Minireview

Extreme genomes

Edward F DeLong

Address: Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA. E-mail: delong@mbari.org

Abstract

The complete genome sequence of *Thermoplasma acidophilum*, an acid- and heat-loving archaeon, has recently been reported. Comparative genomic analysis of this 'extremophile' is providing new insights into the metabolic machinery, ecology and evolution of thermophilic archaea.

Genomics has invigorated diverse biological fields, from biomedicine to biophysics to evolutionary biology. Important biomedical goals and drivers have done much to shape contemporary genome biology, but fundamental biological questions also play a central role in genomics. The considerable effort directed towards completing microbial genome sequences across a broad range of phylogenetically diverse microbial lineages, for example, has been an important strategic approach in the emerging field of comparative microbial genomics. Interestingly, some of the first whole-genome sequences have been derived from microorganisms that thrive in 'extreme' habitats of elevated temperature, low pH, or high salinity. At least ten complete genome sequences are now available from such archaeal and bacterial 'extremophiles'. A recent addition to the list of completed genome sequences is that of *Thermoplasma acidophilum*, a thermoacidophilic archaeon [1,2]. This newly completed whole-genome sequence adds interesting and important grist to the mill of comparative microbial genomics, shedding light not only on metabolism in extreme conditions but also on the evolution and ecology of archaea.

The extreme team

The limits of temperature, pressure, pH, and water activity at which life can thrive are not yet precisely defined. It is now well documented, however, that microbes can flourish in settings that chemically and physically challenge the very fabric of life. Microbial life is now known to exist at temperatures of at least 113°C, at pH values approaching zero, and at salinities found in saturated brines. The majority of microbial record holders for growth at extremes all belong to the domain Archaea. Their unusual phenotypic properties, unique evolutionary relationships, remarkable adaptations, and small genome size have made these microbes a natural focus of whole-genome sequencing efforts, ever since the method was developed. To date, at least ten complete genome sequences from archaeal and bacterial extremophiles have been reported (Table 1) [1,3-11]. Many more genome sequences from similar microbes are either complete and undergoing annotation or are currently being sequenced [12].

Thermoplasma acidophilum

Imagine a discarded coal pile in a rural Indiana setting, burning slowly and steadily over time, like an unattended barbecue. This seems an unlikely place to look for life. Yet this is just the habitat from which the acid-loving archaeon *T. acidophilum* was first isolated by Darland and colleagues [2], in the late 1960s. *T. acidophilum* is an unusual microorganism from a variety of perspectives. It grows at very low pH values, ranging from 0.8 to 4.0, and at temperatures ranging from 45 to 63°C. Unlike most microorganisms, *T. acidophilum* has no cell wall - its cytoplasm is separated from the outside world by a thin membrane composed of phytanyl and biphytanyl ether lipids bound to a glycerol backbone. This membrane lipid composition, along with the evolutionary affinities of its ribosomal RNAs and some conserved proteins, identify *T. acidophilum* as an archaeon. Although *T. acidophilum* appears to fall within the archaeal subgroup Euryarchaeota, it is only distantly related to the extremely halophilic and methanogenic archaea. Interestingly, a number

Table 1

Extremophiles with completely sequenced genomes

Organism	Domain*	Genome size (base pairs)	Phenotype†	Reference
Methanococcus jannaschii	AE	1,644,970	H	[3]
Methanobacterium thermoautotrophicum	AE	1,751,377	T	[4]
Archaeoglobus fulgidus	AE	2,178,400	H	[5]
Aquifex aeolicus	B	1,551,335	H	[6]
Pyrococcus horikoshii	AE	1,738,505	H	[7]
Thermotoga maritima	B	1,860,725	H	[8]
Aeropyrum pernix	AC	1,699,695	H	[9]
Thermoplasma acidophilum	AE	1,564,905	T	[1]
Halobacterium strain NRC-1	AE	2,014,239	EH	[10]
Pyrococcus abyssii	AE	1,765,505	H	[11]

*Domains: A, Archaea; E, Euryarchaeota; C, Crenarchaeota; B, Bacteria.
†Phenotypes: H, hyperthermophile (growth optima > 80°C);
T, thermophile (growth optima > 40°C); EH, extreme halophile (absolute growth requirement for NaCl > 2 M NaCl).

of new relatives of *T. acidophilum* have recently been identified: some grow at pH values approaching zero [13], some live at lower temperatures and low pH [14], and some thrive at near-neutral pH in cold, aerobic marine habitats [15].

