

RESEARCH ARTICLE

# Complete Genome Sequence of *Thermus aquaticus* Y51MC23

Phillip J. Brumm<sup>1\*</sup>, Scott Monsma<sup>2</sup>, Brendan Keough<sup>2</sup>, Svetlana Jasinovica<sup>2</sup>, Erin Ferguson<sup>2</sup>, Thomas Schoenfeld<sup>2</sup>, Michael Lodes<sup>2</sup>, David A. Mead<sup>2</sup>

**1** C5-6 Technologies LLC, Fitchburg, Wisconsin, United States of America, **2** Lucigen Corporation, Middleton, Wisconsin, United States of America

\* [pbrumm@c56technologies.com](mailto:pbrumm@c56technologies.com)



**OPEN ACCESS**

**Citation:** Brumm PJ, Monsma S, Keough B, Jasinovica S, Ferguson E, Schoenfeld T, et al. (2015) Complete Genome Sequence of *Thermus aquaticus* Y51MC23. PLoS ONE 10(10): e0138674. doi:10.1371/journal.pone.0138674

**Editor:** Hodaka Fujii, Osaka University, JAPAN

**Received:** April 22, 2015

**Accepted:** September 2, 2015

**Published:** October 14, 2015

**Copyright:** © 2015 Brumm et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequence files are available from the GenBank database (<http://www.ncbi.nlm.nih.gov/genbank/>) (Chr.sqn Chr, CP010822; Chr.sqn pTA14, CP010823; Chr.sqn pTA16, CP010824; Chr.sqn pTA69, CP010825; Chr.sqn pTA78, CP010826).

**Funding:** This project was supported by an SBIR grant award (1R43HG007797-01A1) from the National Human Genome Research Institute. C5<sup>6</sup> Technologies LLC, and Lucigen Corporation provided support in the form of salaries for authors PB, SM, BK, SJ, EF, TS, ML & DM, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the

## Abstract

*Thermus aquaticus* Y51MC23 was isolated from a boiling spring in the Lower Geyser Basin of Yellowstone National Park. Remarkably, this *T. aquaticus* strain is able to grow anaerobically and produces multiple morphological forms. Y51MC23 is a Gram-negative, rod-shaped organism that grows well between 50°C and 80°C with maximum growth rate at 65°C to 70°C. Growth studies suggest that Y51MC23 primarily scavenges protein from the environment, supported by the high number of secreted and intracellular proteases and peptidases as well as transporter systems for amino acids and peptides. The genome was assembled *de novo* using a 350 bp fragment library (paired end sequencing) and an 8 kb long span mate pair library. A closed and finished genome was obtained consisting of a single chromosome of 2.15 Mb and four plasmids of 11, 14, 70, and 79 kb. Unlike other *Thermus* species, functions usually found on megaplasmids were identified on the chromosome. The Y51MC23 genome contains two full and two partial prophage as well as numerous CRISPR loci. The high identity and synteny between Y51MC23 prophage 2 and that of *Thermus* sp. 2.9 is interesting, given the 8,800 km separation of the two hot springs from which they were isolated. The anaerobic lifestyle of Y51MC23 is complex, with multiple morphologies present in cultures. The use of fluorescence microscopy reveals new details about these unusual morphological features, including the presence of multiple types of large and small spheres, often forming a confluent layer of spheres. Many of the spheres appear to be formed not from cell envelope or outer membrane components as previously believed, but from a remodeled peptidoglycan cell wall. These complex morphological forms may serve multiple functions in the survival of the organism, including food and nucleic acid storage as well as colony attachment and organization.

## Introduction

*Thermus aquaticus* YT-1 holds a special place in the history of microbiology. The thermophile was first isolated and cultured from a hot spring in Yellowstone National Park in 1969 [1]. The discovery of life at high temperatures was controversial at that time, but later shown to be quite

manuscript. The specific roles of these authors are articulated in the 'author contributions' section.

**Competing Interests:** Authors SM, BK, SJ, EF, TS, ML & DM were employees of Lucigen at the time of the study and own stock or stock options in Lucigen. PB was an employee of C5-6 Technologies, Inc at the time of the study and owned stock and stock options in the company. C5-6 Technologies, Inc was dissolved in December 2014 by the board of directors and no longer exists. No patents or products were developed from this work. This does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

prevalent as demonstrated by the isolation of *Thermus* strains from hot water heaters and other sources [2]. The subsequent discovery and characterization of *Thermus aquaticus* DNA polymerase resulted in the development of amplification and sequencing tools that have revolutionized nearly every field of biology and medicine [3].

*Thermus* species that have been isolated from hot springs around the globe include *T. Brockianus*, *T. thermophilus* [3], *T. oshimai* [4], *T. caliditerrae* [5], *T. arciformis* [6], *T. islandicus* [7], *T. igniterrae* [8] and *T. antranikianii* [8]. Not all *Thermus* species have been found in hot springs. *T. composti* was isolated from an oyster mushroom compost [9] and *T. scotoeductus* from a South African gold mine [10]. A number of isolates initially classified as *Thermus* species have been reclassified as *Meiothermus* species [11], based on phylogenetic and physiological differences such as lower optimum growth temperatures.

A unique feature of *Thermus aquaticus* Y51MC23 is its unusual cellular morphology. Shortly after the initial *T. aquaticus* report, Brock described the presence of "rotund bodies", which appeared to be an association of multiple cells connected by a combined outer envelope as visualized by electron microscopy [12]. Little is known about the prevalence or function of these rotund bodies; however, the only other organism reported to form them is *Meiothermus ruber* (formerly *Thermus ruber*) [13–15]. Because of the low levels of rotund bodies observed in cultures, it is unclear if their production is limited to *T. aquaticus* and a few other related species, or if it is a trait shared by all *Thermus* species. Micrographs demonstrating the remarkable morphological diversity of *Thermus aquaticus* Y51MC23 are presented here.

There were 28 *Thermus* genome projects as of April 2015 (Genomes online database, <https://gold.jgi-psf.org/index>), 23 of which have retrievable sequence data. Complete genome sequences have been reported for *T. thermophilus* HB8 [16], *T. thermophilus* HB27 [17], *T. scotoeductus* SA-01 [10], *T. oshimai* JL-2 and *T. thermophilus* JL-18 [18, 19]. Most *Thermus* genomes have been left unfinished in permanent draft status (17/23), including *Thermus* sp. strain RL [20], *Thermus* sp. strain CCB\_US3\_UF1 [21], *Thermus* sp. 2.9 [22], *T. thermophilus* ATCC 33923 [23], and *T. aquaticus* Y51MC23 (<http://www.ncbi.nlm.nih.gov/nuccore/218297404>). The ability to close and finish microbial genomes is constrained by the lack of tools for connecting short read sequence data across long repeats. This paper describes the complete genome sequence for *Thermus aquaticus* Y51MC23 and its four plasmids, the first for this species, which was closed and finished using a new mate pair library construction tool. This complete genome is a valuable reference for the comparative analysis of the organization and structure of *Thermus* genomes.

## Methods

### Isolation, Growth Conditions and DNA Isolation

*Thermus aquaticus* Y51MC23 (Y51MC23) was isolated from a sample of hot spring water by enrichment and plating on YTP-2 agar at 70°C [24] and subsequently maintained on YTP-2 agar plates. The sample was collected under Scientific Research and Collecting Permit YELL-2001-SCI-0221 issued by the US Department of the Interior National Park Service, Yellowstone National Park. The culture is available from the ATCC™. For preparation of genomic DNA, cultures of Y51MC23 were grown from a single colony in 1000 ml YTP-2 medium in a 2000 ml Erlenmeyer flask at 70°C, 200 rpm for 18 hours. Cells were collected by centrifugation at 4°C and stored frozen until used for DNA preparation. The cell concentrate was lysed using a combination of SDS and proteinase K, and genomic DNA was isolated using a phenol/chloroform extraction method [25].

Aerobic growth was performed using either 50 ml YTP-2 media in a 250 ml flask, or 1000 ml YTP-2 media in a 2000 ml flask, at 70°C, 200 rpm, with silicone foam closure.

Anaerobic growth was performed using YTP-2 medium with nitrate omitted. Cultures were grown in either 50 ml of media in 50 ml conical, screw cap tube or 1000 ml media in 1000 ml screw cap bottle, at 70°C with no agitation.

## Microscopy

Culture samples were treated using either a 5  $\mu$ M solution of SYTO<sup>®</sup> 9 fluorescent stain, Live-Dead<sup>®</sup> Stain (SYTO<sup>®</sup> 9 and propidium bromide), or ViaGram<sup>™</sup> Red<sup>+</sup> Bacterial Gram Stain and Viability Kit (SYTOX<sup>®</sup> Green nucleic acid stain, 4,6-diamidino-2-phenylindole and Texas Red<sup>®</sup>-X dye-labeled wheat germ agglutinin) in sterile water (Life Technologies). Dark field fluorescence microscopy was performed using a Nikon Eclipse TE2000-S epifluorescence microscope at 20 $\times$  or 2000 $\times$  magnification using a high-pressure Hg light source.

## Genome Sequencing and Assembly

A permanent draft genome of *Thermus aquaticus* Y51MC23 containing 22 contigs was deposited at NCBI in 2008 (<http://www.ncbi.nlm.nih.gov/nuccore/218297404>) by the Joint Genome Institute (JGI) (Walnut Creek, CA). The draft genome was sequenced using a combination of Illumina fragment libraries and Sanger chemistry. The same culture isolate used for that effort was re-sequenced using the methods described below. Both fragment and 8 kb mate pair libraries were constructed for sequencing using NxSeq<sup>®</sup> DNA Sample Prep and Long Mate Pair Library Kits (Lucigen, Middleton, WI), respectively. Fragment libraries were constructed by shearing genomic DNA to approximately 300–400 with a Covaris LE220 Focused-Ultrasonicator (Covaris, Woburn, MA), end repairing, A-tailing and ligating Illumina compatible adaptors. The adapted DNA was then size selected with AMPure XP beads (Beckman Coulter, Brea CA). An eight kb mate pair library was constructed by shearing genomic DNA with a g-TUBE (Covaris), followed by end repair, A-tailing and ligation of adaptors. Adapted DNA was then size-selected with AMPure XP beads, ligated to a coupler, exonuclease treated, digested with restriction endonucleases and purified prior to circularization with a junction adaptor and PCR amplification. Libraries were sequenced on a MiSeq using Reagent Kit v2 (Illumina, San Diego, CA). Genomic sequence assembly and analysis were carried out using CLC Genomics Workbench v7 (Qiagen, Boston, MA).

DNA extracts were loaded onto a 1% Pulsed Field Certified (BioRad) agarose gel along with LR PFG markers (NEB) and run on a CHEF MAPPER XA system (BioRad) under the following conditions calculated by the MAPPER AutoAlgorithm function for 1–200 kb separation: 0.06–17.33 second linear ramping over 7.53 h run time at 6 V/cm in 1X TAE at 12°C with a 120° reorientation angle. The gel was post-stained with EtBr, destained, and visualized on a GelDoc XR+ system (BioRad).

## PCR Validation of JGI Draft and Lucigen Finished Assemblies of Y51MC23

Both the JGI draft and the Lucigen finished genome assemblies were verified via PCR. Primer pairs were designed from high coverage areas of the sequence assemblies to amplify across regions of ambiguity or low coverage (S1 Table); in nearly all cases the expected amplicon size was observed, except for several regions of possible collapsed repeats. The same general method was applied to verify the order and orientation of the JGI contigs and to validate and strengthen assembly in regions of low sequence coverage for the chromosome and the plasmids in the Lucigen contigs (S1 Fig). All PCR validations were optimized using Taq98 DNA polymerase (Lucigen) or KAPA 2G Robust polymerase (Kapa Biosystems). The amplicons were purified by

SPRI beads (Beckman), and used as template for Sanger sequencing for comparison to and inclusion in the existing assemblies.

## Genome Annotation

The assembled Y51MC23 genome was annotated using Rapid Annotations using Subsystems Technology (RAST) [26, 27]. Manual annotation and curation of the genome was also performed. Corresponding genes in the IMG database [28] were identified by BLASTp [29] analysis using protein sequences generated in RAST. Signal peptides were determined using SignalP 4.0 [30]. Peptidases were predicted with MEROPS [31, 32] and confirmed by BLASTp [29] analysis of the UniProt database [33, 34]. Metabolic reconstructions were performed using PRIAM [35] and SEED [27] metabolic pathway construction software. Genomic islands were identified using Island Viewer [36, 37] software and the closed Y51MC23 chromosome. For assembly comparison purposes, the permanent draft genome for *T. aquaticus* containing 22 contigs (NZ\_ABVK02000001 through NZ\_ABVK020000022) was assembled into a single contig and annotated using RAST. Likewise, the chromosome of *Thermus scotoductus* SA-01, ATCC 700910 (NC\_014974.1) was re-annotated using RAST. Prophage sequences were identified using the program Prophage Finder [38].

The evolutionary history of Y51MC23 was inferred by using the maximum likelihood method based on the Tamura-Nei model [39] on 16S rDNA sequences. Evolutionary analyses were conducted in MEGA5 [40]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 18 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1184 positions in the final dataset. Accession numbers for 16S rRNA gene sequences are: *T. antranikianii* ATCC12462<sup>T</sup>, Y18411; *T. aquaticus* YT-1<sup>T</sup>, L09663; *T. arciformis* strain TH92<sup>T</sup>, EU247889; *T. brockianus*<sup>T</sup>, Z15062; *T. caliditerrae* YIM 77925<sup>T</sup>, KC852874; *T. igniterrae* ATCC 700962<sup>T</sup>, Y18406; *T. oshimai* strain SPS17<sup>T</sup>, Y18416; *M. ruber* DSM 1279<sup>T</sup>, Z15059; *T. scotoductus* SA-01, EU330195.1; *T. scotoductus* ATCC 51532<sup>T</sup>, AF032127; *T. scotoductus* K12, NZ\_JQLJ01000001.1; *T. tengchongensis* YIM 77924<sup>T</sup>, JX112365; *T. thermophilus* 33923<sup>T</sup>, X07998; *T. thermophilus* HB27, NR\_074423.1; *T. thermophilus* JL-18, NC\_017587.1; and *Thermus* sp. CCB\_US3\_UF1, CP003126.1.

## Results

*Thermus aquaticus* Y51MC23 is one of a number of novel thermophilic species isolated from 88°C water in the northern outflow channel of Bath hot spring in the Lower Geyser Basin of Yellowstone National Park [41]. The general features of the organism are summarized in Table 1. The temperature of Bath hot spring is 93.5°C at the source, which is the boiling point at the spring's elevation. The pH of the spring is 8.9 with SiO<sub>2</sub> (244.8 mg/l) and Cl<sup>-</sup> (297.1 mg/l) the dominant dissolved minerals [42]. Y51MC23 is a gram-negative, rod-shaped organism that grows well between 50°C and 80°C with maximum growth rate at 65°C to 70°C (optimum pH for growth is 7.6). The organism forms yellow colonies on YTP-2 agar, and cells isolated from liquid cultures are yellow. Colonies are catalase positive, forming bubbles when tested with aqueous hydrogen peroxide. Like most *Thermus* species, Y51MC23 grows well on media containing low concentrations of salts (2 to 4 g/l), yeast extract and protein hydrolysates such as tryptone or peptone and it does not grow in standard laboratory media such as Luria broth. Unlike the type strain *T. aquaticus* YT-1 [1], Y51MC23 is unable to grow on a minimal

**Table 1. Features of *Thermus aquaticus* Y51MC23.**

Property	Term
Strain	<i>Thermus aquaticus</i> Y51MC23
Gram stain	Negative
Cell shape	Rod
Motility	Non-motile
Temperature range	50°C to 80°C
Optimum temperature	70°C
Carbon source	Amino acids and proteins
Energy source	Chemoorganotroph
Terminal electron acceptor	O <sub>2</sub>
Habitat	Hot spring
Oxygen	Facultative anaerobe
Biotic relationship	Free living
Pathogenicity	Non-pathogenic
Geographic location	Wyoming, USA
Latitude	44.5376262
Longitude	-110.690383
Altitude	2195 m

