sp. genome sequence, over 600 other genomes have been reported, including several other haloarchaeal genomes.

One reason that the genome of R-1 was so eagerly anticipated was that the classic papers by Sapienza and Doolittle in *Nature* [14,15] had reported that the genomes of both of these strains were highly novel. These authors reported, based on probing with random fragments cloned with EcoRI, HindIII, and BamHI, that the genomes were full of repeated elements, estimated to belong to more than 50 families. They also reported that greater than 4×10^{-3} recombinational events occur per family per generation and that two daughter cells produced by a single cell division have only an 80% chance of bearing identical genomes. Although the genome sequence of NRC-1 showed only 91 IS elements belonging to 12 families, the original results being skewed by the high GC content and the selection of restriction enzymes (especially EcoRI and HindIII with relatively AT-rich recognition sequences), the unusual genome structure and instability were still remarkable. The comparison of the genome of NRC-1 to the genome of R-1, the latter of which was reportedly isolated as a spontaneously occurring gas vesicle deficient variant of the wild-type NRC-1 strain, was still of great interest [16,17].

However, after a wait of nearly 8 years, the paper by Pfeiffer et al. reporting the genome sequence of R-1 published in a recent issue of *Genomics* [18] is, unfortunately, extremely disappointing because of what we believe are subjective and unfounded claims made regarding the NRC-1 genome assembly. Instead of a focus on the science with a comprehensive comparison of similarities and differences between the NRC-1 and R-1 genomes, the authors insinuate that the NRC-1 genome was not properly assembled, while at the same time ignoring the fact that a decade of cloning, mapping, and sequencing had clearly established both the structures and the rearrangements of these extrachromosomal replicons in strain NRC-1. These results have been documented in over a dozen peer-reviewed publications [11,19–33]. In so doing, they not only ignore the published record from our group, they also ignore resequencing of the NRC-1 genome by the 454 company published over 2 years ago, which confirmed the original conclusions [34].

Pfeiffer et al. characterize the differences between strains R-1 and NRC-1 as "vanishingly" small, and object to the name of the organism as a species of *Halobacterium*, since it suggests to them that it is a "distinct species." Indeed, our choice was to drop the species designation in the absence of definitive genomic data on *Halobacterium* isolates, a fact that was clearly stated on the title page footnote in the NRC-1 genome paper [6]:

"*Halobacterium* species are referred to in the literature by a variety of designations, including *H. halobium*, *H. cutirubrum*, *H. salinarium*, and *H. salinarum*. The precise relationships among these organisms and *Halobacterium* sp. strain NRC-1 are not entirely clear. Strain NRC-1

was a gift from W.F. Doolittle, Dalhousie University, Halifax, Canada. The strain has been deposited with the American Type Culture Collection, Manassas, VA (reference no. ATCC700922)."

The purpose of removing the species designation, contrary to the suggestion of Pfeiffer et al., was to clarify the differences between these related but distinct isolates. Given the long history of isolation of haloarchaea from salted food and leather, and the presence of many repeated elements in the genomes of these strains, it is not surprising that there are many related but distinct isolates. A recent study of prokaryotic taxonomy at the American Type Culture Collection indicated that there is as much diversity within the *Halobacterium* genus (among strains of *Halobacterium 'salinarum'*) as there is in the entire family of *Halobacteriaceae* and that NRC-1 is more distantly related to R-1 than some that are members of other genera [35,36]. Because there is no universally accepted standard for prokaryotic species designation, some clades have much more diversity than others. Among halophiles, some isolates have multiple divergent 16S rRNA genes in the same genome, while other diverse isolates have similar or identical 16S rRNA [37,38]. The fact that the true identities of NRC-1 and R-1 are quite unclear (see, e.g., [16,36]; also R.D. Simon and W.F. Doolittle, personal communications), our designation for NRC-1 was simply a way to show caution and objectivity to avoid confusion from naming and renaming a widely used strain.

