ABSTRACT

FROCK, ANDREW DAVID. Functional Genomics Analysis of the Microbial Physiology and Ecology of Hyperthermophilic *Thermotoga* Species. (Under the direction of Dr. Robert Kelly.)

Generation of renewable fuels through biological conversion of inexpensive carbohydrate feedstocks has been the focus of intense research, which has mostly involved metabolic engineering of model mesophilic microorganisms. However, hyperthermophilic (T_{opt} ≥ 80°C) bacteria of the genus *Thermotoga* exhibit several desirable traits for a potential biofuel-producing host. They ferment a variety of carbohydrates to hydrogen with high yields approaching the theoretical limit for anaerobic metabolism of 4 mol H_2/mol glucose. High temperature growth facilitates biomass feedstock degradation, decreases risk of contamination, and eliminates the reactor cooling costs associated with growth of mesophiles. To further understand the glycoside hydrolase and carbohydrate transporter inventory of these bacteria, four closely-related *Thermotoga* species were compared. *T. maritima*, *T. neapolitana*, *T. petrophila*, and *T.* sp. strain RQ2 share approximately 75% of their genomes, yet each species exhibited distinguishing features during growth on simple and complex carbohydrates that correlated with their genotypes. *T. maritima* and *T. neapolitana* displayed similar monosaccharide utilization profiles, with a strong preference for glucose and xylose. Only *T.* sp. strain RQ2 has a fructose specific phosphotransferase system (PTS), and it was the only species found to utilize fructose. *T. petrophila* consumes glucose to a lesser extent and up-regulates a transporter normally used for xylose (XylE1F1K1) to compensate for the absence of the primary glucose transporter in these species (XylE2F2K2).

In some cases, the use of mixed cultures has been proposed to combine the complementary strengths of different microbes for the development of improved

bioprocesses. For this reason, bioprocessing would benefit from better insight into the factors that trigger stationary phase and affect cell viability in both pure and mixed cultures. Analysis of population dynamics of mixed cultures of these *Thermotoga* species showed that they do not form stable co-cultures under the conditions tested. This illustrated that establishment of stable, synthetic, mixed cultures requires an existing (or engineered) dependence of one species upon another. To cast light on signaling and survival mechanisms, especially the likely involvement of poorly characterized small ORFs (< 100 AA), *T. maritima* was cultivated in pure culture and nutrient levels were varied by changing the yeast extract concentration. The transition from exponential phase to stationary phase on richer medium resulted in differential transcription of 359 genes involved in cell division, carbon metabolism, and biosynthesis of flagella, amino acids, and vitamins. Growth on leaner medium triggered a comparable number of genes (254) with similar functional roles, but also a ~3-fold lower cell density, with many cells becoming bloated and coccoidal upon entry into stationary phase. An ORF corresponding to a putative 31-amino acid peptide (TM1316) was dramatically up-regulated in stationary phase on the lean medium, such that it was the most highly transcribed gene in the transcriptome. TM1316 is associated with a unique region of the genome characterized by repetitive organization of toxin/antitoxin pairs, small hypothetical ORFs, radical SAM proteins, and ABC-2 transporters. The only functionally characterized loci with similar organization were found to be responsible for bacteriocin production in *Bacillus* species. TM1316, consistent with transcriptomic data, was produced at high levels during stationary phase and localized to the cell envelope.

Taken together, this study highlights the fact that subtle genomic variations can have important consequences for industrially relevant phenotypes such as carbohydrate utilization. This insight supports the value of studies targeted at such variations and the

value of high-throughput screening of even closely related strains. Furthermore, the high-level transcription of TM1316 and associated phenotype contributes to the body of literature on physiology outside of central metabolic processes, the understanding of which is paramount for the development of robust bioprocesses.

Functional Genomics Analysis of the Microbial Physiology and Ecology of Hyperthermophilic *Thermotoga* Species

by
Andrew David Frock

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Chemical Engineering

Raleigh, North Carolina

2013

APPROVED BY:

_____	_____
David F. Ollis	Balaji M. Rao
_____	_____
Amy M. Grunden	Robert M. Kelly Chair of Advisory Committee

BIOGRAPHY

Andrew David Frock was born in Fletcher, NC on April 20, 1985. He grew up living on-site at Lutheridge Conference Center and Camp in Arden, NC. He graduated from T.C. Roberson High School in 2003. As a high school senior, he went on his first date with his future wife, Brittany Nicole Watson, on February 14, 2003. In 2007, he received a Bachelor of Science degree from Virginia Polytechnic Institute and State University (Blacksburg, VA) with a major in Biological Systems Engineering, a concentration in Bioprocess Engineering, and a minor in Chemistry. He then entered the graduate program in Chemical and Biomolecular Engineering at North Carolina State University. During this time, he happily married Brittany Watson Frock on October 3, 2009. Following completion of his doctoral degree in early 2013, he began work as a process engineer for the production of enzymes at Novozymes in Franklinton, NC.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF FIGURES

Chapter 1

The genus *Thermotoga*: Recent developments

Andrew D. Frock, Jaspreet S. Notey, and Robert M. Kelly

Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

Published in *Environmental Technology* 31(10): 1169-1181 (2010)

ABSTRACT

The genus Thermotoga comprises extremely thermophilic (Topt ≥ 70 °C) and hyperthermophilic (Topt ≥ 80 °C) bacteria, which have been extensively studied for insights into the basis for life at elevated temperatures and for biotechnological opportunities (e.g. biohydrogen production, biocatalysis). Over the past decade, genome sequences have become available for a number of Thermotoga species, leading to functional genomics efforts to understand growth physiology as well as genomics-based identification and characterization of novel high-temperature biocatalysts. Discussed here are recent developments along these lines for this group of microorganisms.