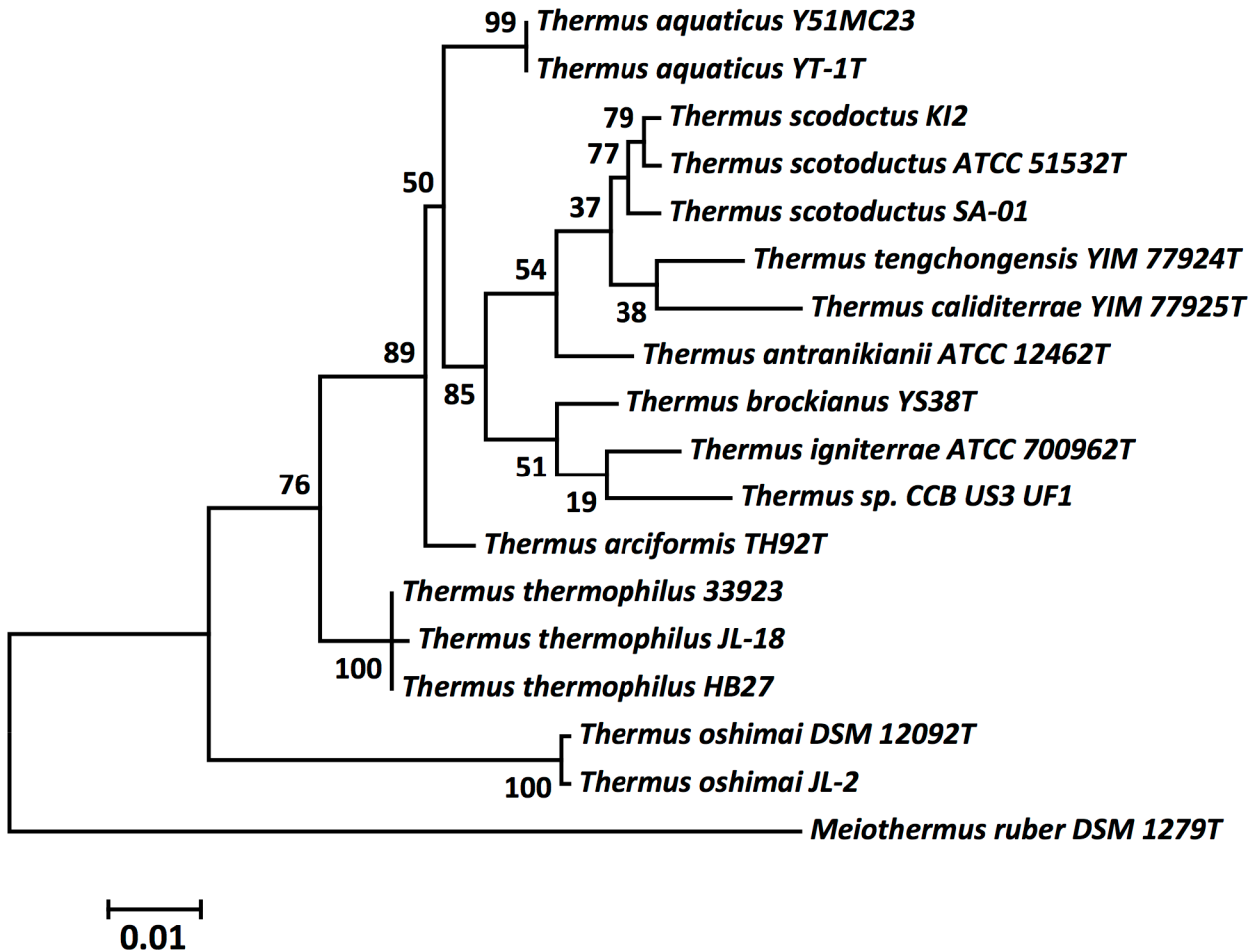
doi:10.1371/journal.pone.0138674.t001

medium containing salts and glucose. Y51MC23 is able to grow under both aerobic and anaerobic conditions, in sharp contrast to YT-1, which was reported to be an obligate aerobe [1]. Cellular morphology in liquid culture is complicated and will be discussed in a later section.

A phylogenetic tree was constructed to determine the relationship between Y51MC23, *Thermus aquaticus* YT-1, and other members (including newly-discovered strains) of the *Thermus* family. The phylogeny of Y51MC23 was determined using its 16S rRNA gene sequence, as well as those of the type strains of all validly described *Thermus* sp. The 16S rRNA gene sequences were aligned using MUSCLE [43], pairwise distances were estimated using the Maximum Composite Likelihood (MCL) approach, and initial trees for heuristic search were obtained automatically by applying the Neighbour-Joining method in MEGA 5 [40]. The alignment and heuristic trees were then used to infer the phylogeny using the Maximum Likelihood method based on Tamura-Nei [39]. The phylogenetic tree identifies Y51MC23 as a *T. aquaticus* strain (Fig 1), in a separate clade from *T. scotoductus*, *T. thermophilus* and *T. oshimai*. The 16S rRNA gene results are also consistent with the analysis of 1,441 orthologous proteins from the currently sequenced *Thermus* genomes [44].

### Closing and Finishing the *Thermus aquaticus* Y51MC23 genome

*Thermus aquaticus* Y51MC23 was re-sequenced by Illumina chemistry using a 300–400 bp ‘fragment’ library and an 8 kb large insert mate pair library, as described in the methods section. Using the fragment library sequence data only, the genome assembled into 82 contigs >1 kb (82 gaps with 164 termini, data not shown). Bioinformatic analysis tabulates 44 repetitive elements > 500 bp in length (discussed in detail later), but only two of them were located internally to the contig termini. These repetitive sequences confound *de novo* assembly programs, causing 84 of 164 contig terminations (51%) in the fragment assembly of *T. aquaticus* (data not shown). The failure to extend and join contigs during *de novo* assembly from fragment sequencing reads is well known [45]. Information that extends beyond the repeat length is required to join contigs in the correct orientation and order. An 8 kb ‘jumping library’, or



**Fig 1. Molecular phylogenetic analysis of *Thermus* species by maximum likelihood method using 16S rRNA gene sequences.** The tree with the highest log likelihood (-3496.7463) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches.

doi:10.1371/journal.pone.0138674.g001

mate pair library was constructed to span these repeats as described in the methods section. A new genome assembly was computed using the fragment plus 8 kb mate pair library, scaffolding the 82 contigs into a single chromosome assembly and four plasmids. To ensure the accuracy of the closed assembly and to finish several gaps caused by problematic GC rich stretches, eighteen pairs of PCR primers were designed and used to cross areas of uncertainty (S1 Table). The gaps were successfully amplified (S1 Fig) and the amplicons were sequenced by Sanger chemistry to finish the complete genome presented here.

Additional manual curation of the draft genome (<http://www.ncbi.nlm.nih.gov/nuccore/218297404>) was performed to systematically compare differences within the finished chromosome and its four attendant plasmids described in this work and the draft assembly (S2 Table, where draft genome genes are identified by their UniProt designations of TaqDRAFT\_3002 through TaqDRAFT\_5541). The manual curation revealed 35 additional annotated genes not present in the draft genome. While most of the newly identified genes are single genes, an insert of six genes (genes 883 through 888) was found in the chromosome. The comparison of the two assemblies also identified 42 genes identified in the IMG (JGI) annotation, but missed by the RAST annotation. The identities of these 42 genes were confirmed and the genes were added to the annotation. The original sequencing effort from 2008 introduced mistakes that

**Table 2. *Thermus aquaticus* Y51MC23 genome summary.**

	Chromosome	pTA14	pTA16	pTA69	pTA78
GenBank Accession	CP010822	CP010823	CP0108224	CP010825	CP010826
DNA, total bases	2158963	14448	16597	69906	78727
Contigs	1	1	1	1	1
G+C	68.0%	64.0%	67.2%	69.1%	66.3%
Protein coding genes	2309	21	20	92	78
tRNA genes	50				
5S rRNA genes	2				
16S rRNA genes	2				
23S rRNA genes	2				

doi:10.1371/journal.pone.0138674.t002

caused 27 frameshift errors, resulting in single genes being split into two separate protein fragments. Correction of these frameshift errors removed 27 genes from the assembly. Elimination of partial and mis-assembled genes at the ends of the twenty-two contigs in the draft assembly removed 17 additional genes, and elimination of questionable small open reading frames removed another 20 genes. Finally, the DNA sequences of 12 genes in the draft genome were not present in the final assembly, reducing the number of actual genes in the draft assembly to 2309 in the final assembly. The complete alignment of the final assembly versus the draft genome is shown in [S2 Table](#).

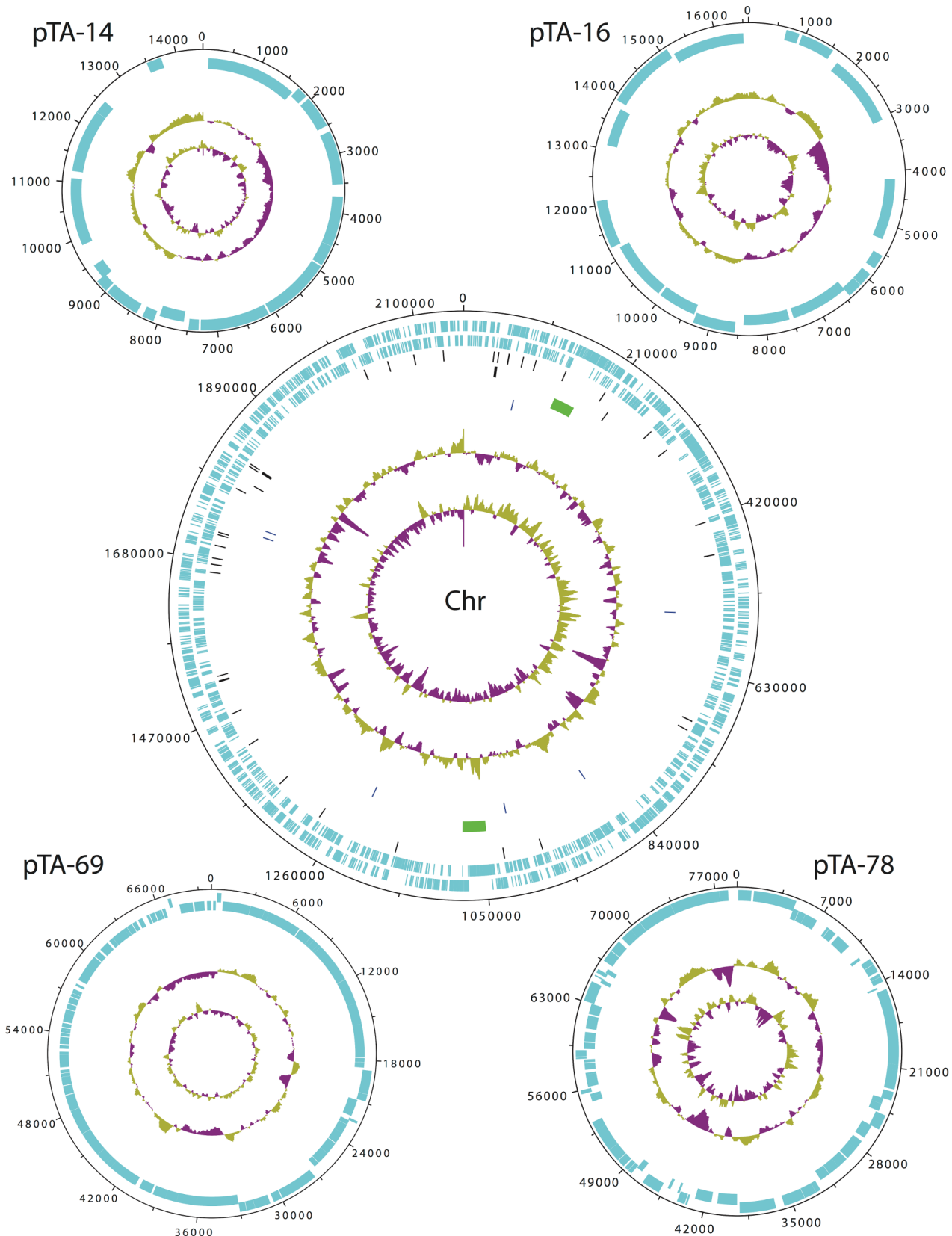
All contigs from the draft assembly were incorporated intact into either the chromosome or one of the four plasmids. A synteny plot was constructed to examine the location of the genes from the draft assembly within the closed chromosome (data not shown). The presence of long stretches of homology in the synteny plot shows that the genes within the contigs of the draft genome were assembled in the correct order, but the contigs themselves could not be assembled. The partial and mis-assembled genes at the ends of the twenty-two contigs prevented recognition of the correct order of these contigs and their assembly in the draft genome from 2008. The use of a mate pair library greater than 5 kb was required to accurately close the genome (data not shown).

The annotation results of the final, closed genome assembly described here are shown in [Table 2](#) and the chromosome plus four plasmids map [46] is shown in [Fig 2](#). The closed genome contains an additional 448 DNA bases not found in the draft sequence analysis. The number of contigs has been reduced from twenty-two to five, consisting of one chromosome and four plasmids ([Table 2](#)). The presence of the four plasmids was confirmed by PFGE ([Fig 3](#)). The 5S and 23S rRNA genes occur in two clusters that are 382 kb apart, and the two 16S genes are separated from these clusters by an additional 172kb and 19kb. This rRNA content differs from the draft assembly, in which three partial 23S genes were called (all at the ends of contigs), plus two complete 5S genes and one complete 16S gene

### Comparison of Y51MC23 to Other *Thermus* Species

The RAST annotations of Y51MC23 (**Taq**), *T. scotoductus* SA-01 (**Tsc**), *T. thermophilus* HB8 (**Tth HB8**) and *T. thermophilus* HB27 (**Tth HB27**) were used to prevent differences in annotation software from influencing genome comparisons. The summary of the annotations ([Table 3](#)) shows that the four organisms have similar genome sizes and similar numbers of genes within the individual subsystems. (The results from the RAST annotations may not agree with published annotation results obtained using other annotation software.)

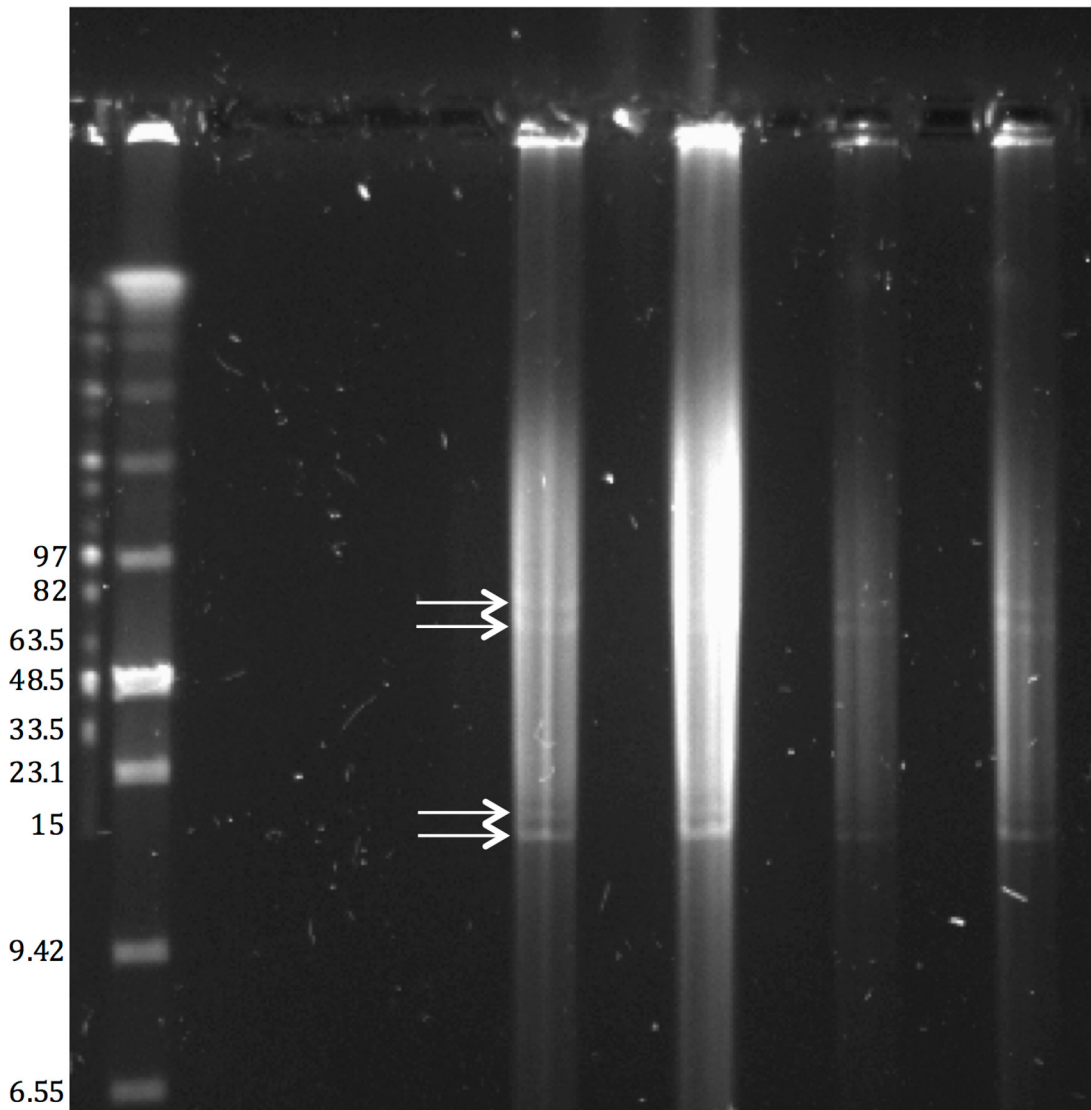
An in-depth comparison was carried out between the genomes of Y51MC23 and *T. scotoductus* SA-01, the closest sequenced relative of Y51MC23. Ninety predicted proteins with