One of the most important results from the R-1 genome sequence by Pfeiffer et al. [18] is that it has reaffirmed the high quality of the *Halobacterium* sp. NRC-1 genome sequence, which was completed by our consortium in 2000 [6,34]. Of the 2,014,239-bp large chromosome of NRC-1, only 4 single nucleotide polymorphisms and 5 single base and 3 large indels were found in the chromosomes of these 2 strains. Though not cited by Pfeiffer et al., the similarities were already established by extensive macrorestriction mapping work [30,39], as well as the cosmid mapping studies [40,41] over a dozen years ago. Although there are variations in the genetic arrangements and sequences of the plasmids, we consider these to be natural variations. Because the transpositions of insertion sequence elements (ISH) may result in insertions, deletions, and other types of DNA rearrangements, it would not be surprising if such events had occurred in the ISH rich plasmids and the data do not clearly establish whether both strains originated from the same isolate. As detailed above, DNA variations including plasmid DNA structural differences have already been well documented in previous studies.

On the question of assembly and misassembly, the Phred, Phrap, and Consed program suite used for the NRC-1 genome has been repeatedly shown to be very reliable for genome sequence assembly [42,43]. Because of the availability of an ordered library and detailed restriction map for pNRC100, ISH elements causing misassemblies
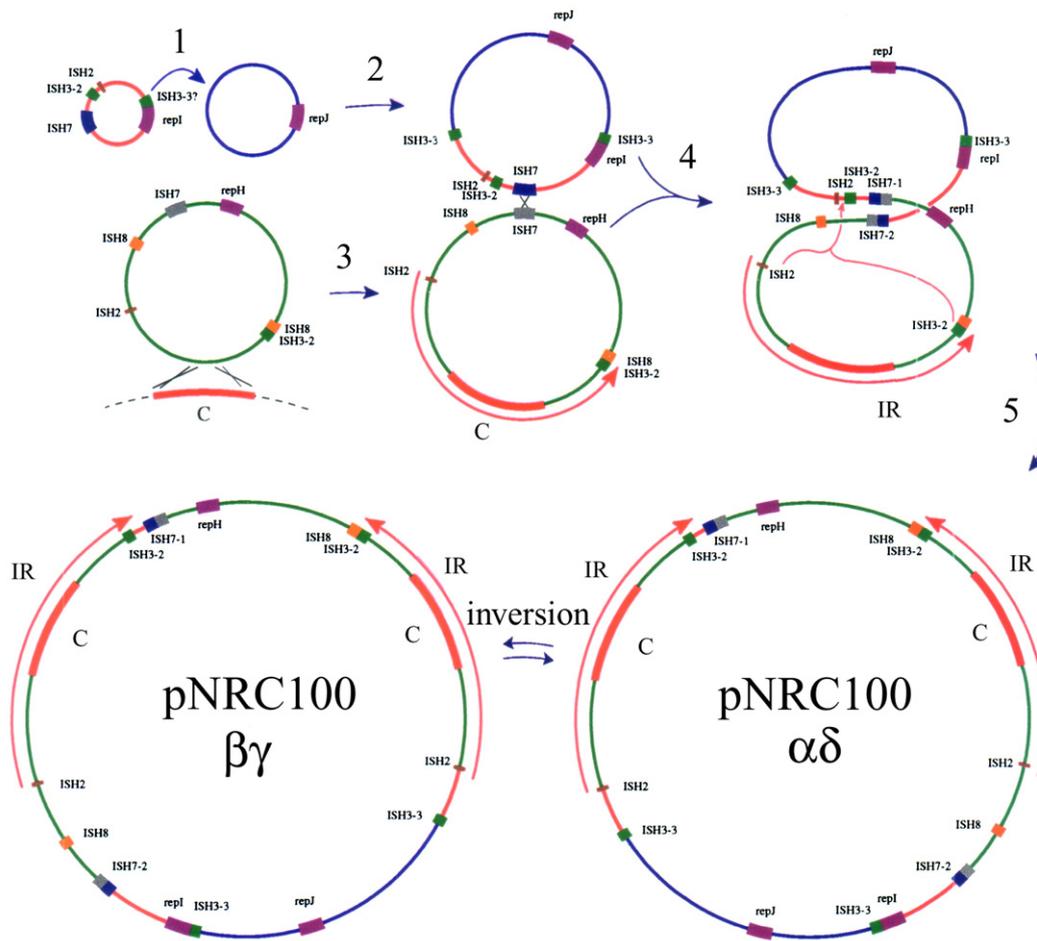