Using different methods from those of most genome sequencing projects, Ruepp *et al.* [1] have recently published the full genome sequence of *T. acidophilum*. The most frequently employed method for acquiring whole genome sequences is the whole-genome 'shotgun' strategy [16]. This now-standard approach couples random shotgun cloning and sequencing of large numbers of small DNA fragments (approximately sevenfold genome coverage in fragments of 1-3 kilobases in size) with subsequent assembly *in silico* to derive contiguous genome sequence. Final gap filling is achieved by sequencing large-fragment clones that form an effective scaffold around the genome, by using PCR to bridge gaps, or by other similar methods. Instead of this approach, Ruepp *et al.* [1] used a more directed approach, 'shotgun primer walking', to obtain the complete *T. acidophilum* genome sequence. This approach combines a preliminary shotgun sequencing effort with constant monitoring to reduce redundant sequencing effort. Subsequent PCR amplification and primer-walking strategies then fill remaining gaps. The approach taken by Ruepp *et al.* required 7,855 sequencing reactions to cover the 1.56 megabase *T. acidophilum* genome, about a third of the number of reactions that would have been required using a standard shotgun sequencing strategy for a similiarly sized genome [1]. The trade-off in the approach is the effort required constantly to monitor results and adjust the sequencing strategy, so as to

maintain sequence quality in the face of the overall reduced redundancy. Ruepp *et al.* [1] point out, however, that high-quality genome sequence data could be obtained with the method, even with just twofold sequence coverage.

Ruepp *et al.* [1] found that the single, circular, 1,564,905 base-pair genome of *T. acidophilum* contained 1,509 open reading frames (ORFs), one third of which had homologs present in all three domains of life. At a density of about one ORF per kilobase pair, genes appear as densely packed in *T. acidophilum* as they are in other archaeal and bacterial genomes. About 16% of the genome, or 240 ORFs, were 'singletons', having no recognizable homologs in the current databases. About 55% (823 ORFs) could be assigned function on the basis of similarity to annotated database sequences. In terms of representation in functional categories, only transport facilitation proteins were notably unusual, comprising a relatively larger proportion of the *T. acidophilum* genome than is found in other sequenced archaeal genomes.

As in other genome studies, what was missing from the *T. acidophilum* genome was perhaps more surprising than what was present. Although *T. acidophilum* can respire anaerobically using sulfur as a terminal electron acceptor, no sulfur-respiratory genes similar to those of the archaeon *Archaeoglobus fulgidus* were found. Instead, bacterium-like sulfur-reducing proteins were identified, and these appear to be responsible for sulfur metabolism in *T. acidophilum*. Additionally, although *T. acidophilum* has flagella and is motile, no recognizable chemotaxis proteins were found. The same is true of a number of other microbes (*Methanococcus jannaschii, Aeropyrum pernix* and *Aquifex aeolicus*, for example), and suggests that the chemotactic signal transduction proteins in these organisms are likely to be found among functionally unassigned ORFs.

Evolutionary considerations
The origin of the eukaryal nucleus remains a mystery, but a number of hypotheses have been proposed, mostly centered on a common theme. This theme derives from the fact that proteins involved in transcription and translation often specifically link archaea with eukarya, to the exclusion of bacteria. This molecular link between eukarya and archaea led to the general notion that the eukaryal nucleus may have resulted from the symbiotic engulfment of an archaeon by a protoeukaryal organism [17]. More recently, it has been proposed that a thermoacidophile, specifically resembling *Thermoplasma*, was the archaeon that contributed genetic material to the evolving eukaryal nucleus [18,19]. The *Thermoplasma*-like archaeon was suggested to contribute, among other things, chromatin and nucleosomal organization, as well as cytoskeletal fibers, as a consequence of the endosymbiotic association [18]. The data of Ruepp *et al.* [1] provide no support for the hypothesis that a *Thermoplasma*-like

ancestor contributed to the formation of the eukaryal nucleus. No eukaryal features unique to *T. acidophilum*, such as proteins of the nuclear pore complex, the exocyst, or the cytoskeleton, were found in the archaeon's genome sequence [1].