**Fig 2. Features of the *Thermus aquaticus* Y51MC23 chromosome and plasmids.** Tracks from outside to inside: CDs forward strand, CDs reverse strand, tRNA genes, rRNA genes, prophage, CRISPRs, GC plot, and GC skew. Prepared using DNA Plotter software [46].

doi:10.1371/journal.pone.0138674.g002





**Fig 3. Identification of four *Thermus aquaticus* Y51MC23 plasmids by PFGE.** Genomic DNA extracts showed a distinct pulsed field gel electrophoresis banding pattern whose sizes closely match 4 contigs with specific coverage depths that do not map to the primary chromosomal contig. The sizes of these bands (arrows) match actual contig lengths from the finished genome assembly (14.4, 16.6, 69.9, and 78.7 kb respectively), suggesting that the material in the bands is of linear form. The native form of these putative plasmids is presumed to be circular.

doi:10.1371/journal.pone.0138674.g003

>95% identity in each species were identified. As expected, most of these ninety predicted proteins are highly conserved, including 31 predicted ribosomal and translation proteins as well as proteins predicted to be involved in electron transport and ATP generation. Other highly conserved predicted proteins are heat and cold shock proteins, a predicted amino acid transporter cluster, a predicted catalase and an unusual SpoVS-related protein. Y51MC23 possesses 369 annotated proteins not present in *T. scotoductus* SA-01. These unique (<10% identity) proteins of Y51MC23 have a variety of predicted functions. Among the 369 genes, there are two large inserts of predicted prophage genes and multiple predicted phage defense CRISPR protein clusters. A pathway for production of phytoene-based pigment is predicted in Y51MC23, but not SA-01, in agreement with the published report on SA-01 [10]. A pathway for production of corrinoid-type molecules is also predicted in Y51MC23, but not SA-01. Finally, over 280

**Table 3. Comparison of RAST annotations of *Thermus* chromosomes.**

Subsystem Feature Count	Taq	Tsc	Tth HB8	Tth HB27
Total coding sequences	2309	2422	2240	2233
Cofactors, vitamins, prosthetic groups, pigments	226	188	243	243
Cell wall and capsule	56	52	50	60
Virulence, disease and defense	35	35	43	39
Potassium metabolism	7	6	17	17
Miscellaneous	29	32	32	32
Phages, prophages, transposable elements, plasmids	6	1	0	0
Membrane transport	75	94	74	70
Iron acquisition and metabolism	4	3	11	12
RNA metabolism	109	116	123	146
Nucleosides and nucleotides	77	83	86	82
Protein metabolism	227	225	213	212
Cell division and cell cycle	35	32	33	35
Motility and chemotaxis	2	2	1	2
Regulation and cell signaling	15	14	24	20
Secondary metabolism	5	5	5	5
DNA metabolism	82	77	93	63
Fatty acids, lipids, and isoprenoids	91	89	98	82
Nitrogen metabolism	5	18	4	0
Dormancy and sporulation	2	2	2	2
Respiration	79	102	89	82
Stress response	38	41	38	36
Metabolism of aromatic compounds	7	13	17	3
Amino acids and derivatives	281	288	273	258
Sulfur metabolism	14	18	19	16
Phosphorus metabolism	24	31	26	29
Carbohydrates	210	231	245	202

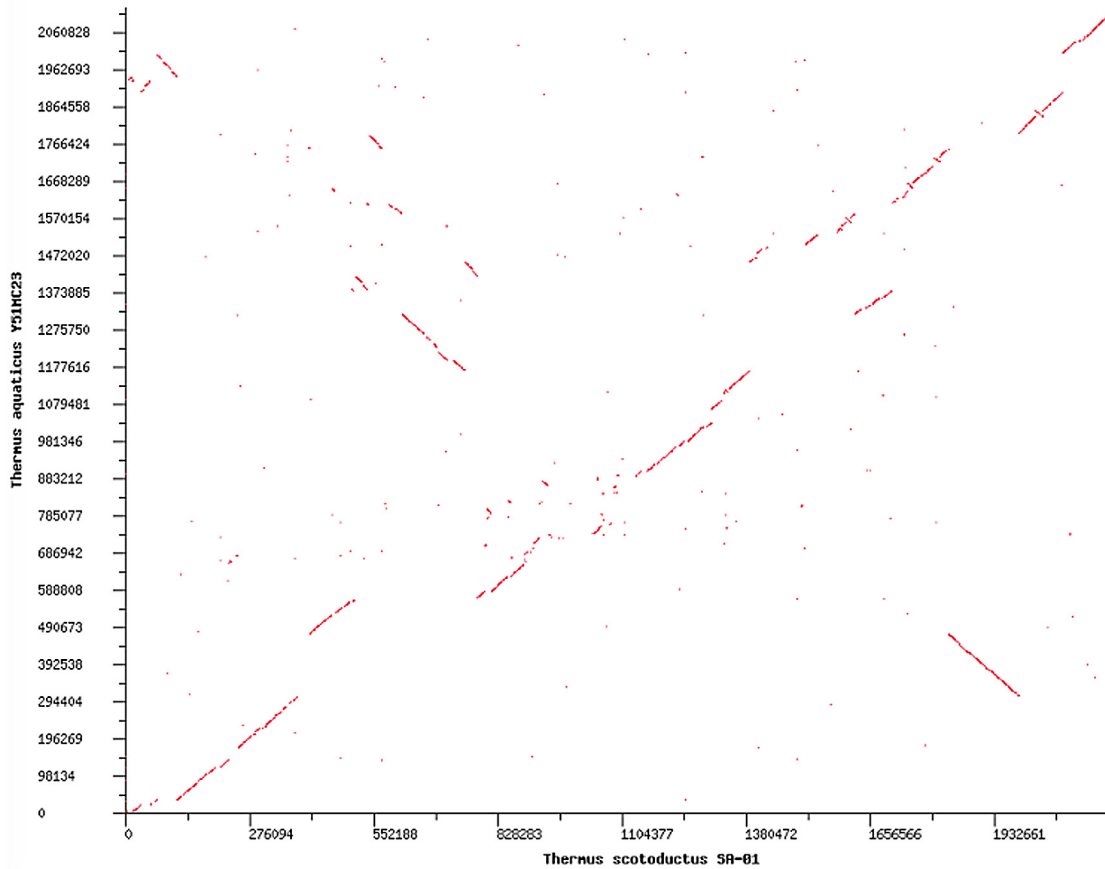
Abbreviations: Taq, *T. aquaticus* Y51MC23; Tsc, *T. scotoductus* SA-01; Tth HB8, *T. thermophilus* HB8; Tth HB27, *T. thermophilus* HB27.

doi:10.1371/journal.pone.0138674.t003

hypothetical or putative proteins with unknown functions are predicted t in Y51MC23, but not SA-01. The proteins unique to *T. scotoductus* SA-01 are associated with genomic islands that are predicted to confer nitrate reduction, aromatic degradation, and metal utilization abilities [10].

The genome of *T. scotoductus* SA-01 has been labeled hyperplastic [10] because of large differences seen when compared to the genomes of *T. thermophilus* HB8 and HB27. Sequence-based comparison of Y51MC23 to *T. scotoductus* SA-01, the phylogenetically-closest sequenced species, revealed a different picture. The synteny plot comparison of these two species shows a high degree of synteny throughout the genomes (Fig 4) suggesting that regions of relative stability exists between certain *Thermus* species. Sequence-based comparison of Y51MC23 to *T. thermophilus* strains shows fewer and smaller regions of synteny, as expected from the greater distance on the phylogenetic tree (Fig 1).

Genomic islands have been implicated in rearrangements [44] and acquisition of new metabolic capabilities [10] in *Thermus* species. Analysis of the Y51MC23 genome using sequence composition-based approaches, such as SIGI-HMM and IslandPath-DIMOB, and the comparative genomics approach IslandPick [36] showed only two small potential genomic islands



**Fig 4. Synteny plot of the closed and finished *T. aquaticus* Y51MC23 genome versus *T. scotoductus* SA-01.**

doi:10.1371/journal.pone.0138674.g004

(data not shown). The first potential island (7473 bp) contains eight genes (1072 through 1079), and the second potential island (7678 bp), six genes (2195 through 2200). Neither island codes for an identifiable metabolic pathway. BLASTn analyses show both regions are highly conserved in other *Thermus* species such as *T. scotoductus*, *T. thermophilus*, *T. oshimai*, and *Thermus* sp. CCB\_US3\_UF1, with 50% to 80% coverage and 75% to 85% identity. No significant (>2% coverage) hits to non-*Thermus* sequences were found, supporting the hypothesis that these genes were not acquired from organisms outside the genus.

### Insights from the *Thermus aquaticus* Y51MC23 genome

**Transport functions.** Y51MC23 possesses genes encoding 35 predicted membrane transporter systems, significantly more than the 22 predicted transporter systems reported in *T. scotoductus* SA-01 [10]. Like *T. scotoductus*, Y51MC23 appears to rely primarily on thirty-one ABC (ATP-binding cassette protein) transporter systems (Table 4) that couple ATP hydrolysis to transport of nutrients from the medium. ABC transporters typically contain four components, two integral membrane proteins and two cytoplasmic proteins [47], though ABC transporters that utilize fewer and more varied components (ECF transporters) have also been described [48]. Of the thirty-one predicted ABC transporter systems, the highest number, thirteen, are predicted to transport peptides and amino acids, indicating Y51MC23 may utilize amino acids as its primary energy source. Y51MC23 possesses genes that could potentially encode three ABC transporter systems for sugars, two are predicted to transport maltose and

**Table 4. Predicted Transporter systems present in the *T. aquaticus* Y51MC23 genome.**

Genes	Transporter Type	Annotated Substrate
36–38	ABC	Maltose-maltodextrin
104–106	ABC	Unknown
240–242	ABC	Molybdenum
311–312	ABC	Thiamine
504–505	ABC	Cytochrome C biogenesis
571–573	ABC	Oligopeptide
574–578	ABC	Branch chain amino acids
589–591	TRAP	C4-dicarboxylic acids
634–637	ABC	Branch chain amino acids /leucine
704	ATPase	Cations
706	Antiporter	Calcium-sodium
804–807	ABC	Inositol
816–818	ABC	Dipeptides
850–852	ABC	Heme
877–878	ABC	Branch chain amino acids
880–882	ABC	Branch chain amino acids
902–906	ABC	Branch chain amino acids
913–915	ABC	Glycerol phosphate
925–928, 930	ABC	Branch chain amino acids
965–968	ABC	Branch chain amino acids
1024–1026	ABC	Taurine
1120	ABC	Glycerol phosphate
1231–1234	ABC	Iron citrate
1258–1262	ABC	Amino acids-lysine, arginine, ornithine
1297–1299	ABC	Unknown
1397–1400	ABC	Spermidine, putrescine
1503–1505	ABC	Copper
1808–1812	ABC	Oligopeptides
1843	ABC	Amino acids-methionine
2031–2033	ABC	Maltose-maltodextrin
2042–2043	ABC	Branch chain amino acids
2079	ABC	Ferric iron
2108–2110	TRAP	C4-dicarboxylic acids
2178–2179	ABC	Glycerol phosphate
2213	ABC	Iron

doi:10.1371/journal.pone.0138674.t004

maltodextrins, and one to transport inositol. Y51MC23 possesses genes that could potentially encode for five ABC transporter systems for essential metal ions, one each for molybdenum, heavy metals, and copper and two for iron. Other potential ABC transporter systems are annotated as transporters for amines such as spermidine, putrescine, and taurine; thiamine; heme; and glycerol phosphate. Most of the potential ABC transporter system clusters contain either three or four genes: two for membrane components and one or two for permease components. The annotated thiamine, cytochrome C biogenesis, and glycerol-phosphate transport genes may be part of ECF transporters. In addition to the potential ABC transporters, Y51MC23 possess two predicted tripartite ATP-independent periplasmic (TRAP) transporters that utilize an ion gradient to transport four carbon dicarboxylic acids, and a predicted PacL-type ATPase

cation transporter that may transport potassium, calcium or heavy metals such as lead, mercury, cadmium, or zinc. An annotated calcium-sodium antiporter is also present. There are no genes that could potentially encode three-component phosphotransferase system (PTS) transporter systems that use phosphoenolpyruvate to transport sugars into the cell and phosphorylate them.

**Sugar Metabolism and Biosynthesis.** Based on the SEED metabolic reconstruction [27], Y51MC23 is predicted to interconvert and utilize glucose, fructose, galactose and mannose. However, based on the annotation, it is unclear if transport systems for galactose, mannose and fructose exist, and the lack of these transport systems may limit the organism's ability to utilize galactose, mannose and fructose. Of the hexose oligosaccharides and polysaccharides, Y51MC23 is predicted to be able to only utilize maltose, maltodextrins, and oligosaccharides derived from starch. The organism possesses three genes predicted to encode glycoside hydrolase family 13 proteins (GH13) and one GH57 amylopullulanase [49]. The GH 57 amylopullulanase gene (2070) codes for a predicted signal peptide[30], suggesting the protein is secreted into the medium where it may generate soluble malto-oligosaccharides from starch. These malto-oligosaccharides may then be transported by two predicted ABC transporters coded by genes 36, 37 and 38, and 2031, 2032 and 2033. The three predicted GH13 genes (388, 408, and 2034) do not code for signal peptides and these three amylases appear to act only on intracellular substrates, either malto-oligosaccharides or storage polysaccharides, converting them to glucose for utilization. There are no annotated genes coding for secreted galactosidases, invertases, cellulases, or mannanases, indicating these substrates cannot be utilized by Y51MC23. The SEED metabolic reconstruction also indicates that Y51MC23 lacks the genes needed to utilize commonly occurring pentose sugars such as xylose, xylulose, or arabinose. A lack of genes coding for the necessary pathway enzymes also prevents utilization of inositol, inositol phosphate, sugar alcohols, glucuronate and other sugar acids. The genes predicted to code for a complete pentose phosphate pathway are present, predicting the ability to produce ribose from fructose.

**Energy Generation.** Y51MC23 is predicted to possess a complete Embden-Meyerhof-Parnas (glycolysis) pathway for the utilization of glucose, fructose and related sugars. Based on the SEED metabolic reconstruction, pyruvate formed in glycolysis may be converted to acetate, lactate and ethanol under anaerobic conditions. While Y51MC23 does not possess genes coding for an annotated NADH-utilizing lactate dehydrogenases for conversion of pyruvate to lactate, pyruvate may be converted to lactate via a predicted cytochrome *c*-dependent D-lactate dehydrogenase (gene 317) and a predicted three-component L-lactate dehydrogenase coupled to an Fe-S oxidoreductase (genes 548, 549, 550). Ethanol may be produced from pyruvate by a predicted alcohol dehydrogenase (gene 1813).

Under aerobic conditions the pyruvate produced in glycolysis is predicted to be utilized via a complete citric acid (TCA) cycle, with ATP generated through oxidative phosphorylation. Y51MC23 possesses a predicted NADH ubiquinone oxidoreductase complex coded by genes 1536 through 1549 (chains A through N) and gene 1789 (NADH-quinone oxidoreductase chain 15). Genes 1497 through 1500 are predicted to code for components of a succinate dehydrogenase. Y51MC23 appears to have genes coding for two separate cytochrome C oxidase complexes (genes 713–714 and 2163–2164) as well as an ubiquinol cytochrome C oxidoreductase complex (genes 2181–2184). Unlike other *Thermus* species [19], Y51MC23 appears unable to utilize nitrate as a terminal electron acceptor, because the strain contains no annotated gene cluster for the reduction of nitrate to nitrous oxide. This was confirmed by fermentation studies, where addition of nitrate did not stimulate growth of Y51MC23 under anaerobic conditions.

ATP appears to be generated using a V-type ATP synthase V complex. V-type ATPases have been reported in *T. aquaticus* YT-1 as well as *T. thermophilus* HB27, while other *Thermus* species possess the more typical bacterial F-type [50]. The genes coding for the predicted ATP synthase are located downstream from the NADH ubiquinone oxidoreductase genes. The ATPase genes, 1552 through 1560, code for (in order) ATPase subunits D, B, F, A, C, E, K, I, G.

**Amino Acid Metabolism and Biosynthesis.** Y51MC23 may utilize a combination of secreted, membrane bound, and cytosolic proteases and peptidases for converting external proteins and peptides into free amino acids. In contrast to the predicted single secreted enzyme for oligosaccharide degradation (the GH 57 amylopullulanase), Y51MC23 possesses thirteen genes predicted to code for secreted peptidases/proteases. Among the thirteen secreted proteins, the annotations predict three subtilisin-like proteases, two chymotrypsin-like proteases, two metalloproteases, and a carboxypeptidase. The two predicted chymotrypsin-like proteases (genes 1448 and 1635) have annotated PDZ domains, that may increase the thermostability of these proteases [51]. The peptides and amino acids generated by these enzymes may be transported into the cytoplasm via the thirteen predicted peptide and amino acid ABC transporter systems described above. Once inside, seventeen predicted soluble peptidases and two predicted membrane-bound peptidases may hydrolyze the peptides to free amino acids for metabolism and reuse.