Fig. 1. Proposed mechanism of evolution of both inversion isomers of pNRC100. The mechanism of evolution of haloarchaeal megaplasmids likely involves a series of fusions of smaller replicons, each carrying a *repH*-like gene. The replicon fusions may have resulted from the intermolecular transposition of IS elements (e.g., depicted in steps 1 and 2) and/or by site-specific or homologous recombination between IS elements (step 4). Additional events proposed in pNRC100 evolution were recombinational exchanges leading to acquisition of chromosomal genes (step 3) and duplication of the acquired chromosomal region (step 5) to generate the large inverted repeats via unequal crossing over. Recombination between the inverted repeats was demonstrated experimentally to show the occurrence of inversion isomers [25] and recombination between IS elements was also detected to form minor circular DNAs [24,31]. Similar recombinational events likely led to replacement of a portion of the large single-copy region of pNRC100 with additional genes to form pNRC200. Three DNA replication genes (*repH, I,* and *J*) and a subset of IS elements are shown in colored boxes as is the chromosome-like region, labeled "C". The large inverted repeats are shown with pink arrows. Adapted from Ng et al. [11] published on the cover of *Genome Research* [11].

were noted and reported in our early assembly of the pNRC100 shotgun sequences [11]. However, we used a very simple but highly effective strategy to circumvent this problem. The simple vector sequence masking function in Phrap allowed successful correct assembly via efficient masking of repetitive elements in sequence assembly. We therefore eliminated the misassembly problem by using a combination of mapping and multistep assembly approaches. We named this method developed to assemble the NRC-1 genome the "hide-and-seek" sequence assembly strategy and also successfully applied it to the *Haloarcula marismortui* genome in 2004 [6,36].

For the quickly evolving megaplasmids (also called minichromosomes or extrachromosomal replicons) of strain NRC-1, pNRC100 and pNRC200, we concluded the plasmid map and sequence should be accomplished first in our genome project. Thus the pNRC100 restriction map was constructed by extensive restriction mapping of the purified plasmid and its recombinant HindIII fragment clones [25]. In the 1980s, using pulsed-field gel electrophoresis, we showed the presence of a 35,000- to 38,000-bp inverted repeat (IR) sequence [24]. Inversion isomers (Fig. 1) of pNRC100 had been demonstrated by Southern hybridization analysis using rare-cutting restriction enzymes (AflII and SfiI), which cut asymmetrically within the intervening small single-copy and large single-copy regions, respectively, but not within the large IRs. In this regard, the pNRC100 structure resembled some chloroplast and mitochondrial genomes which also contain large IRs [44]. In a deletion derivative of pNRC100 lacking one copy of the IRs, no inversion isomers were observed, indicating that both copies are required for inversion to occur and likely mediated by recombination between the large inverted repeat [31]. These early studies also established the identities and approximate positions of nearly a dozen IS elements in pNRC100, using Southern hybridization

and limited nucleotide sequence analysis across the IS element-target site junctions. Four prevalent classes of IS elements were initially found—ISH2, a 0.5-kb element; ISH3, a heterologous family of 1.4-kb elements; ISH4, a 1.0-kb element; and ISH8, a 1.4-kb element—representing a relatively small fraction of all the elements in the genome. The large IRs of pNRC100 terminated at an ISH2 element at one end and an ISH3 element at the other end. Subsequently the pNRC100 sequencing was conducted on shotgun libraries of plasmid and ordered HindIII fragment clones of pNRC100. This sequencing analysis of pNRC100 showed the presence of important genes, such as *trxAB* coding for thioredoxin and thioredoxin reductase, involved in protein reduction, and *cydAB* specifying cytochrome oxidase, important in aerobic respiration [11]. However, no rRNA genes were located in the pNRC IRs or elsewhere on the replicons. In the publication that reported the complete sequence of pNRC100 [11], an evolutionary model for how such a plasmid structure may have arisen, possibly in the laboratory or possibly in nature (Fig. 1), was provided and featured on the cover of *Genome Research*.