Lateral gene transfer is an increasingly hot topic in the field of comparative genome analyses. Interestingly, the genome of *T. acidophilum* has a relatively high number of ORFs (252, or 17%) that were most similar to those found in a single organism, the crenarchaeote *Sulfolobus solfataricus*. This result is unexpected, as the full genome sequence of *S. solfataricus* is not yet available, whereas the completed genomes of quite a few other archaea (crenarchaeotes and euryarchaeotes) were available for comparison. Why are 17% of the protein-encoding genes of *T. acidophilum* most similar to a single distantly related crenachaeote, *S. solfataricus*? Ruepp *et al.* [1] propose that lateral gene transfer is the answer. Gene exchange between these microbes is presumably facilitated by their adaptation to the same unique niche (a high temperature, low pH environment), which is occupied by few other species. A similar argument may hold for two hyperthermophilic bacteria, *Aquifex aeolicus* and *Thermotoga maritima*, each with 20-25% of their genes being more similar to archaeal homologs than to bacterial ones [6,8]. The lesson is intuitive: organisms that live together swap genes at higher frequency. It appears that ecology and genomics, disciplines that on the surface appear quite removed from one another, are beginning to converge. It may well be that environmental microbiology will contribute importantly to our understanding of gene distributions among different microbial genomes. Conversely, whole-genome analyses may provide significant information about the population dynamics and specific interactions of naturally ocurring microbial species.

References

1. Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W: **The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*.** *Nature* 2000, **407:**508-513.
2. Darland G, Brock TD, Samsonoff W, Conti SF: **A thermoacidophilic mycoplasma isolated from a coal refuse pile.** *Science* 1970, **170:**1416-1418.
3. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, *et al.*: **Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.** *Science* 1996, **273:**1058-1073.
4. Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, *et al.*: **Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics.** *J Bacteriol* 1997, **179:**7135-7155.
5. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, *et al.*: **The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*.** *Nature* 1997, **390:**364-370.
6. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, *et al.*: **The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.** *Nature* 1998, **392:**353-358.
7. Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, *et al.*: **Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3.** *DNA Res* 1998, **5:**55-156.
8. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, *et al.*: **Evidence for lateral gene transfer between Archaea and bacteria from the genome sequence of *Thermotoga maritima*.** *Nature* 1999, **399:**323-329.
9. Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, *et al.*: **Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1.** *DNA Res* 1999, **6:**83-101, 145-52.
10. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pana M, Shuklac HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, *et al.*: **Genome sequence of *Halobacterium* species NRC-1.** *Proc Natl Acad Sci USA* 2000, **97:**12176-12181.
11. ***Pyrococcus abyssi* home page at Genoscope** [http://www.genoscope.cns.fr/Pab/]
12. **TIGR Microbial Database: a listing of microbial genomes and chromosomes in progress** [http://www.tigr.org/tdb/mdb/mdbinprogress.html]
13. Schleper C, Puhler G, Kuhlmorgen B, Zillig W: **Life at extremely low pH.** *Nature* 1995, **375:**741-742.
14. Golyshina OV, Pivovarova TA, Karavaiko GI, Kondrateva TF, Moore ER, Abraham WR, Lunsdorf H, Timmis KN, Yakimov MM, Golyshin PN: ***Ferroplasma acidiphilum* gen. nov., sp. nov., an acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic member of the Ferroplasmaceae fam. nov., comprising a distinct lineage of the Archaea.** *Int J Syst Evol Microbiol* 2000, **50:**997-1006.
15. DeLong EF: **Archaea in coastal marine environments.** *Proc Natl Acad Sci USA* 1992, **89:**5685-5689.
16. Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, *et al.*: **The minimal gene complement of *Mycoplasma genitalium*.** *Science* 1995, **270:**397-403.
17. Sogin ML: **Early evolution and the origin of eukaryotes.** *Curr Opin Genet Dev* 1991, **1:**457-463.
18. Margulis L: **Archaeal-eubacterial mergers in the origin of Eukarya: phylogenetic classification of life.** *Proc Natl Acad Sci USA* 1996, **93:**1071-1076.
19. Margulis L, Dolan MF, Guerrero R: **The chimeric eukaryote: origin of the nucleus from the karyomastigont in amitochondriate protists.** *Proc Natl Acad Sci USA* 2000, **97:**6954-6959.