Y51MC23 recycles misfolded proteins into amino acids using two ATP-dependent pathways. The organism possesses genes predicted to code for a two-subunit 20S proteasome system (gene 769, HslV and gene 770 HslU) [52, 53] as well as three annotated Lon proteases (genes 332, 737, and 1458) [54]. There are no genes predicted to code for recycling transglutaminase/proteases annotated in the genome. Y51MC23 also possesses genes predicted to code for an ATP-dependent TldE/TldD proteolytic complex (genes 1447 and 1448) with an unknown function.

Based on the metabolic reconstruction, Y51MC23 appears to be able to metabolize sixteen of the twenty amino acids. Based on genes present, pathways for degradation of Glu, Gln, Asp, Asn, Ala, Gly, Ser, Thr, Met, Cys, Val, Leu, Ile, Arg, Pro, and His are predicted, while no predicted pathways exist for degradation of Lys, Trp, Tyr, Phe. Under aerobic conditions, the degradation products of these amino acids appears to funnel predominantly into the citric acid cycle and generate ATP via oxidative phosphorylation. While Y51MC23 is able to grow on amino acids under anaerobic conditions, it is unclear how ATP generation is coupled to amino acid degradation.

The metabolic reconstruction indicates that Y51MC23 possesses all genes necessary for the biosynthesis of all twenty amino acids via conventional microbial pathways. Among the amino acids with multiple known biosynthetic pathways, lysine is predicted to be synthesized via the N-2 acetyl-L-amino adipate pathway and cysteine is predicted to be synthesized from acetyl serine and H<sub>2</sub>S. Genes within biosynthetic pathways are not organized into a single operon controlled by regulatory units. The thirteen Y51MC23 tryptophan biosynthesis genes are scattered throughout the genome (genes 520, 521, 522, 1058, 1353, 1354, 1447, 1733, 1734, 2080, 2187, 2188, and 2189), unlike the tryptophan operon in *E. coli* [55] or thirteen-gene operon (GY4MC1\_1353 through GY4MC1\_1364) for tryptophan biosynthesis present in *Geobacillus thermoglucosidasius* Y4.1MC1, a Gram-positive organism isolated from the same hot spring as Y51MC23 [41].

**Biosynthesis of Vitamins, Cofactors, and Pigments.** Y51MC23 possesses genes potentially encoding biosynthetic pathways for production of three porphyrin products, heme, siroheme, and corrinoids. In addition to the predicted heme biosynthetic pathway, Y51MC23 also possesses genes potentially encoding an ABC transporter cluster for uptake of heme (genes 850, 851, and 852). The predicted biosynthetic pathway for corrinoid biosynthesis is similar to

that found in *T. thermophilus* HB8 megaplasmid pTT27. In Y51MC23, the predicted pathway is primarily coded by an eleven gene cluster (genes 976 through 986) that is syntenous to pTT27 genes TTHB051 through TTHB060. Remaining orthologs to the pTT27 pathway are scattered through the Y51MC23 chromosome. Y51MC23 encodes genes for the pathways involved in biosynthesis of riboflavin, nicotinate/NAD, and folate. Genes coding for pathways involved in biosynthesis of thiamine, B6, pantothenate, and biotin are absent, leading to the growth requirement for yeast extract.

Like most *Thermus* species, Y51MC23 cells are yellow, the result of carotenoid biosynthesis [56]. Y51MC23 possesses a chromosomal gene cluster for phytoene biosynthesis (genes 1210 through 1222) similar to the phytoene gene cluster in *T. thermophilus* HB8 megaplasmid pTT27 (genes TTHB098 through TTHB110) [57]. The two clusters show the same organization, with the pTT27 cluster containing an additional annotated plasmid stability protein (gene TTHB108) within the cluster. Light regulates carotenoid biosynthesis via a LitR-dependent regulatory protein in the Crp/Fnr family (gene 1211) similar to that observed in *T. thermophilus* [58], and the light is detected via a proteorhodopsin (gene 1599) [59].

**Plasmids.** Y51MC23 possesses four plasmids. The two smaller plasmids, pTA14 and pTA16, contain 21 and 20 genes respectively. The genes predominantly code for hypothetical proteins with no significant BLAST hits to proteins with known functions. Plasmid pTA78 contains 78 annotated genes. Among the genes of interest in this plasmid are clusters predicted to encode a heme transport and utilization system (pTA78-27 through 31), an aerobic-type, class 1a ribonucleotide reductase (pTA78-34 and pTA78-35), and chromosome partitioning proteins ParA and ParB (pTA78-74 and pTA78-75). The plasmid also contains two partial copies of the pilT gene (pTA78-41 and pTA78-53, 137 and 138 a.a. respectively), which may allow integration of the plasmid into the two chromosomal copies of pilT (genes 50 and 210). The entire heme transport gene cluster is also present in the chromosome of *Thermus* sp. CCB\_US3\_UF1 (accession CP003126, nt. 228390–224668), the megaplasmid pTT27 of *Thermus thermophilus* HB8 and HB27 (accession AP008227, nt. 223,442–227,102 and accession AE017222, nt. 176,277–179,767, respectively) and the megaplasmid pTHEOS01 of *Thermus oshimai* (accession CP003250, nt. 156812–160493).

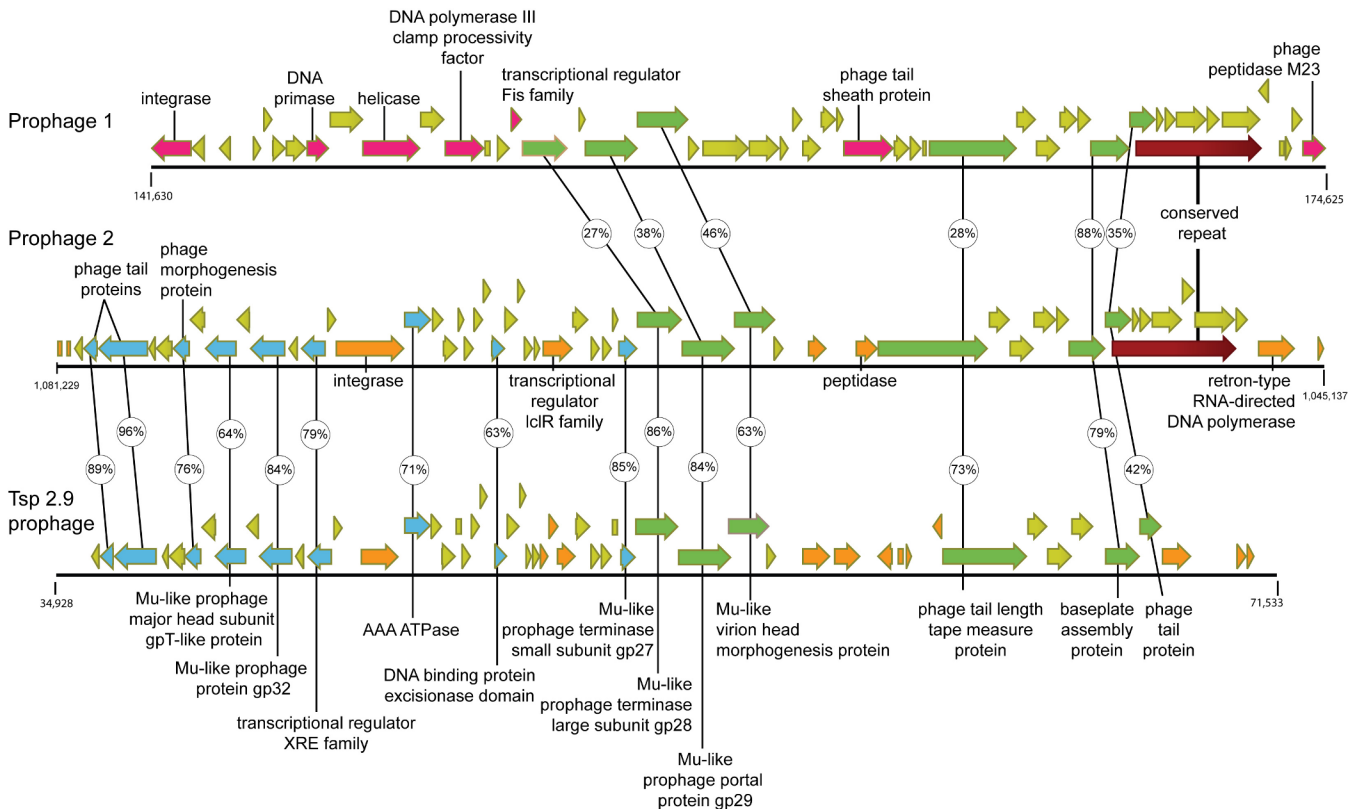
The second largest plasmid, pTA69, contains 92 annotated genes, and appears to code for a *Thermus* conjugation system. The conjugation system consists of relaxosome [60] and transferosome components. The transferosome is made of genes predicted to encode a type IV secretion system (T4SS) [61]. Because of the diverse structures of the protein components of relaxosomes and transferosomes, the function of each plasmid protein cannot be definitively ascertained. Based on structural and sequence homologies, the plasmid encodes predicted orthologs of relaxosome proteins including primase (pTA69-32), relaxase (pTA69-45), ssDNA binding protein (pTA69-70) and ds break repair protein (pTA69-62). The plasmid encodes identifiable orthologs of T4SS components including virB1 (pTA69-13), virB2 (pTA69-22), virB3 (pTA69-27), virB4 (pTA69-20), virB5 (pTA69-22), virB6 (pTA69-18), virB7 (pTA69-26), virB9 (pTA69-30), virB10 (pTA69-31), virB11 (pTA69-34), VirD4 protein (pTA69-16), and conjugal transfer protein TraD (pTA69-19) as well as additional annotated membrane proteins that may be components of the secretion system [62]. Orthologs of these genes are present in the *Thermus thermophilus* JL-18 plasmid pTTJL1802 (accession CP003254) and SGO,5JP17-16 plasmid pTHTHE1601 (accession CP002778). The pTA-69 plasmid also contains genes for two predicted nucleases (pTA69-41 and pTA69-77), a Type I restriction-modification system (pTA69-42), a serine protease (pTA69-46), peptidase (pTA69-3), chromosome partitioning proteins ParA and ParB (pTA69-43 and pTA69-44), and RNA polymerase sigma-70 factor (pTA69-91).

*Thermus thermophilus* strains HB8 and HB27 both possess “megaplasms” that contain over 200 kb of DNA. The *T. thermophilus* HB8 plasmid pTT27 (NC\_006462) possesses 228 RAST-annotated genes. Of these 228 genes, 133 have orthologs on the Y51MC23 chromosome, and 95 have no orthologs. Orthologs of the genes of the HB8 megaplasms include the cobalamin synthesis cluster (TT\_P001 through TT\_P023), which is present on the Y51MC23 chromosome (0832 through 0839 and 0976 through 0985). The HB8 megaplasms hemin transport cluster (TT\_P175 through TT\_P179) also has a Y51MC23 chromosomal ortholog (0849 through 0852), which differs from the heme transport/utilization cluster on plasmid pTA-78. Orthologs to the genes of the phytoene synthetic cluster on the HB8 megaplasms (TT\_P101 through TT\_P111) are also found in the Y51MC23 chromosome (1213 through 1222). Of the 95 megaplasms genes with no orthologs in the Y51MC23 chromosome, 58 code for hypothetical proteins. Other genes with no orthologs in the Y51MC23 chromosome are four annotated *beta*-galactosidases and one *alpha*-glucosidase. The ability to map orthologs of over half of the megaplasms genes and clusters to the Y51MC23 chromosome (often with a high degree of synteny) suggests that Y51MC23 may have stably incorporated large sections of a megaplasms into its chromosome. Further work is needed to establish if this chromosomal incorporation is a rare or common event in *Thermus* species.

**Prophage.** Numerous lytic phages that infect *Thermus* species have been described [63–71], but little is known about temperate bacteriophages and their putative prophage elements. Y51MC23 is unusual among the sequenced *Thermus* species in possessing two large prophages. *Thermus* sp. RL [20], *Thermus* sp. CCB\_US3\_UF1 [21], and *Thermus* sp. 2.9 [22] each have one prophage, while none of the other *Thermus* genomes appear to have any. Y51MC23 prophage 1 is 32,996 bp (positions 141,630–174,625) and prophage 2 is 36,093 bp (positions 1,045,137–1,081,229). Prophage 1 contains 48 annotated genes (23 hypothetical) whereas prophage 2 contains 55 (28 hypothetical) (Fig 5). The architecture of both elements is similar to many known phage where structural and replication associated genes are arranged together in modules. Both prophage contained multiple structural genes for predicted head/capsid proteins (174, 1175), tail (177, 180, 185, 190, 194, 1138, 1180, 1181), baseplate (118, 189, 1139), and virion processing genes for terminase packaging (168, 1150, 1151), tape measure (184, 1144), baseplate assembly (188, 189), lysozyme (188, 1140), and portal/morphogenesis proteins (168, 171, 1148, 1149, 1177). Many of these are annotated as Mu-like; bacteriophage Mu is a dsDNA temperate phage that uses DNA-based transposition in its lysogenic cycle. Predicted replication and repair genes were more distinct between the two prophage elements. Prophage 1 contains annotated genes for a predicted DNA primase (160), helicase (162), DNA polymerase III beta clamp processivity factor (164), and Holliday junction resolvase (166), whereas prophage 2 only contains an annotated gene for a predicted RNA-directed DNA polymerase (1131).

The two Y51MC23 prophages are unique across most of their length, but they do share 99.6% identity (3528/3562 identical) across 3,562 bp (prophage 1 positions 169,281–172,822; prophage 2 positions 1,047,643–1,051,184) (Fig 5). The region of 99.6% homology is 1803 or 2506 bp from the boundaries of the prophage elements. A novel RNA-directed DNA polymerase (1131) near the terminal boundary for prophage 2 could explain the apparent duplication of five genes in two separate prophage separated by 874,821 bp in Y51MC23. The RNA-directed DNA polymerase at locus 1131 is a special type of reverse transcriptase classified as a “diversity generating retroelement” (DGR) [72]. DGRs encode a reverse transcriptase (RT) as well as a template repeat (TR) and a variable repeat (VT) [72], all three of which were identified in Taq Y51MC23. TR1 is located at 1,047,173–1,047,258 (GGAATGGCGGTAACGCGGGGC TCGCCGCGTTGAACCTGCTCAACCCGCGCGGCAACCGGAACCTGGAGTGTCCGG GCCCCCCCCGCT), and VR1 is located at 1,047,693–1,047,778 (GGAATGGCGGTGTC





**Fig 5. Diagram of Y51MC23 prophage 1 and 2 versus *Thermus* sp. 2.9 prophage synteny.** Annotated ORFs are shown as block arrows for *Thermus aquaticus* prophage 1 (top), prophage 2 (center), and *Thermus* sp. 2.9 (bottom). Gold indicates hypothetical ORFs; Green indicates ORFs shared between all three prophage; magenta indicates ORFs specific to prophage 1; blue indicates ORFs shared between prophage 2 and TSP2.9 prophage; orange indicates regions that differ between prophage 2 and TSP2.9 prophage. The 3,652 bp region of nucleic acid identity between prophage 1 and 2 is indicated in brick-red color. Amino acid identity between pairs of encoded proteins are indicated in small circles as determined by blastp.

doi:10.1371/journal.pone.0138674.g005

GCGGGGCTCGCCGCGTTGTACCTGCTCAACCCGCGCGGCTCCCGGCGCTGGGGT GT CGGGGCCCGCCCGCT), just upstream of the reverse transcriptase (1,045,973–1,047,016). VR1 is located in the carboxy-terminal end of the putative formylglycine-generating sulfatase enzyme (FGE sulfatase), which is also the case 27% of the time in the other 155 identified DGR elements [72]. The role that FGE sulfatase might play as a target protein for diversity generation is unknown [72].