For pNRC200 a similar but less laborious approach was undertaken in the mid to late 1990s to map and end sequence cosmid clones of pNRC200, to establish its identity and structure. A HindIII clone library was created for pNRC200 and the structure of this plasmid was validated by sequencing [6,45].

The authors of the R-1 paper write as if they are unaware of the extensive previous work on the NRC-1 plasmids and the assembly strategy used for pNRC100 and the entire NRC-1 genome. Thus, their comments on NRC-1 plasmid sequence fidelity are not justified and should be reevaluated, especially given the extensive published and unpublished work performed to analyze both megaplasmids' structures and sequences. We believe strongly in our published, peer-
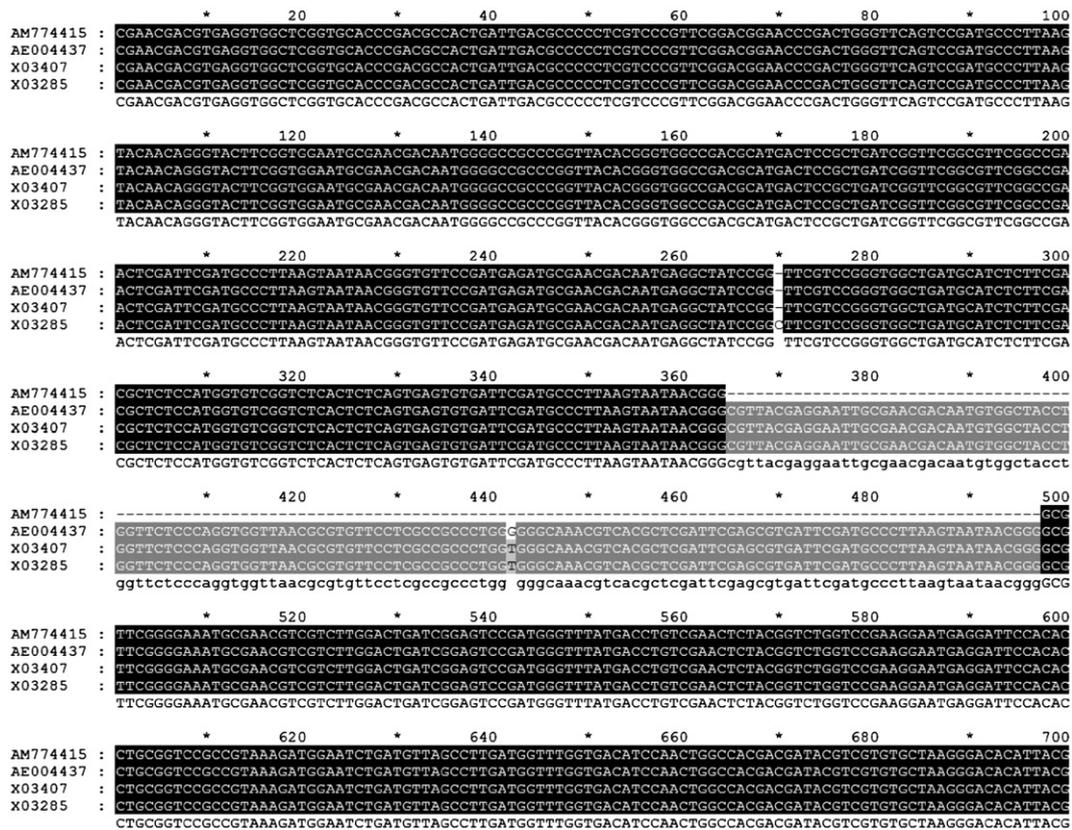

Fig. 2. Multiple sequence alignment of a 700-bp region containing the *Halobacterium* rRNA promoter The *Halobacterium salinarium* strain R-1 (AM774415: 1,863,001 to 1,863,566 bp), *Halobacterium* sp. NRC-1 (AE004437: 1,874,751 to 1,875,449 bp), *Halobacterium halobium* (X03407: 156 to 854 bp), and *Halobacterium cutirubrum* (X03285:155 to 854 bp) were aligned using ClustalX2.0.3 software [58]. The sequences present in AE004437, X03407, and X03285 but not in the *Halobacterium salinarium* R-1 strain are shaded in gray and the highly conserved sequences are shaded in black. A consensus sequence is shown at the bottom.

reviewed assembly of strain NRC-1′s megaplasmids and have already put forth a viable interpretation of the "creation" of pNRC100 from three smaller "hypothetical" plasmids including a segment from the large chromosome. Painstaking work was performed to understand and interpret the genome structure of *Halobacterium* sp. strain NRC-1, which should not be undervalued as Pfeiffer et al. attempt to do.

The application of a comparative genomics approach to enhance gene annotation is a powerful and appropriate strategy for the second genome and underscores the importance of availability of multiple related species in genome analysis. While ORPHEUS may arguably not be the best program for predicting genes, Pfeiffer et al. were able to eliminate most, if not all, of the "spurious" ORFs (6517) it predicted in the R-1 genome by using a comparative genomic and proteomic approach. Without the completion of NRC-1 and the other halophile genomes, the selection of the R-1 ORFs might have been very challenging even with the aid of proteomics data. In contrast, the Glimmer program efficiently predicted most of the *Halobacterium* sp. NRC-1 genes with a limited number of spurious predictions in the absence of comparative genomic data (see Supplementary Table S5 [18]). The difficult problem of accurate translation start site prediction for some genes, however, is largely unavoidable and this problem is still present in commonly used gene prediction programs [46–48]. Because NRC-1 was the first completed halophile genome, the advantage of extensive gene curation using a comparative genomic approach was not available at the time of the initial genome sequence, released in 2000. The gene prediction problem is likely not restricted to NRC-1 but to most of the initial genomes sequenced. However, it is the early release of the NRC-1 genome sequence, and not the R-1 sequence, that has nurtured the development of several important postgenomic and systems biology studies [8,10,49–56].

It would be helpful if all the true translation start sites were correctly predicted but it is not an absolute prerequisite for proteome analysis and mapping of the N-terminal peptides by mass spectrometry. Thus the Pfeiffer et al. claim of "identification of N-terminal peptides by proteomics would be hampered if the start codon is misassigned" is moot. A proteomics analysis should not be deferred because the above issue can be easily solved by mapping the MS/MS spectra against six-frame translated protein sequences database, which is already a built-in function in programs such as SEQUEST [57]. Thus it is still feasible to identify the N-terminal peptides and curate the ORF sequences. Regarding the proteome analysis, the statement "a whole class of proteins was discovered that has not been studied in *Hbt. salinarium* or any other species yet" is controversial. An earlier *Halobacterium* sp. NRC-1 proteome analysis had identified more than 400 proteins in the membrane and soluble proteomes in 2003 [49]. A subsequent in-depth soluble proteome analysis that identified approximately 888 proteins was documented in 2006 [52].

Given that we wanted to rapidly respond to what we believe are shortcomings in the Pfeiffer et al. R-1 genome sequence report, we have not yet fully analyzed strain R-1 and cannot comment on the validity of genome assembly or annotation. To be cautious, we have not made any comments on the sequence variations between NRC-1 and R-1 strains unless third-party sequences were available. However, we have analyzed Table 2 of the R-1 paper; the readers should be careful about the row "insertion/deletion 3" where "113 additional bases in NRC-1 affecting rRNA promoter region" are described. At least 2 sequences covering this region from *H. halobium* NRC817 (Accession No. X03407) and *H. cutirubrum* (X03285) had been determined and deposited in GenBank. Using the ClustalX2.0.3 program to perform a multiple sequence analysis [58], it became clear that the "133 additional bases" present in NRC-1 are also conserved in these two sequences (Fig. 2). Because this segment is within the promoter region, it may be useful to reconsider what might have happened to this DNA segment in the R-1 genome assembly.