There are seven *Thermus*-specific bacteriophages whose genome has been determined (phiYS40, TMA, phiOH2, P23-77, P23-45, p74-26, In93), none of which have homology to Y51MC23 prophage 1 or 2 genes. However, there is a significant degree of synteny and homology between Y51MC23 prophage 2 (55 annotated genes) and a prophage embedded in the *Thermus* species 2.9 [22] genome (55 annotated genes) (Fig 5). *Thermus aquaticus* Y51MC23 was isolated from Bath Hot Spring in Yellowstone National Park, whereas *Thermus* species 2.9 (Tsp 2.9) was isolated from a hot spring in Salta, Argentina [22]. Thirty eight genes share homology between the two prophages, ranging from 42–98% identity (Fig 5). The synteny between the two prophages is colinear except for three regions between Taq Y51MC23 1137–1138, 1144–1150, and 1155–1157 (Fig 5). The single gap at Taq Y51MC23 1168–1169 is most likely a sequence error in the Tsp 2.9 genome at QT17\_04420, as this gene appears to have a frame shift mistake that would make it part of 4415. Both prophages have identical flanking bacterial genes on both sides of their respective insertion sites: cell division protein

**Table 5. CRISPR elements found in *T. aquaticus* Y51MC23.**

Name	Genomic location	# DRs	Left flank	Right flank
CRISPR 1	81689..82168	7	<-cas2	<-cas1
CRISPR 4	550600..550809	4	<-hypothetical	<-disulfide bond formation protein B gene
CRISPR 5	869567..871070	23	cas1->	Transcriptional regulator <i>tetR</i> ->
CRISPR 7	1010468..1010734	4	Mobile element gene->	<-hypothetical
CRISPR 8	1232917..1233864	13	<-putative esterase gene	CRISPR-associated <i>csm1</i> ->
CRISPR 9	1731966..1732521	8	<-3'-5' oligoribonuclease A gene	CRISPR-associated <i>TM1812</i> ->
CRISPR 10	1742839..1743245	6	hypothetical->	<-hypothetical

doi:10.1371/journal.pone.0138674.t005

gene *mraZ* (gene locus 4350 for Tsp 2.9, 1185 for Taq Y51MC23) on one side, and 50S ribosomal protein L31 (gene locus 4640 for Tsp 2.9, 1129 for Taq Y51MC23) on the other side.

Two more putative prophage remnants can be found from gene locus 773–796 and 2148–2161, which includes a phage associated primase (784), phage recombination protein Bet (785), and a hypothetical protein homolog to DNA repair proteins (783) for the first one, and a conserved hypothetical protein with homology to primase (2151), a bifunctional DNA primase/polymerase family protein (2152) and a nuclease domain protein (2158) for the second one.

**CRISPR Elements.** CRISPR-Cas modules (Clustered regularly interspersed short palindromic repeats-CRISPR associated proteins) comprise the adaptive immune system in many bacteria and archaea. In contrast to the typical single CRISPR locus in most bacterial genomes, Y51MC23 contains 7 definite CRISPR loci (Table 5), similar to the hyperthermophile *Sulfolobus solfataricus*, which contains 6 CRISPR loci [73]. Similar numbers of CRISPR loci were found in organisms isolated from Yellowstone hot springs and sequenced by our group, including *Geobacillus thermoglucosidasius* Y41MC1 (Bath hot spring, 6 CRISPRs) [41], *Geobacillus thermocatenulatus* strains Y41MC52 and Y41MC61 (Obsidian Pool, 6 CRISPRs each), and *Geobacillus thermoglucosidasius* C56-YS93 (Obsidian Pool, 6 CRISPRs). Features of the definite CRISPR loci and flanking ORFs (arrows indicate reading frame orientation) are listed below. CRISPR 1 is flanked by *cas2*, *cas1*, and CRISPR-associated *cas02710* (genes 92–94). CRISPR 5 is flanked by *cas2*, CRISPR-associated genes *csd2/csh2*, *csh1*, CRISPR repeat RNA endoribonuclease gene *cas6*, CRISPR-associated RecB-family exonuclease gene, and *cas1* (genes 947–954). CRISPR 8 is flanked by CRISPR-associated genes *csm1* through *csm5*, CRISPR-associated *cas02710*, CRISPR repeat RNA endoribonuclease *cas6*, and Exonuclease *sbcC* and *sbcD* (genes 1341–1353). CRISPR 9 and 10 are located very near each other and are separated by CRISPR-associated gene *TM1812* (gene 1849) plus CRISPR-associated RAMPs *cmr1* through *cmr6* (genes 1850–1855).

**Other Repetitive Elements.** Overall the genome contains 44 repetitive elements > 500 b in length, the largest of which are the prophage repeats (2 copies, 3544 bp) and rRNA operons (2 copies, 3361 bp). The remaining 40 repeats fall into 12 classes with lengths ranging from 757 bp to 1695 bp, and copy numbers ranging from 2 to 8. The average identity for the 44 repeats is 98.16%, with 16 of the 44 repeats inverted relative to the first copy. Altogether the 44 elements account for 41,282 bp, equal to 1.91% of the genome. The highest copy number elements are mobile element protein WP\_003043735.1 (8 copies, 1690 bp) and an ORF with similarity to IS4 family transposase WP\_003043670.1 (7 copies, 923 bp). These 44 repeat elements were the predominant cause of contig extension failure during *de novo* assembly from fragment sequencing reads, accounting for 84 of 164 contig termini (51%) in the fragment assembly.

**Cell Envelope and Cell Shape.** A 95 kd protein, SlpA, is primarily responsible for formation of the S-layer in *T. thermophilus* HB8 and HB27 [74, 75], and constitutes most of the

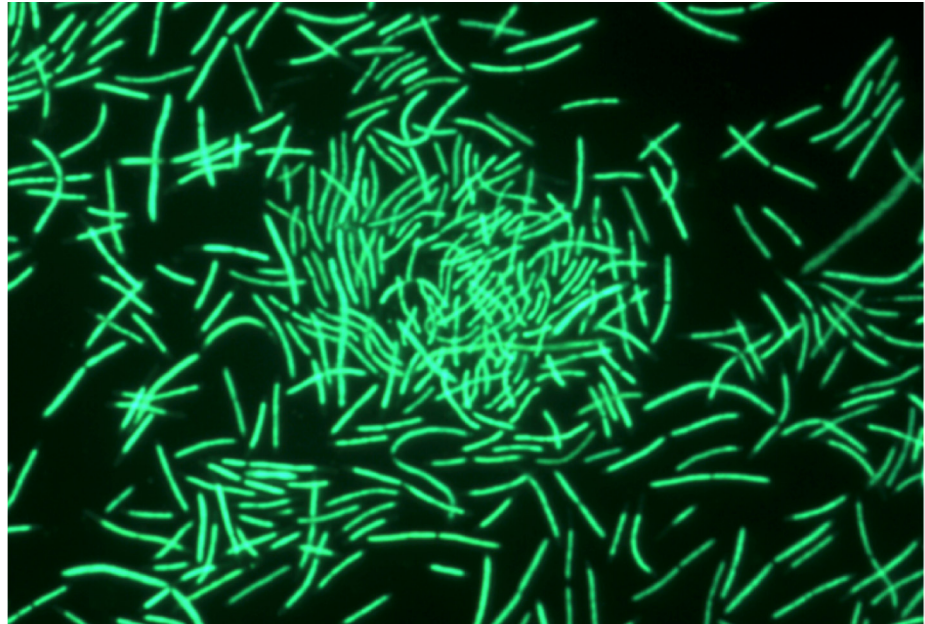
protein recovered from the cell envelope. Y51MC23 has a SlpA ortholog (gene 2228) with 911 amino acids. Immediately downstream of the *slpA* gene is an annotated glucosamine-6-P synthase gene, an ortholog of the *glmSth* gene found downstream of the S-layer protein gene in *T. thermophilus* [76]. The product of the glucosamine-6-P synthase gene is also a component of the cell envelope. Based on SignalP analysis [30], Y51MC23 also possesses a number of genes coding for large, non-catalytic secreted proteins that may be involved in cell envelope structure or formation of multiple cellular forms observed in micrographs. These proteins include the product of gene 1363, coding for a predicted secreted protein of 2665 amino acids, and the products of genes 932, 933, 934 and 935, coding for predicted secreted proteins of 912, 882, 1793, and 1356 amino acids respectively. BLASTp analyses of the sequences of these proteins show they are conserved in members of the *Deinococcus-Thermus* phylogenetic group, but not in other organisms.

The Y51MC23 genome codes for a number of proteins that may be involved in generating the unusual morphologies seen in micrographs described below. Y51MC23 possesses two genes predicted to be *lytR\_cpsA\_psr* family members (genes 206 and 1577). The *lytR\_cpsA\_psr* family members are putative membrane-bound proteins that contain an additional predicted extracellular domain. The LytR\_CpsA family of proteins has been implicated in a variety of functions in other organisms including cell capsule formation in *Streptococcus pneumoniae* [77] and regulation of *Bacillus anthracis* cell length through S-layer assembly and attachment of secondary cell wall polysaccharide to peptidoglycan [78]. The two annotated *lytR\_cpsA\_psr* genes may regulate the formation of single cells versus chains of cells in Y51MC23 and the overall length of the chains. In addition, Y51MC23 possesses annotated *spoVR* and *spoVS* genes (genes 11651 and 1644) and a *prkA* gene (gen 1649). In *B. subtilis*, *spoVR* [79] and *spoVS* [80] genes are involved in the asymmetrical cell differentiation that ultimately leads to spore formation. The *prkA* gene codes for a serine protein kinase [81] that is also involved in spore formation in *B. subtilis* [82], where PrkA accelerated sporulation and the expression of  $\sigma^K$  by suppression of Hpr [83]. Y51MC23 possesses a gene coding for an annotated, secreted, stage II sporulation protein D ortholog (SpoIID, gene 406). In *B. subtilis*, SpoIID is a membrane-anchored enzyme that degrades peptidoglycan and is required for sporulation [84]. Two predicted cell wall endopeptidases (genes 214 and 1610) may be involved in shape determination. Finally, Y51MC23 possesses a gene coding for a secreted protein (gene 391) that contains two sporulation related domains that bind to peptidoglycan. These domains are found in proteins such as FtsN, DedD, and CwlM that are involved in cell division and spore formation. Taken together, the presence of these proteins suggests a more primitive system with similarities to sporulation in Gram-positive organisms is responsible for the formation of the unique, multiple morphologies seen in the micrographs below.

## Unique Cellular Morphologies

Unlike *T. aquaticus* YT-1, Y51MC23 grows well under both aerobic and anaerobic conditions. Under aerobic conditions (50 ml media in 250 ml flask, 70°C, 200 rpm, silicone foam closure) the cells grew as single short rods and rosettes (Fig 6) similar to the growth observed by Brock for YT-1 [1].

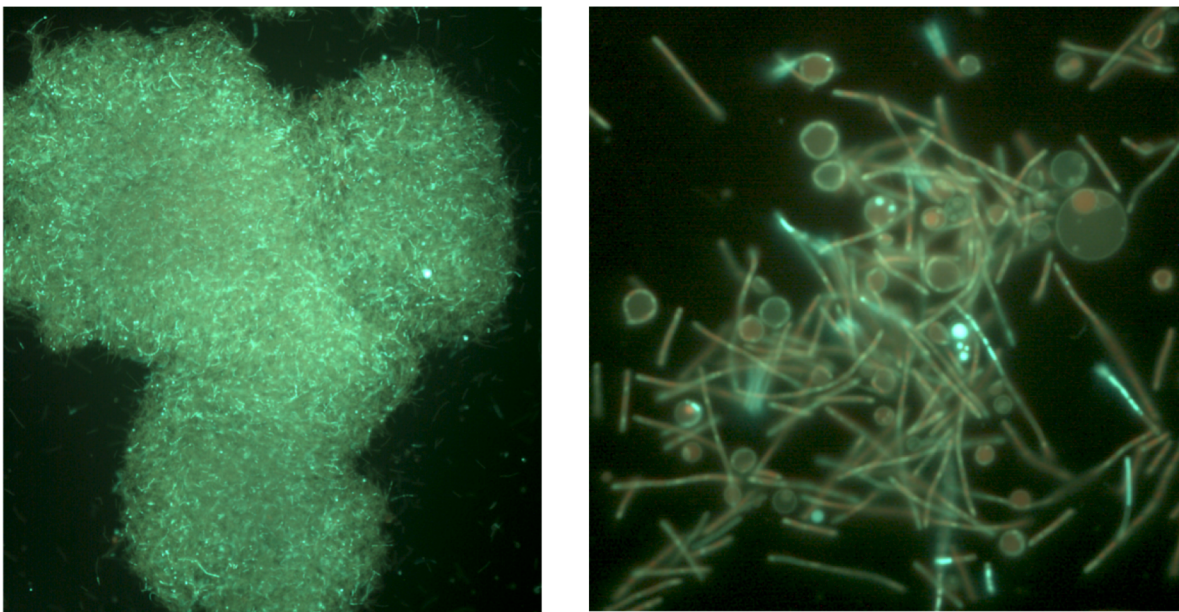
When grown under anaerobic conditions (50 ml media in 50 ml conical, screw cap tube or 1000 ml media in 1000 ml screw cap bottle, 70°C no agitation), the Y51MC23 grows in yellow flocculent clumps on the bottom of the container that are visible to the naked eye (Fig 7). Fluorescence microscopy of these clumps shows Y51MC23 produces a mixture of unique cell structures in high concentrations (Fig 8). The microscopic examination shows the presence of rods, but also large and small spherical bodies. Some of the spherical bodies appear strongly



**Fig 6. Micrograph of *Thermus aquaticus* Y51MC23 cells from aerobic cultures.** Culture samples were stained with SYTO<sup>®</sup> 9 fluorescent stain in sterile water (Molecular Probes). Dark field fluorescence microscopy was performed using a Nikon Eclipse TE2000-S epifluorescence microscope at 2000x magnification and a high-pressure Hg light source.

doi:10.1371/journal.pone.0138674.g006

fluorescent, while others have a fluorescent outline and dark center. The spheres appear to have a random distribution of sizes within the layers, and some spheres appear to have smaller spheres within their walls. In addition to the numerous small spherical bodies, rod-shaped cells



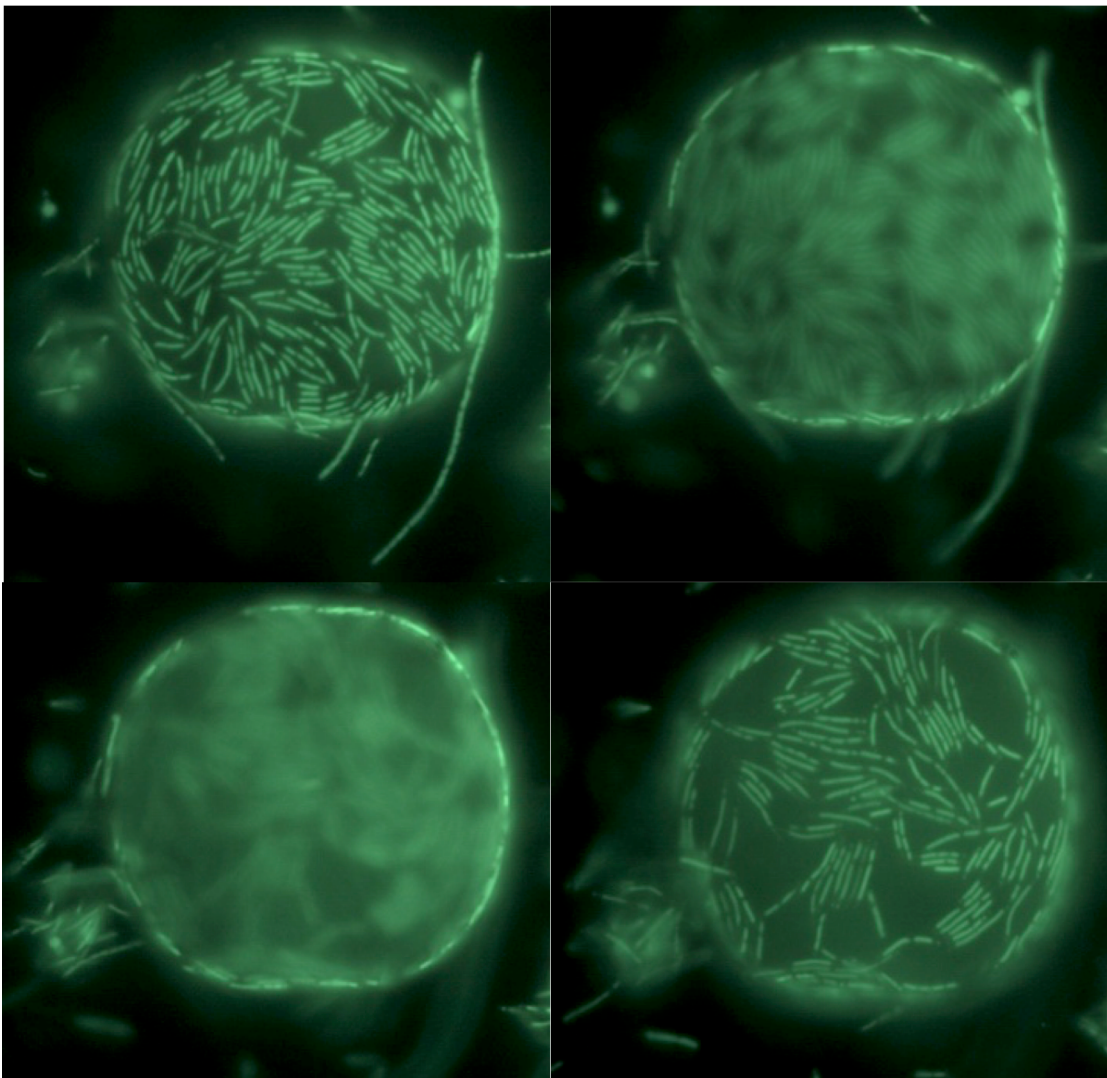
**Fig 7. Micrograph of *Thermus aquaticus* Y51MC23 from anaerobic cultures.** Culture sample was stained with SYTO<sup>®</sup> 9 fluorescent stain in sterile water (Molecular Probes). Dark field fluorescence microscopy was performed using a Nikon Eclipse TE2000-S epifluorescence microscope at 200X magnification (left) or 2000X magnification (right) and a high-pressure Hg light source (484 nm excitation and 500 nm emission filters).

doi:10.1371/journal.pone.0138674.g007

overlying a tile-like layer of spherical cell bodies can be found (Fig 7), similar to but much larger than the rotund bodies described by Brock [12].