As a whole, research groups interested in the biology and evolution of halophilic archaea should be excited and overjoyed to have a second complete genome sequence available to the scientific community. As one such research group, we are pleased with the research opportunities a second *Halobacterium* sp. genome provides, although we are quite dismayed at the disparaging, subjective, and inappropriate comments regarding the first genome sequence of a haloarchaeon, *Halobacterium* sp. strain NRC-1. The authors here believe that science should be and is critical, precise, and objective. Here, we have presented these comments and addressed the concerns raised in Pfeiffer et al. regarding the genome assembly procedures used, the gene prediction methodologies, and the proteomic pitfalls for *Halobacterium* sp. strain NRC-1. As discussed above, the lineage and history of specific haloarchaeal isolates have been muddled and continue to be confusing and points of contention for many labs studying this group of organisms [36]. Specifically, neither we nor Pfeiffer et al. can be certain of *Halobacterium* sp. strain NRC-1′s origins and therefore we all should withhold judgment regarding the lineages and histories of strains R-1 and NRC-1. Clearly, both genome sequences can be placed within the genus of *Halobacterium* (based upon nearly identical large chromosome sequences). Apparent also from both genome sequences is that the makeup and genome structure of extrachromosomal replicons are different in these two strains. Whether that is a result of evolution inside or outside a laboratory setting is unknown, but it is certainly not a result of misassembly of the *Halobacterium* sp. strain NRC-1 genome sequence. We believe that this specific point has already been amply demonstrated and will continue to be borne out through time and additional research in the haloarchaeal community.

References

[1] E. Pennisi, Breakthrough of the year. Genomics comes of age, Science 290 (2000) 2220–2221.

[2] K.E. Nelson, I.T. Paulsen, C.M. Fraser, Microbial genome sequencing: a window into evolution and physiology, ASM News 67 (2001) 310–317.

[3] S. McCready, et al., UV irradiation induces homologous recombination genes in the model archaeon, *Halobacterium* sp. NRC-1, Saline Systems 1 (2005) 3.

[4] S. DasSarma, P. Arora, Halophiles, Encyclopedia of Life Sciences, Wiley, Chichester, 2006.

[5] L.C. DeVeaux, et al., Extremely radiation-resistant mutants of a halophilic archaeon with increased single-stranded DNA-binding protein (RPA) gene expression, Radiat. Res. 168 (2007) 507–514.

[6] W.V. Ng, et al., Genome sequence of *Halobacterium* species NRC-1, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 12176–12181.

[7] R.F. Peck, S. DasSarma, M.P. Krebs, Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker, Mol. Microbiol. 35 (2000) 667–676.

[8] S. DasSarma, B.R. Berquist, J.A. Coker, P. DasSarma, J.A. Müller, Post-genomics of the model haloarchaeon *Halobacterium* sp. NRC-1, Saline Systems 2 (2006) 3.

[9] The HaloWeb database, http://halo2.umbi.umd.edu/cgi-bin/scripts/form2.pl and http://halo3.umbi.umd.edu/cgi-bin/scripts/form2.pl.

[10] S. DasSarma, Extreme microbes, American Scientist 95 (2007) 224–231.

[11] W.V. Ng, et al., Snapshot of a large dynamic replicon in a halophilic archaeon: megaplasmid or minichromosome? Genome Res 8 (1998) 1131–1141.

[12] S.P. Kennedy, W.V. Ng, S.L. Salzberg, L. Hood, S. DasSarma, Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence, Genome Res. 11 (2001) 1641–1650.

[13] The HaloLex database, www.halolex.mpg.de/public/.

[14] C. Sapienza, W.F. Doolittle, Unusual physical organization of the *Halobacterium* genome, Nature 295 (1982) 384–389.

[15] C. Sapienza, M.R. Rose, W.F. Doolittle, High-frequency genomic rearrangements involving archaebacterial repeat sequence elements, Nature 299 (1982) 182–185.

[16] W. Stoeckenius, R.H. Lozier, R.A. Bogomolni, Bacteriorhodopsin and the purple membrane of halobacteria, Biochim. Biophys. Acta 505 (1979) 215–278.

[17] B.R. Berquist, J.A. Müller, S. DasSarma, Genetic systems for halophilic archaea, in: A. Oren, F. Rainey (Eds.), Extremophiles, Elsevier/Academic Press, 2006, pp. 649–680.

[18] F. Pfeiffer, S.C. Schuster, A. Broichera, M. Falb, P. Palm, K. Rodewalda, A. Ruepp, J. Soppa, J. Tittor, D. Oesterhelt, Evolution in the laboratory: The genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1, Genomics 91 (2008) 335–346.

[19] S. DasSarma, T. Damerval, J.G. Jones, N. Tandeau de Marsac, A plasmid-encoded gas vesicle protein gene in a halophilic archaebacterium, Mol. Microbiol. 1 (1987) 365–370.

[20] S. DasSarma, et al., High-frequency mutations in a plasmid-encoded gas vesicle gene in *Halobacterium halobium*, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 6861–6865.

[21] S. DasSarma, Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations, Can. J. Microbiol. 35 (1989) 65–72.

[22] J.G. Jones, et al., Analysis of insertion mutants reveals two new genes in the pNRC100 gas vesicle gene cluster of *Halobacterium halobium*, Nucleic Acids Res. 17 (1989) 7785–7793.

[23] J.G. Jones, D.C. Young, S. DasSarma, Structure and organization of the gas vesicle gene cluster on the *Halobacterium halobium* plasmid pNRC100, Gene 102 (1991) 117–122.

[24] W.-L. Ng, S. DasSarma, Physical and genetic mapping of the unstable gas vesicle plasmid in *Halobacterium halobium* NRC-1, in: F. Rodriguez-Valera (Ed.), Genetic and applied aspects of halophilic microorganisms, Plenum Press, 1991, pp. 305–311.

[25] 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) 3933.

[26] J.T. Halladay, W.-L. Ng, S. DasSarma, Genetic transformation of a halophilic archaebacterium with a gas vesicle gene cluster restores its ability to float, Gene 119 (1992) 131–136.

[27] S. DasSarma, Identification and analysis of the gas vesicle gene cluster on an unstable plasmid of *Halobacterium halobium*, Experientia 49 (1993) 482–486.

[28] J.T. Halladay, J.G. Jones, F. Lin, A.B. MacDonald, S. DasSarma, The rightward gas vesicle operon in *Halobacterium* plasmid pNRC100: identification of the *gvpA* and *gvpC* gene products by use of antibody probes and genetic analysis of the region downstream of *gvpC*, J. Bacteriol. 175 (1993) 684–692.

[29] W.-L. Ng, S. DasSarma, Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100, J. Bacteriol. 175 (1993) 4584–4596.

[30] Y. Bobovnikova, W.-L. Hg, S. DasSarma, N.R. Hackett, Restriction mapping the genome of *Halobacterium halobium* strain NRC-1, Syst. Appl. Microbiol. 16 (1994) 97–104.

[31] W.-L. Ng, P. Arora, S. DasSarma, Large deletions in class III gas vesicle-deficient mutants of *Halobacterium halobium*, Syst. Appl. Microbiol. 16 (1994) 560–568.

[32] S. DasSarma, Natural plasmids and plasmid vectors of halophiles, in: S. Dassarma, E.M. Fleischmann (Eds.), Halophiles, Cold Spring Harbor Laboratory Press, 1995, pp. 243–250.

[33] S. DasSarma, P. Arora, Genetic analysis of the gas vesicle gene cluster in haloarchaea, FEMS Microbiol. Lett. 153 (1997) 1–10.

[34] R. Pinard, et al., Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing, BMC Genomics 7 (2006) 216.

[35] D. Cleland, P. Krader, D. Emerson, Use of the DiversiLab repetitive sequence-based PCR system for genotyping and identification of Archaea, J. Microbiol. Methods 73 (2008) 172–178.

[36] P. DasSarma, S. DasSarma, On the origin of prokaryotic "species": the taxonomy of halophilic Archaea, Saline Systems 4 (2008) 5.

[37] N.S. Baliga, et al., Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea, Genome Res. 14 (2004) 2221–2234.