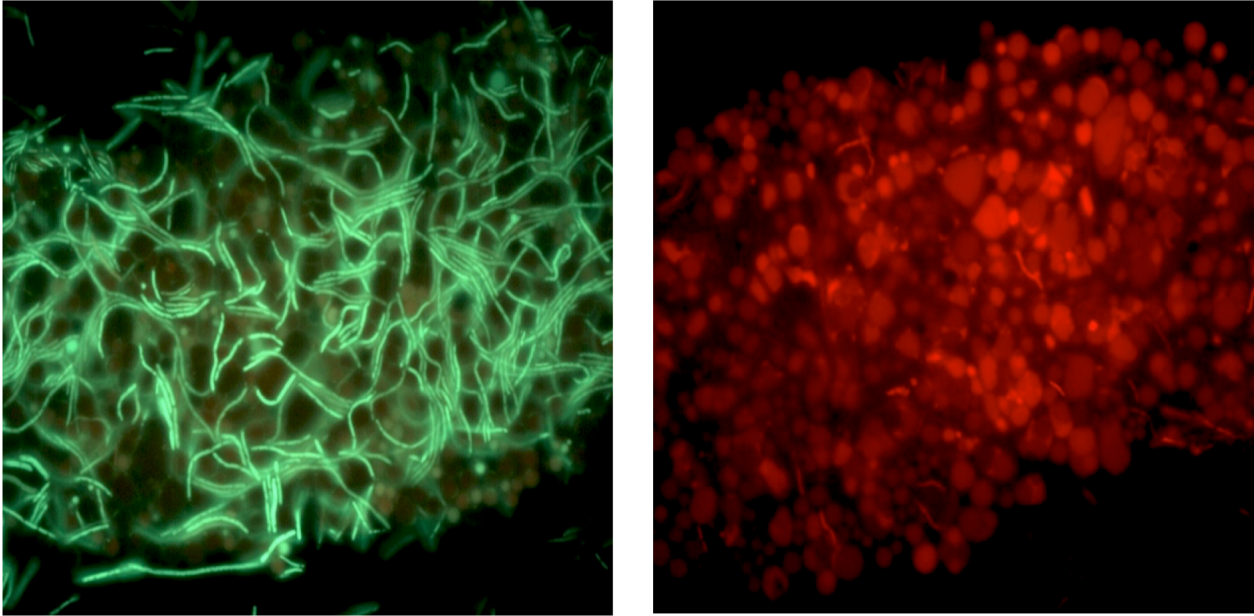
Clumps of cells were treated with Live-Dead<sup>®</sup> stain to improve visualization of the spherical bodies which often appear significantly less fluorescent than rod-shaped cells when stained with SYTO<sup>®</sup> 9. The multilayer structure is clearly seen as a layer of green rods on top of a matrix of red-staining spheres (Fig 9). This suggests that the multiple types of spheres seen in the micrographs may serve multiple discrete functions, including a nutrient reservoir (as in *Meiothermus ruber* [13, 14]), a biofilm-like structure to anchor cells to each other and solid surfaces, and as a source of persister cells or primitive spores [85, 86].

Microscopy suggests that some of the highly fluorescent spherical bodies are formed by asymmetrical division of cells (Fig 10). In one scenario, an elongated cell forms in the culture (A). The elongated cell develops a swollen section, either terminal or subterminal (B, C). The swollen end separates as an irregular shaped-object (D) and appears to gradually change shape to spherical. The final shapes are spherical, opaque fluorescent spheres of various sizes.



**Fig 8. Micrograph of *Thermus aquaticus* Y51MC23 round body found in anaerobic cultures.** Clockwise from left to right, views through a single large round body. Culture samples were stained with SYTO<sup>®</sup> 9 fluorescent stain in sterile water (Molecular Probes).

doi:10.1371/journal.pone.0138674.g008



**Fig 9. Micrograph of *Thermus aquaticus* Y51MC23 layered structure in anaerobic culture.** Clumps of cells were re-suspended in sterile water and stained with SYTO<sup>®</sup> 9 (green fluorescence) using 484 nm excitation and 500 nm emission filters (left panel) or propidium iodide (red fluorescence) using 536 nm excitation and 617 nm emission filters (right panel). Dark field fluorescence microscopy was performed using a Nikon Eclipse TE2000-S epifluorescence microscope at 2000X magnification and a high-pressure Hg light source.

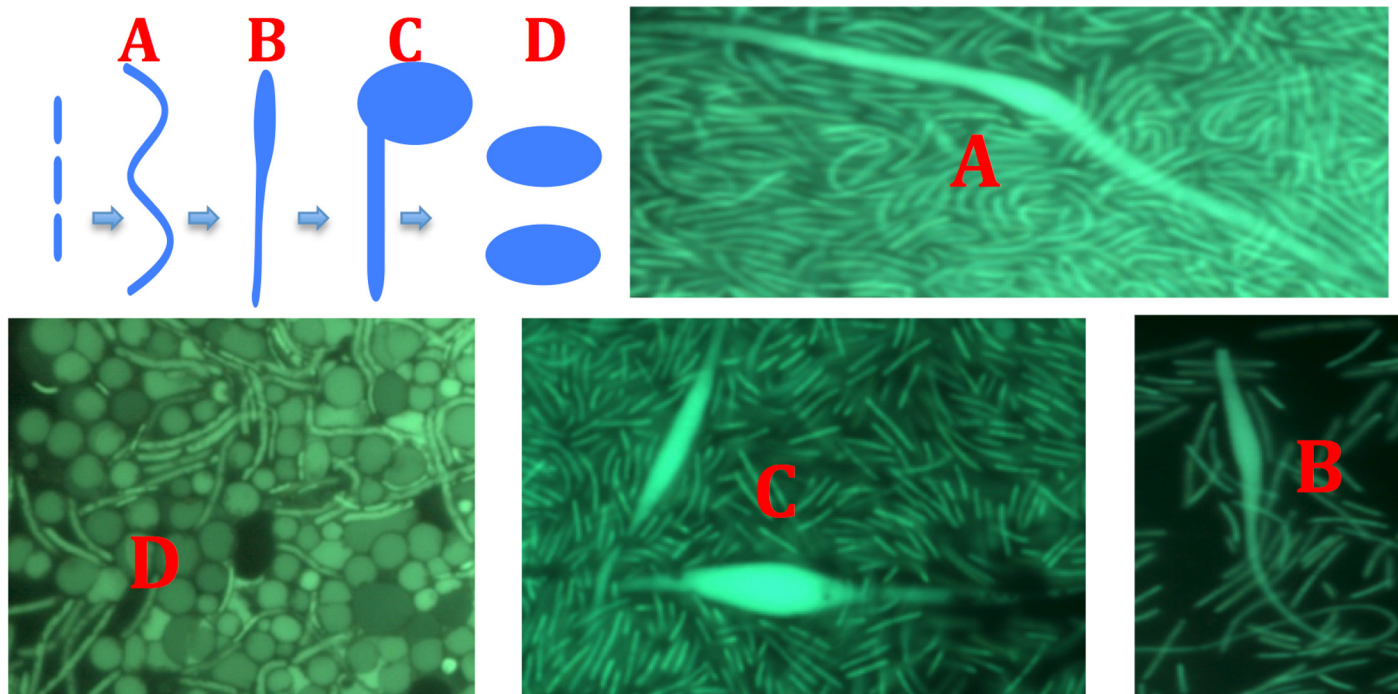
doi:10.1371/journal.pone.0138674.g009

Previous work using electron microscopy and visible light microscopy led to the theory that *Thermus* and *Meiothermus* spheres were formed by dissolution of the cell wall and retention of the outer envelope [12, 15]. The large spheres (rotund bodies) were formed by condensation of multiple outer envelopes. The presence of strongly-staining rings around the spheres seen in fluorescent micrographs (Fig 6) suggested that at least some of the spheres retained a thick outer structure, possibly the peptidoglycan wall. Cultures were stained with ViaGram Red<sup>+</sup> Bacterial Gram Stain and Viability Kit, and examined for red fluorescence from binding of Texas Red-X dye-labeled wheat germ agglutinin (WGA) to exposed peptidoglycan. *T. aquaticus* vegetative cells, being Gram negative, would be expected to not bind the Texas Red-X dye-labeled WGA. The results (Fig 11) show Texas Red-X dye-labeled WGA binds strongly to the outside of many of the spheres. This indicates that spheres are formed by the loss of the outer membrane and cell envelope and remodeling of the peptidoglycan to the new shape. The presence of an intact, possibly thicker, peptidoglycan wall may also explain the stability of these spheres to osmotic and other stresses.

## Conclusions

*Thermus aquaticus*, the first thermophilic bacterium shown to grow at temperatures well over 55°C, started a revolution in thermal biology with its initial description in 1969 [1]. The ubiquitous presence of *Thermus* species [2] suggests the genus has remarkable survival adaptations not present in other organisms, resulting in extensive research to understand its role as a model organism for life at high temperatures [87].

We report here the isolation and complete genome sequence of *T. aquaticus* Y51MC23, a new strain isolated from the outflow of Bath Hot Spring in Yellowstone National Park. This genome was closed and finished utilizing two NGS libraries, a 350 bp fragment library, and an 8 kb mate pair library, which produced a single chromosome scaffold and four plasmids. The



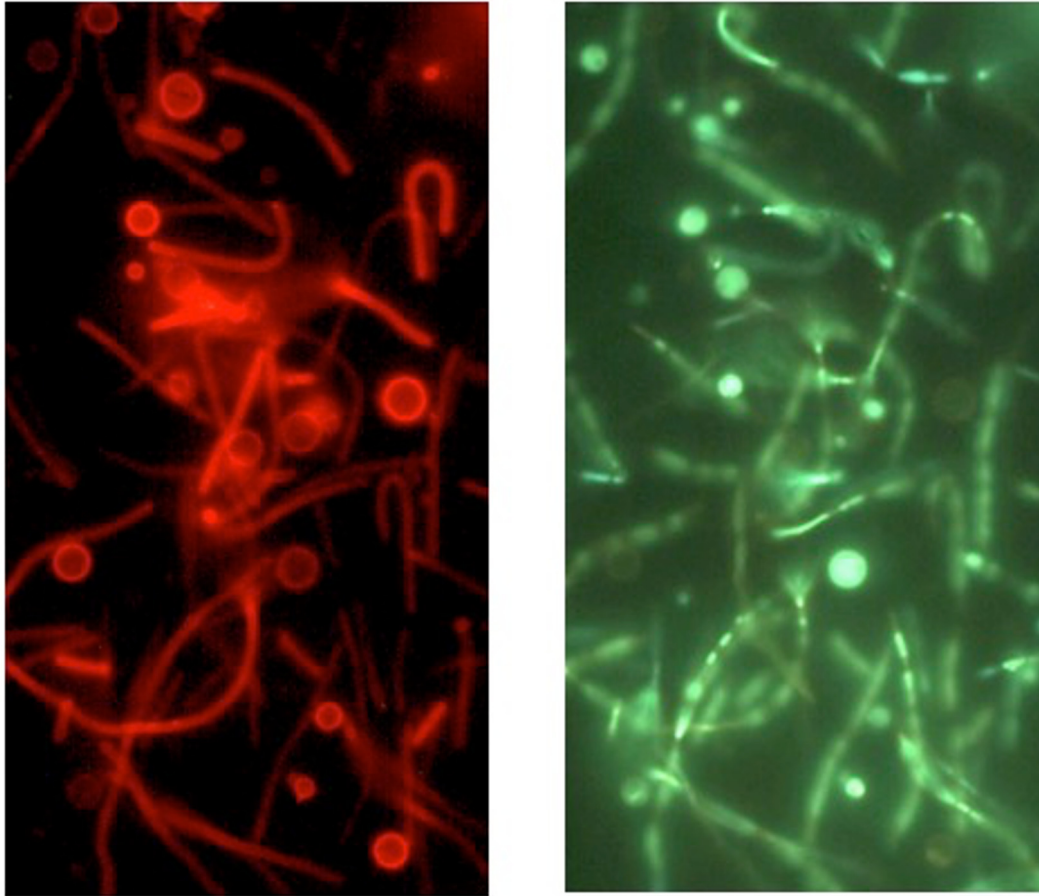
**Fig 10. Formation of highly fluorescent spheres in Y51MC23 culture.** Clockwise from top left. 1. Schematic of proposed mechanism of sphere formation. 2. Elongated cell (A) and swollen regions (B, C). 3. Highly fluorescent spheres (D). Culture samples were stained with SYTO<sup>®</sup> 9 fluorescent stain in sterile water (Molecular Probes).

doi:10.1371/journal.pone.0138674.g010

previous draft version could not distinguish the five separate elements and the ends of the numerous contigs often resulted in misassembled genes. A closed and finished genome simplifies comparative genomics and accurate gene annotation.

The genomes of *Thermus* species are significantly different from those of typical bacteria. *Thermus thermophilus* has been reported to possess multiple copies of the chromosome [88] as well as a 230 kb “megaplasmid” [17]. The role of the megaplasmids is believed to act as a storage site for genes related to thermophily [16]. The closed and finished genome sequence of *Thermus aquaticus* Y51MC23 shows similarities to other sequenced *Thermus* species in both the number and function of genes present in the genomes as well as short-range genomic organization. *T. aquaticus* Y51MC23 possesses four distinct plasmids, but no detectable megaplasmids. The generation of a finished, closed genome allowed the determination that many of the functions present on the *T. thermophilus* megaplasmids, such as cobalamin and phytoene biosynthetic clusters are present and located on the Y51MC23 chromosome. The organization of these clusters suggests they have arisen from stable integration of a megaplasmid into the Y51MC23 chromosome.

*T. aquaticus* Y51MC23 possesses significant metabolic differences from the Brock YT-1 strain, and most strains of *Thermus* in being capable of both aerobic and true fermentative growth. Most strains of *Thermus* are obligate aerobes that can also utilize nitrate as an alternative terminal electron acceptor [87, 89, 90]. The ability to utilize nitrate as an electron acceptor is conferred by a DNA fragment, dubbed the “nitrate respiratory conjugative element”, which encodes nitrate reductase and various proteins required for its activity [91]. *Thermus aquaticus* Y51MC23 possesses no orthologs of these nitrate reductase genes. Metabolic analysis suggests that Y51MC23 primarily scavenges protein from the environment, based on the high number of secreted and intracellular proteases and peptidases as well as transporter systems for amino



**Fig 11. Peptidoglycan staining of Y51MC23 culture.** Clumps of cells were re-suspended in sterile water and stained with SYTOX Green (green fluorescence) using 484 nm excitation and 500 nm emission filters (right panel) or Texas Red-X dye-labeled WGA (red fluorescence) using 536 nm excitation and 617 nm emission filters (left panel). Dark field fluorescence microscopy was performed using a Nikon Eclipse TE2000-S epifluorescence microscope at 2000X magnification and a high-pressure Hg light source.

doi:10.1371/journal.pone.0138674.g011

acids and peptides. Under aerobic conditions, amino acids are most likely deaminated and the carbon backbone is oxidized using the citric acid cycle. It is unclear how ATP is generated and redox balance is maintained under anaerobic conditions.

The Y51MC23 genome shows the presence of two recognizable prophage inserts and an additional two putative prophage inserts as well as numerous CRISPR loci. The high homology and synteny between Y51MC23 prophage 2 and that of *Thermus sp.* 2.9 is ecologically interesting, given the 8,800 km separation of the two hot springs.