[38] Y.A. Goo, et al., Low-pass sequencing for microbial comparative genomics, BMC Genomics 5 (2004) 3.

[39] N.R. Hackett, Y. Bobovnikova, N. Heyrovska, Conservation of chromosomal arrangement among three strains of the genetically unstable archaeon *Halobacterium salinarium*, J. Bacteriol. 176 (1994) 7711–7718.

[40] A. St Jean, B.A. Trieselmann, R.L. Charlebois, Physical map and set of overlapping cosmid clones representing the genome of the archaeon *Halobacterium* sp. GRB, Nucleic Acids Res. 22 (1994) 1476–1483.

[41] A. St Jean, R.L. Charlebois, Comparative genomic analysis of the *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB contig maps reveals extensive rearrangement, J. Bacteriol. 178 (1996) 3860–3868.

[42] B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res. 8 (1998) 186–194.

[43] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, Genome Res. 8 (1998) 175–185.

[44] M.W. Gray, Evolution of organellar genomes, Curr. Opin. Genet. Dev. 9 (1999) 678–687.

[45] S. Kennedy, Understanding genome structure, function, and evolution in the halophilic archaeon *Halobacterium* NRC-1, Department of Microbiology, University of Massachusetts, Amherst, 2003, p. 171.

[46] A.C. McHardy, A. Goesmann, A. Puhler, F. Meyer, Development of joint application strategies for two microbial gene finders, Bioinformatics 20 (2004) 1622–1631.

[47] P. Nielsen, A. Krogh, Large-scale prokaryotic gene prediction and comparison to genome annotation, Bioinformatics 21 (2005) 4322–4329.

[48] H. Zhu, G.Q. Hu, Y.F. Yang, J. Wang, Z.S. She, MED: a new non-supervised gene prediction algorithm for bacterial and archaeal genomes, BMC Bioinformatics 8 (2007) 97.

[49] Y.A. Goo, et al., Proteomic analysis of an extreme halophilic archaeon, *Halobacterium* sp. NRC-1. Mol. Cell Proteomics 2 (2003) 506–524.

[50] N.S. Baliga, et al., Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium* NRC-1, Genome Res. 14 (2004) 1025–1035.

[51] J.A. Müller, S. DasSarma, Genomic analysis of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine N-oxide as terminal electron acceptors, J. Bacteriol. 187 (2005) 1659–1667.

[52] R.R. Gan, et al., Proteome analysis of *Halobacterium* sp. NRC-1 facilitated by the biomodule analysis tool BMSorter, Mol. Cell. Proteomics 5 (2006) 987–997.

[53] R. Bonneau, et al., A predictive model for transcriptional control of physiology in a free living cell, Cell 131 (2007) 1354–1365.

[54] J.A. Coker, P. DasSarma, J. Kumar, J.A. Müller, S. DasSarma, Transcriptional profiling of the model Archaeon *Halobacterium* sp. NRC-1: responses to changes in salinity and temperature, Saline Systems 3 (2007) 6.

[55] J.A. Coker, S. DasSarma, Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA, BMC Genet. 8 (2007) 61.

[56] A.K. Schmid, et al., The anatomy of microbial cell state transitions in response to oxygen, Genome Res. 17 (2007) 1399–1413.

[57] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino sequences in protein database, J. Am. Soc. Mass Spectrom. 5 (1994) 976–989.

[58] R. Chenna, et al., Multiple sequence alignment with the Clustal series of programs, Nucleic Acids Res 31 (2003) 3497–3500.

Wailap Victor Ng, Timothy H. Wu
Department of Biotechnology and Laboratory Science in Medicine,
Institute of Biomedical Informatics,
National Yang Ming University,
Taipei 112, Taiwan, ROC

Brian R. Berquist
Unit of Structure and Function in Base Excision Repair,
National Institute on Aging,
National Institutes of Health,
Baltimore, MD 21224, USA

James A. Coker, Melinda Capes,
Priya DasSarma,
Shiladitya DasSarma*
University of Maryland Biotechnology Institute,
Center of Marine Biotechnology,
Baltimore MD 21202, USA
*Corresponding author.

27 March 2008