The anaerobic lifestyle of Y51MC23 is complex, with multiple morphologies present in cultures. The use of fluorescence microscopy reveals new details about these unusual morphological features not seen before, including the presence of multiple types of large and small spheres, often forming a confluent layer of spheres. Previous work showing rotund bodies was performed using electron microscopy, which may have destroyed many of the structural details demonstrated here using conventional light microscopy. Use of multiple dyes with different molecular specificities allowed multiple features on the same cells to be visualized for the first time. Many of the spheres seen in the micrographs appear to be formed not from cell envelope or outer membrane components [12, 15], but from a remodeled peptidoglycan cell wall. These morphological forms may serve multiple functions in the survival of the organism. Some of the



spherical bodies may be a food storage system as suggested for *M. ruber* [13, 14]. The interlocking layer of spheres may represent an extremely resilient biofilm backbone, capable of anchoring cells to surfaces and resisting high temperatures and high liquid flow rates. Finally, some of the spheres may serve as a protective coating for genomic DNA, analogous to spores formed in Gram-positive organisms.

Numerous questions remain to be answered about Y51MC23. The source of ATP generation from amino acid fermentation remains to be discovered. How the individual cellular morphologies are generated and what molecular controls are involved in selecting morphologies remains to be elucidated. The function of the unusual high MW hypothetical proteins that are conserved in the *Deinococcus-Thermus* phylogenetic group also remains as a challenge to better understanding the ability to these organisms to grow and survive in high temperature, nutrient-poor, fast-flowing environments.

Thermophilic enzymes such as Taq DNA polymerase and others have proven useful for numerous industrial applications [92]. The presence of an annotated reverse transcriptase located in prophage 2 of *T. aquaticus* Y51MC23 (gene 1131) is of potential utility for molecular biology applications, but the protein is completely insoluble during preliminary expression experiments in our hands (data not shown). The presence of an annotated bifunctional primase/polymerase (gene 2152) also has potential utility for whole genome amplification depending on the appropriate biochemical characteristics.

## Supporting Information

**S1 Fig. PCR verification of contig order and orientation.**  
(DOCX)

**S1 Table. PCR primers used to cross areas of uncertainty.**  
(DOCX)

**S2 Table. Comparison of the finished chromosome and its four attendant plasmids and the draft assembly.**  
(PDF)

## Acknowledgments

This project was supported by an SBIR grant award (1R43HG007797-01A1) from the National Human Genome Research Institute. We thank Brian Hedlund and Scott Thomas for critical reading of the manuscript.

## Author Contributions

Conceived and designed the experiments: DAM PJB SM BK TS. Performed the experiments: PJB SM BK SJ EF ML. Analyzed the data: PJB SM BK DAM. Wrote the paper: PJB SM DAM BK.

## References

1. Brock TD, Freeze H. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J Bacteriol.* 1969; 98(1):289–97. PMID: [5781580](#); PubMed Central PMCID: PMC249935.
2. Brock TD, Boylen KL. Presence of thermophilic bacteria in laundry and domestic hot-water heaters. *Applied microbiology.* 1973; 25(1):72–6. Epub 1973/01/01. PMID: [4568892](#); PubMed Central PMCID: PMC249935.
3. Williams RA, Smith KE, Welch SG, Micallef J, Sharp RJ. DNA relatedness of *Thermus* strains, description of *Thermus brockianus* sp. nov., and proposal to reestablish *Thermus thermophilus* (Oshima and Imahori). *Int J Syst Bacteriol.* 1995; 45(3):495–9. Epub 1995/07/01. PMID: [8590676](#).

4. Williams RA, Smith KE, Welch SG, Micallef J. *Thermus oshimai* sp. nov., isolated from hot springs in Portugal, Iceland, and the Azores, and comment on the concept of a limited geographical distribution of *Thermus* species. *Int J Syst Bacteriol*. 1996; 46(2):403–8. Epub 1996/04/01. PMID: [8934898](#).
5. Ming H, Yin YR, Li S, Nie GX, Yu TT, Zhou EM, et al. *Thermus caliditerrae* sp. nov., a novel thermophilic species isolated from a geothermal area. *Int J Syst Evol Microbiol*. 2014; 64(Pt 2):650–6. Epub 2013/10/26. doi: [10.1099/ijs.0.056838-0](#) PMID: [24158953](#).
6. Zhang XQ, Ying Y, Ye Y, Xu XW, Zhu XF, Wu M. *Thermus arciformis* sp. nov., a thermophilic species from a geothermal area. *Int J Syst Evol Microbiol*. 2010; 60(Pt 4):834–9. Epub 2009/08/08. doi: [10.1099/ijs.0.007690-0](#) PMID: [19661520](#).
7. Bjornsdottir SH, Petursdottir SK, Hreggvidsson GO, Skirmisdottir S, Hjorleifsdottir S, Arnfinnsson J, et al. *Thermus islandicus* sp. nov., a mixotrophic sulfur-oxidizing bacterium isolated from the Torfajokull geothermal area. *Int J Syst Evol Microbiol*. 2009; 59(Pt 12):2962–6. Epub 2009/07/25. doi: [10.1099/ijs.0.007013-0](#) PMID: [19628590](#).
8. Chung AP, Rainey FA, Valente M, Nobre MF, da Costa MS. *Thermus igniterrae* sp. nov. and *Thermus antranikianii* sp. nov., two new species from Iceland. *Int J Syst Evol Microbiol*. 2000; 50 Pt 1:209–17. Epub 2000/05/29. PMID: [10826806](#).
9. Vajna B, Kanizsai S, Keki Z, Marialigeti K, Schumann P, Toth EM. *Thermus composti* sp. nov., isolated from oyster mushroom compost. *Int J Syst Evol Microbiol*. 2012; 62(Pt 7):1486–90. Epub 2011/08/23. doi: [10.1099/ijs.0.030866-0](#) PMID: [21856987](#).
10. Gounder K, Brzuszkiewicz E, Liesegang H, Wollherr A, Daniel R, Gottschalk G, et al. Sequence of the hyperplastic genome of the naturally competent *Thermus scotoeductus* SA-01. *BMC Genomics*. 2011; 12:577. Epub 2011/11/26. doi: [10.1186/1471-2164-12-577](#) PMID: [22115438](#); PubMed Central PMCID: PMC23235269.
11. Nobre MF, Truper HG, da Costa MS. Transfer of *Thermus ruber* (Loginova et al. 1984), *Thermus silvanus* (Tenreiro et al. 1995), and *Thermus chliarophilus* (Tenreiro et al. 1995) to *Meiothermus* gen. nov. as *Meiothermus ruber* comb. nov., *Meiothermus silvanus* comb. nov., and *Meiothermus chliarophilus* comb. nov., Respectively, and Emendation of the Genus *Thermus*. *Int J Syst Evol Microbiol*. 1996; 46:604–6.
12. Brock TD, Edwards MR. Fine structure of *Thermus aquaticus*, an extreme thermophile. *J Bacteriol*. 1970; 104(1):509–17. PMID: [5473907](#); PubMed Central PMCID: PMC248237.
13. Golovacheva RS, Pivovarova TA. [Speroplast behavior in cultures of *Thermus ruber*]. *Mikrobiologiya*. 1977; 46(6):1019–27. Epub 1977/11/01. PMID: [600101](#).
14. Golovacheva RS. [Complex spherical bodies of *Thermus ruber*]. *Mikrobiologiya*. 1977; 46(3):506–12. Epub 1977/05/01. PMID: [895560](#).
15. Golovacheva RS, Pivovarova TA. [Cocoid cells and spheroplasts in cultures of the genus *Thermus*]. *Mikrobiologiya*. 1977; 46(4):695–702. Epub 1977/07/01. PMID: [909468](#).
16. Bruggemann H, Chen C. Comparative genomics of *Thermus thermophilus*: Plasticity of the megaplasmid and its contribution to a thermophilic lifestyle. *J Biotechnol*. 2006; 124(4):654–61. doi: [10.1016/j.jbiotec.2006.03.043](#) PMID: [16713647](#).
17. Henne A, Bruggemann H, Raasch C, Wiezer A, Hartsch T, Liesegang H, et al. The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nature biotechnology*. 2004; 22(5):547–53. Epub 2004/04/06. doi: [10.1038/nbt956](#) PMID: [15064768](#).
18. Murugapiran SK, Huntemann M, Wei CL, Han J, Detter JC, Han C, et al. *Thermus oshimai* JL-2 and *T. thermophilus* JL-18 genome analysis illuminates pathways for carbon, nitrogen, and sulfur cycling. *Stand Genomic Sci*. 2013; 7(3):449–68. Epub 2013/09/11. doi: [10.4056/signs.3667269](#) PMID: [24019992](#); PubMed Central PMCID: PMC23764938.
19. Murugapiran SK, Huntemann M, Wei CL, Han J, Detter JC, Han CS, et al. Whole Genome Sequencing of *Thermus oshimai* JL-2 and *Thermus thermophilus* JL-18, Incomplete Denitrifiers from the United States Great Basin. *Genome Announc*. 2013; 1(1). Epub 2013/02/14. doi: [10.1128/genomeA.00106-12](#) PMID: [23405355](#); PubMed Central PMCID: PMC23569359.
20. Dwivedi V, Sangwan N, Nigam A, Garg N, Niharika N, Khurana P, et al. Draft genome sequence of *Thermus* sp. strain RL, isolated from a hot water spring located atop the Himalayan ranges at Manikaran, India. *J Bacteriol*. 2012; 194(13):3534. Epub 2012/06/13. doi: [10.1128/jb.00604-12](#) PMID: [22689228](#); PubMed Central PMCID: PMC23434741.
21. Teh BS, Abdul Rahman AY, Saito JA, Hou S, Alam M. Complete genome sequence of the thermophilic bacterium *Thermus* sp. strain CCB\_US3\_UF1. *J Bacteriol*. 2012; 194(5):1240. Epub 2012/02/14. doi: [10.1128/jb.06589-11](#) PMID: [22328745](#); PubMed Central PMCID: PMC23294796.
22. Navas LE, Berretta MF, Ortiz EM, Benintende GB, Amadio AF, Zandomeni RO. Draft Genome Sequence of *Thermus* sp. Isolate 2.9, Obtained from a Hot Water Spring Located in Salta, Argentina.

- Genome Announc. 2015; 3(1). doi: [10.1128/genomeA.01414-14](https://doi.org/10.1128/genomeA.01414-14) PMID: [25593256](https://pubmed.ncbi.nlm.nih.gov/25593256/); PubMed Central PMCID: PMC4299898.
23. Jiang L, Lin M, Li X, Cui H, Xu X, Li S, et al. Genome Sequence of *Thermus thermophilus* ATCC 33923, a Thermostable Trehalose-Producing Strain. *Genome Announc.* 2013; 1(4). Epub 2013/07/28. doi: [10.1128/genomeA.00493-13](https://doi.org/10.1128/genomeA.00493-13) PMID: [23887916](https://pubmed.ncbi.nlm.nih.gov/23887916/); PubMed Central PMCID: PMC3735056.
  24. Mead DA, Lucas S, Copeland A, Lapidus A, Cheng JF, Bruce DC, et al. Complete Genome Sequence of *Paenibacillus* strain Y4.12MC10, a Novel *Paenibacillus lautus* strain Isolated from Obsidian Hot Spring in Yellowstone National Park. *Stand Genomic Sci.* 2012; 6(3):381–400. Epub 2013/02/15. doi: [10.4056/sigs.2605792](https://doi.org/10.4056/sigs.2605792) PMID: [23408395](https://pubmed.ncbi.nlm.nih.gov/23408395/); PubMed Central PMCID: PMC3558958.
  25. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. NY: Cold Spring Harbor Laboratory Press; 1989.
  26. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* 2008; 9:75. Epub 2008/02/12. doi: [10.1186/1471-2164-9-75](https://doi.org/10.1186/1471-2164-9-75) PMID: [18261238](https://pubmed.ncbi.nlm.nih.gov/18261238/); PubMed Central PMCID: PMC2265698.
  27. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C. Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Methods Mol Biol.* 2013; 985:17–45. Epub 2013/02/19. doi: [10.1007/978-1-62703-299-5\\_2](https://doi.org/10.1007/978-1-62703-299-5_2) PMID: [23417797](https://pubmed.ncbi.nlm.nih.gov/23417797/).
  28. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 2014; 42(Database issue):D560–7. Epub 2013/10/30. doi: [10.1093/nar/gkt963](https://doi.org/10.1093/nar/gkt963) PMID: [24165883](https://pubmed.ncbi.nlm.nih.gov/24165883/); PubMed Central PMCID: PMC3965111.
  29. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25(17):3389–402. Epub 1997/09/01. PMID: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/); PubMed Central PMCID: PMC146917.
  30. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011; 8(10):785–6. Epub 2011/10/01. doi: [10.1038/nmeth.1701](https://doi.org/10.1038/nmeth.1701) PMID: [21959131](https://pubmed.ncbi.nlm.nih.gov/21959131/).
  31. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 2014; 42(Database issue):D503–9. doi: [10.1093/nar/gkt953](https://doi.org/10.1093/nar/gkt953) PMID: [24157837](https://pubmed.ncbi.nlm.nih.gov/24157837/); PubMed Central PMCID: PMC3964991.
  32. Barrett AJ, Rawlings ND, O'Brien EA. The MEROPS database as a protease information system. *Journal of structural biology.* 2001; 134(2–3):95–102. doi: [10.1006/jsbi.2000.4332](https://doi.org/10.1006/jsbi.2000.4332) PMID: [11551172](https://pubmed.ncbi.nlm.nih.gov/11551172/).
  33. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 2004; 32(Database issue):D115–9. Epub 2003/12/19. doi: [10.1093/nar/gkh131](https://doi.org/10.1093/nar/gkh131) PMID: [14681372](https://pubmed.ncbi.nlm.nih.gov/14681372/); PubMed Central PMCID: PMC308865.
  34. Consortium TU. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2014; 42(1):D191–8. Epub 2013/11/21. doi: [10.1093/nar/gkt1140](https://doi.org/10.1093/nar/gkt1140) PMID: [24253303](https://pubmed.ncbi.nlm.nih.gov/24253303/).
  35. Claudel-Renard C, Chevalet C, Faraut T, Kahn D. Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* 2003; 31(22):6633–9. PMID: [14602924](https://pubmed.ncbi.nlm.nih.gov/14602924/); PubMed Central PMCID: PMC275543.
  36. Langille MG, Brinkman FS. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics.* 2009; 25(5):664–5. Epub 2009/01/20. doi: [10.1093/bioinformatics/btp030](https://doi.org/10.1093/bioinformatics/btp030) PMID: [19151094](https://pubmed.ncbi.nlm.nih.gov/19151094/); PubMed Central PMCID: PMC2647836.
  37. Hsiao W, Wan I, Jones SJ, Brinkman FS. IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics.* 2003; 19(3):418–20. Epub 2003/02/14. PMID: [12584130](https://pubmed.ncbi.nlm.nih.gov/12584130/).
  38. Bose M, Barber RD. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol.* 2006; 6(3):223–7. PMID: [16922685](https://pubmed.ncbi.nlm.nih.gov/16922685/).
  39. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1993; 10(3):512–26. Epub 1993/05/01. PMID: [8336541](https://pubmed.ncbi.nlm.nih.gov/8336541/).
  40. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28(10):2731–9. Epub 2011/05/07. doi: [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121) PMID: [21546353](https://pubmed.ncbi.nlm.nih.gov/21546353/); PubMed Central PMCID: PMC3203626.
  41. Brumm P, Land M, Hauser LJ, Jeffries C, Chang YJ, Mead D. Complete Genome Sequence of *Geobacillus* strain Y4.1MC1, a Novel CO-Utilizing *Geobacillus thermoglucosidasius* Strain Isolated from Bath Hot Spring in Yellowstone National Park. *Bioenerg Res.* 2015. Epub 2/10/2015. doi: [10.1007/s12155-015-9585-2](https://doi.org/10.1007/s12155-015-9585-2)

42. McCleskey C, Ball J, Nordstrom DK, Holloway JM, Taylor HE. Water-Chemistry Data for Selected Hot Springs, Geysers, and Streams in Yellowstone National Park, Wyoming, 2001–2002. Open-File Report 2004–1316 [Internet]. 2004. Available: <http://pubs.usgs.gov/of/2004/1316/>.
43. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32(5):1792–7. Epub 2004/03/23. doi: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340) PMID: [15034147](https://pubmed.ncbi.nlm.nih.gov/15034147/); PubMed Central PMCID: PMC390337.
44. Kumwenda B, Litthauer D, Reva O. Analysis of genomic rearrangements, horizontal gene transfer and role of plasmids in the evolution of industrial important *Thermus* species. *BMC Genomics.* 2014; 15(1). doi: [10.1186/1471-2164-15-813](https://doi.org/10.1186/1471-2164-15-813) PMID: [25257245](https://pubmed.ncbi.nlm.nih.gov/25257245/); PubMed Central PMCID: PMC34180962.
45. van Heesch S, Kloosterman WP, Lansu N, Ruzius FP, Levandowsky E, Lee CC, et al. Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *BMC Genomics.* 2013; 14:257. doi: [10.1186/1471-2164-14-257](https://doi.org/10.1186/1471-2164-14-257) PMID: [23590730](https://pubmed.ncbi.nlm.nih.gov/23590730/); PubMed Central PMCID: PMC3648348.
46. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics.* 2009; 25(1):19–20. doi: [10.1093/bioinformatics/btn578](https://doi.org/10.1093/bioinformatics/btn578) PMID: [18990721](https://pubmed.ncbi.nlm.nih.gov/18990721/); PubMed Central PMCID: PMC2612626.
47. Saurin W, Hofnung M, Dassa E. Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *Journal of molecular evolution.* 1999; 48(1):22–41. Epub 1999/01/05. PMID: [9873074](https://pubmed.ncbi.nlm.nih.gov/9873074/).
48. Karpowich NK, Wang D-N. Assembly and mechanism of a group II ECF transporter. *Proc Natl Acad Sci U S A.* 2013; 110(7):2534–9. doi: [10.1073/pnas.1217361110](https://doi.org/10.1073/pnas.1217361110) PMC3574940. PMID: [23359690](https://pubmed.ncbi.nlm.nih.gov/23359690/)
49. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014; 42(Database issue):D490–5. Epub 2013/11/26. doi: [10.1093/nar/gkt1178](https://doi.org/10.1093/nar/gkt1178) PMID: [24270786](https://pubmed.ncbi.nlm.nih.gov/24270786/); PubMed Central PMCID: PMC3965031.
50. Radax C, Sigurdsson O, Hreggvidsson GO, Aichinger N, Gruber C, Kristjansson JK, et al. F- and V-ATPases in the genus *Thermus* and related species. *Syst Appl Microbiol.* 1998; 21(1):12–22. Epub 1998/09/19. doi: [10.1016/s0723-2020\(98\)80003-9](https://doi.org/10.1016/s0723-2020(98)80003-9) PMID: [9741106](https://pubmed.ncbi.nlm.nih.gov/9741106/).
51. Ponting CP. Evidence for PDZ domains in bacteria, yeast, and plants. *Protein Sci.* 1997; 6(2):464–8. doi: [10.1002/pro.5560060225](https://doi.org/10.1002/pro.5560060225) PMID: [9041651](https://pubmed.ncbi.nlm.nih.gov/9041651/); PubMed Central PMCID: PMC2143646.
52. Valas RE, Bourne PE. Rethinking proteasome evolution: two novel bacterial proteasomes. *Journal of molecular evolution.* 2008; 66(5):494–504. Epub 2008/04/05. doi: [10.1007/s00239-008-9075-7](https://doi.org/10.1007/s00239-008-9075-7) PMID: [18389302](https://pubmed.ncbi.nlm.nih.gov/18389302/); PubMed Central PMCID: PMC3235984.
53. Valas RE, Bourne PE. Rethinking proteasome evolution: two novel bacterial proteasomes. (0022–2844 (Print)). doi: D—NLM: PMC3235984 EDAT- 2008/04/05 09:00 MHDA- 2008/10/01 09:00 CRDT- 2008/04/05 09:00 PHST- 2007/09/18 [received] PHST- 2008/01/25 [accepted] PHST- 2008/01/23 [revised] PHST- 2008/04/04 [aheadofprint] AID— doi: [10.1007/s00239-008-9075-7](https://doi.org/10.1007/s00239-008-9075-7) PST—ppublish.
54. Gur E, Sauer RT. Recognition of misfolded proteins by Lon, a AAA(+) protease. (0890–9369 (Print)). doi: D—NLM: PMC2518814 EDAT- 2008/08/19 09:00 MHDA- 2008/09/17 09:00 CRDT- 2008/08/19 09:00 AID—22/16/2267 [pii] AID— doi: [10.1101/gad.1670908](https://doi.org/10.1101/gad.1670908) PST—ppublish.
55. Xie G, Keyhani NO, Bonner CA, Jensen RA. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev.* 2003; 67(3):303–42, table of contents. Epub 2003/09/11. PMID: [12966138](https://pubmed.ncbi.nlm.nih.gov/12966138/); PubMed Central PMCID: PMC3193870.
56. Tian B, Hua Y. Carotenoid biosynthesis in extremophilic *Deinococcus*–*Thermus* bacteria. *Trends in microbiology.* 18(11):512–20. doi: [10.1016/j.tim.2010.07.007](https://doi.org/10.1016/j.tim.2010.07.007) PMID: [20832321](https://pubmed.ncbi.nlm.nih.gov/20832321/)
57. Tabata K, Ishida S Fau—Nakahara T, Nakahara T Fau—Hoshino T, Hoshino T. A carotenogenic gene cluster exists on a large plasmid in *Thermus thermophilus*. (0014–5793 (Print)).
58. Takano H, Agari Y, Hagiwara K, Watanabe R, Yamazaki R, Beppu T, et al. LdrP, a cAMP receptor protein/FNR family transcriptional regulator, serves as a positive regulator for the light-inducible gene cluster in the megaplasmid of *Thermus thermophilus*. (1465–2080 (Electronic)).
59. Fuhrman JA, Schwalbach Ms Fau—Stingl U, Stingl U. Proteorhodopsins: an array of physiological roles? (1740–1534 (Electronic)).
60. de la Cruz F, Frost LS, Meyer RJ, Zechner EL. Conjugative DNA metabolism in Gram-negative bacteria. *FEMS Microbiol Rev.* 2010; 34(1):18–40. Epub 2009/11/19. doi: [10.1111/j.1574-6976.2009.00195.x](https://doi.org/10.1111/j.1574-6976.2009.00195.x) PMID: [19919603](https://pubmed.ncbi.nlm.nih.gov/19919603/).
61. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. *Cellular microbiology.* 2010; 12(9):1203–12. doi: [10.1111/j.1462-5822.2010.01499.x](https://doi.org/10.1111/j.1462-5822.2010.01499.x) PMID: [20642798](https://pubmed.ncbi.nlm.nih.gov/20642798/); PubMed Central PMCID: PMC3070162.
62. Chandran V. Type IV secretion machinery: molecular architecture and function. *Biochem Soc Trans.* 2013; 41(1):17–28. Epub 2013/01/30. doi: [10.1042/bst20120332](https://doi.org/10.1042/bst20120332) PMID: [23356253](https://pubmed.ncbi.nlm.nih.gov/23356253/).

63. Yu MX, Slater MR, Ackermann HW. Isolation and characterization of *Thermus* bacteriophages. *Archives of virology*. 2006; 151(4):663–79. Epub 2005/11/26. doi: [10.1007/s00705-005-0667-x](https://doi.org/10.1007/s00705-005-0667-x) PMID: [16308675](https://pubmed.ncbi.nlm.nih.gov/16308675/).
64. Hong W, Han J, Dai X, Ji X, Wei Y, Lin L. [Isolation and characterization of a *Thermus* bacteriophage lytic from Tengchong Rehai Hot Spring lytic]. *Wei Sheng Wu Xue Bao*. 2010; 50(3):322–7. PMID: [20499636](https://pubmed.ncbi.nlm.nih.gov/20499636/).
65. Berdygulova Z, Westblade LF, Florens L, Koonin EV, Chait BT, Ramanculov E, et al. Temporal regulation of gene expression of the *Thermus thermophilus* bacteriophage P23-45. *J Mol Biol*. 2011; 405(1):125–42. doi: [10.1016/j.jmb.2010.10.049](https://doi.org/10.1016/j.jmb.2010.10.049) PMID: [21050864](https://pubmed.ncbi.nlm.nih.gov/21050864/); PubMed Central PMCID: PMC3018760.
66. Lin L, Hong W, Ji X, Han J, Huang L, Wei Y. Isolation and characterization of an extremely long tail *Thermus* bacteriophage from Tengchong hot springs in China. *J Basic Microbiol*. 2010; 50(5):452–6. doi: [10.1002/jobm.201000116](https://doi.org/10.1002/jobm.201000116) PMID: [20806260](https://pubmed.ncbi.nlm.nih.gov/20806260/).
67. Matsushita I, Yanase H. The genomic structure of *thermus* bacteriophage {phi}IN93. *J Biochem*. 2009; 146(6):775–85. doi: [10.1093/jb/mvp125](https://doi.org/10.1093/jb/mvp125) PMID: [19675097](https://pubmed.ncbi.nlm.nih.gov/19675097/).
68. Naryshkina T, Liu J, Florens L, Swanson SK, Pavlov AR, Pavlova NV, et al. *Thermus thermophilus* bacteriophage phiYS40 genome and proteomic characterization of virions. *J Mol Biol*. 2006; 364(4):667–77. doi: [10.1016/j.jmb.2006.08.087](https://doi.org/10.1016/j.jmb.2006.08.087) PMID: [17027029](https://pubmed.ncbi.nlm.nih.gov/17027029/); PubMed Central PMCID: PMC1773054.
69. Overman SA, Bondre P, Maiti NC, Thomas GJ Jr. Structural characterization of the filamentous bacteriophage PH75 from *Thermus thermophilus* by Raman and UV-resonance Raman spectroscopy. *Biochemistry*. 2005; 44(8):3091–100. doi: [10.1021/bi048163d](https://doi.org/10.1021/bi048163d) PMID: [15723554](https://pubmed.ncbi.nlm.nih.gov/15723554/).
70. Sakaki Y, Oshima T. Isolation and characterization of a bacteriophage infectious to an extreme thermophile, *Thermus thermophilus* HB8. *Journal of virology*. 1975; 15(6):1449–53. PMID: [1142476](https://pubmed.ncbi.nlm.nih.gov/1142476/); PubMed Central PMCID: PMC354612.
71. Tamakoshi M, Murakami A, Sugisawa M, Tsuneizumi K, Takeda S, Saheki T, et al. Genomic and proteomic characterization of the large Myoviridae bacteriophage varphiTMA of the extreme thermophile *Thermus thermophilus*. *Bacteriophage*. 2011; 1(3):152–64. doi: [10.4161/bact.1.3.16712](https://doi.org/10.4161/bact.1.3.16712) PMID: [22164349](https://pubmed.ncbi.nlm.nih.gov/22164349/); PubMed Central PMCID: PMC3225780.
72. Schillinger T, Zingler N. The low incidence of diversity-generating retroelements in sequenced genomes. *Mobile genetic elements*. 2012; 2(6):287–91. doi: [10.4161/mge.23244](https://doi.org/10.4161/mge.23244) PMID: [23481467](https://pubmed.ncbi.nlm.nih.gov/23481467/); PubMed Central PMCID: PMC3575424.
73. Zhang J, White MF. Hot and crispy: CRISPR-Cas systems in the hyperthermophile *Sulfolobus solfataricus*. *Biochem Soc Trans*. 2013; 41(6):1422–6. doi: [10.1042/BST20130031](https://doi.org/10.1042/BST20130031) PMID: [24256231](https://pubmed.ncbi.nlm.nih.gov/24256231/).
74. Caston JR, Berenguer J, de Pedro MA, Carrascosa JL. S-layer protein from *Thermus thermophilus* HB8 assembles into porin-like structures. *Mol Microbiol*. 1993; 9(1):65–75. Epub 1993/07/01. PMID: [8412672](https://pubmed.ncbi.nlm.nih.gov/8412672/).
75. Fernández-Herrero LA, Olabarria G, Castón JR, Lasa I, Berenguer J. Horizontal transference of S-layer genes within *Thermus thermophilus*. *J Bacteriol*. 1995; 177(19):5460–6. PMID: [7559330](https://pubmed.ncbi.nlm.nih.gov/7559330/); PubMed Central PMCID: PMCPmc177352.
76. Fernandez-Herrero LA, Badet-Denisot MA, Badet B, Berenguer J. glmS of *Thermus thermophilus* HB8: an essential gene for cell-wall synthesis identified immediately upstream of the S-layer gene. *Mol Microbiol*. 1995; 17(1):1–12. Epub 1995/07/01. PMID: [7476196](https://pubmed.ncbi.nlm.nih.gov/7476196/).
77. Bender MH, Cartee RT, Yother J. Positive correlation between tyrosine phosphorylation of CpsD and capsular polysaccharide production in *Streptococcus pneumoniae*. *J Bacteriol*. 2003; 185(20):6057–66. Epub 2003/10/04. PMID: [14526017](https://pubmed.ncbi.nlm.nih.gov/14526017/); PubMed Central PMCID: PMCPmc225014.
78. Liszewski Zilla M, Chan YG, Lunderberg JM, Schneewind O, Missiakas D. LytR-CpsA-Psr Enzymes as Determinants of *Bacillus anthracis* Secondary Cell Wall Polysaccharide Assembly. *J Bacteriol*. 2015; 197(2):343–53. Epub 2014/11/12. doi: [10.1128/jb.02364-14](https://doi.org/10.1128/jb.02364-14) PMID: [25384480](https://pubmed.ncbi.nlm.nih.gov/25384480/); PubMed Central PMCID: PMCPmc4272586.
79. Beall B, Moran CP Jr. Cloning and characterization of spoVR, a gene from *Bacillus subtilis* involved in spore cortex formation. *J Bacteriol*. 1994; 176(7):2003–12. Epub 1994/04/01. PMID: [8144469](https://pubmed.ncbi.nlm.nih.gov/8144469/); PubMed Central PMCID: PMCPmc205306.
80. Resnekov O, Driks A, Losick R. Identification and characterization of sporulation gene spoVS from *Bacillus subtilis*. *J Bacteriol*. 1995; 177(19):5628–35. Epub 1995/10/01. PMID: [7559352](https://pubmed.ncbi.nlm.nih.gov/7559352/); PubMed Central PMCID: PMCPmc177374.
81. Fischer C, Geourjon C, Bourson C, Deutscher J. Cloning and characterization of the *Bacillus subtilis* prkA gene encoding a novel serine protein kinase. *Gene*. 1996; 168(1):55–60. Epub 1996/02/02. PMID: [8626065](https://pubmed.ncbi.nlm.nih.gov/8626065/).

82. Eichenberger P, Jensen ST, Conlon EM, van Ooij C, Silvaggi J, Gonzalez-Pastor JE, et al. The sigmaE regulon and the identification of additional sporulation genes in *Bacillus subtilis*. *J Mol Biol.* 2003; 327(5):945–72. Epub 2003/03/29. PMID: [12662922](#).
83. Yan J, Zou W, Fang J, Huang X, Gao F, He Z, et al. Eukaryote-like Ser/Thr Protein Kinase PrkA Modulates Sporulation via Regulating the Transcriptional Factor σK in *Bacillus subtilis*. *Frontiers in Microbiology.* 2015; 6. doi: [10.3389/fmicb.2015.00382](#) PMID: [25983726](#)
84. Gutierrez J, Smith R, Pogliano K. SpoIID-mediated peptidoglycan degradation is required throughout engulfment during *Bacillus subtilis* sporulation. *J Bacteriol.* 2010; 192(12):3174–86. Epub 2010/04/13. doi: [10.1128/jb.00127-10](#) PMID: [20382772](#); PubMed Central PMCID: PMCPmc2901704.
85. Wood TK, Knabel SJ, Kwan BW. Bacterial persister cell formation and dormancy. *Appl Environ Microbiol.* 2013; 79(23):7116–21. Epub 2013/09/17. doi: [10.1128/aem.02636-13](#) PMID: [24038684](#); PubMed Central PMCID: PMCPmc3837759.
86. Barth VC Jr, Rodrigues BA, Bonatto GD, Gallo SW, Pagnussatti VE, Ferreira CA, et al. Heterogeneous persister cells formation in *Acinetobacter baumannii*. *PLoS One.* 2013; 8(12):e84361. Epub 2014/01/07. doi: [10.1371/journal.pone.0084361](#) PMID: [24391945](#); PubMed Central PMCID: PMCPmc3877289.
87. Cava F, Hidalgo A, Berenguer J. *Thermus thermophilus* as biological model Extremophiles. 2009;13:213–31.
88. Ohtani N, Tomita M, Itaya M. An extreme thermophile, *Thermus thermophilus*, is a polyploid bacterium. *J Bacteriol.* 2010; 192(20):5499–505. Epub 2010/08/24. doi: [10.1128/jb.00662-10](#) PMID: [20729360](#); PubMed Central PMCID: PMCPmc2950507.
89. Munster MJ, Munster AP, Woodrow JR, Sharp RJ. Isolation and preliminary taxonomic studies of *Thermus* strains isolated from Yellowstone National Park, USA. *Journal of general microbiology.* 1986; 132(6):1677–83. Epub 1986/06/01. PMID: [3806053](#).
90. Costa KC, Navarro JB, Shock EL, Zhang CL, Soukup D, Hedlund BP. Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. *Extremophiles.* 2009; 13(3):447–59. Epub 2009/02/28. doi: [10.1007/s00792-009-0230-x](#) PMID: [19247786](#).
91. Ramirez-Arcos S, Fernandez-Herrero LA, Marin I, Berenguer J. Anaerobic growth, a property horizontally transferred by an Hfr-like mechanism among extreme thermophiles. *J Bacteriol.* 1998; 180(12):3137–43. PMID: [9620963](#); PubMed Central PMCID: PMC107814.
92. Bergquist PL, Morgan HW, Saul D. Selected enzymes from extreme thermophiles with applications in biotechnology. *Current Biotechnology.* 2014; 3:45–59.