INTRODUCTION

The genus *Thermotoga* consists of some of the most thermophilic bacteria known, with optimum growth temperatures up to 80 °C [1–3]. Their ability to degrade a wide range of simple and complex carbohydrates, produce fermentative hydrogen at high yield, and catalyze a variety of high-temperature reactions has been the basis for numerous biotechnological applications [4]. These bacteria are also viewed as model systems for studying adaptation to high temperature and microbial evolution, as they present a challenge to conventional classification [5,6]. A number of genome sequences for Thermotoga species have become available in the past few years [7], offering additional insights into the biology of these interesting bacteria and suggesting biotechnological opportunities.

Members of the genus Thermotoga are anaerobic, rod-shaped bacteria encapsulated

by a unique ‘toga’-like outer membrane. Reported substrates for growth include hexoses, pentoses, disaccharides, glucans, xylans, glucomannan, galactomannan, pectin, chitin and amorphous cellulose [1,8]. This diversity of carbon sources correlates with the unusually large fraction of Thermotoga genes involved in carbohydrate degradation and utilization [9]. The primary products of fermentation are acetate, CO_2 and H_2, although lactate, ethanol, alanine and α-aminobutyrate have also been detected [1,10,11]. Thiosulfate, sulfur and Fe(III) are reduced by some *Thermotoga* species, although the microbes do not appear to gain energy from respiration in any of these cases [12].

Thermotoga species have been isolated from geothermally heated environments across the globe, including oil reservoirs, submarine hot springs and continental solfataric springs (Table 1.1). When the *Thermotoga maritima* genome sequence was completed, 24% of the genes were found to be most similar to archaea, suggesting that significant lateral gene transfer has occurred between these two groups [13]. However, when this issue was revisited recently with more genome sequence data available, only 7.7–11.0% of genes were found to be most similar to archaea, whereas 42.3– 48.2% of genes were most similar to Firmicutes [7]. The decrease in genes most similar to archaea is probably related to disproportionate expansion in the number of bacterial, compared with archaeal, genome sequences. Nevertheless, evidence for archaea–bacteria gene transfer remains, and several studies have provided evidence for frequent gene transfer within the Thermotoga genus [5,6,14]. These studies of genus-level diversity revealed that genes involved in essential processes and central metabolism are more highly conserved than those responsible for utilization of specific carbohydrates, especially polysaccharides.

DIVERSITY WITHIN THE *THERMOTOGA* GENUS

The formally identified *Thermotoga* species (Table 1.1) can essentially be split into two groups based on optimum growth temperature and the ability to reduce elemental sulfur. Four species have optimum growth temperatures of 77 °C and above (*T. maritima*, *T. petrophila*, *T. neapolitana* and *T. naphthophila*), whereas the other five species grow optimally at 70 °C and below (*T. elfii*, *T. thermarum*, *T. subterranea*, *T. hypogea* and *T. lettingae*) [15–17]. With one exception (*T. lettingae*), sulfur reduction is limited to the group of higher temperature species. Based on 16S rRNA, the higher temperature group is very closely related compared with the remaining species (Figure 1.1). Of the lower temperature group, *T. subterranea*, *T. elfii* and *T. lettingae* appear to be more related, whereas *T. hypogea* and *T. thermarum* do not cluster with any members of the genus.

Thermotoga lettingae was isolated from a sulphate-reducing bioreactor, where methanol was the only carbon source [11]. In addition to *T. lettingae*, it was reported that *T. subterranea*, *T. elfii*, *T. thermarum* and *T. maritima* are all able to use methanol [11]. Methanol is produced by pectin degradation, a known capability of *T. maritima* [18,19]. During growth of *T. maritima* on pectin, expression of a Zn-dependent alcohol dehydrogenase (TM0436) is upregulated, raising the possibility that this enzyme may catalyze the oxidation of methanol to formaldehyde [19]. Of the six *Thermotoga* species with complete genome sequences, the presence of this specific alcohol dehydrogenase correlates with the reported ability to degrade methanol in four instances. In the other two instances, *T. petrophila* has this gene, but was reported to not use methanol [10], whereas *T.* sp. RQ2 has this gene, but has not been tested for methanol use. *T. lettingae* also has a catalase/peroxidase enzyme (Tlet_1209), which could catalyze hydrogen-peroxide-

dependent oxidation of methanol, but this presumably would only occur in the presence of oxygen.

HYDROGEN PRODUCTION AND REDUCTANT DISPOSAL DURING CARBOHYDRATE FERMENTATION

Thermotoga species have been targeted for biohydrogen production because of reported yields approaching the Thauer limit for anaerobic fermentation [20–27]. This limit (4 mol H_2/mol glucose) can only be attained if all of the reducing equivalents from glucose oxidation are used to reduce protons to H_2 [28]. In practice, these reducing equivalents are also used for biosynthetic purposes and the formation of other fermentation products, including lactate, ethanol, butyrate and butanol. In particular, NADH produced by glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is usually not used for hydrogen production, as reduced ferredoxin and formate are stronger reducing agents [28,29]. Therefore, model fermenters like *Escherichia coli* and *Clostridia* species often yield 1–2 mol H_2/mol glucose, where the hydrogen formed is a result of pyruvate oxidation [29,30]. In *E. coli*, pyruvate is converted to acetyl-CoA, H_2 and CO_2 by the combined action of pyruvate formate lyase and formate hydrogen lyase. In *Clostridia* species, pyruvate ferredoxin oxidoreductase (PFOR) produces reduced ferredoxin (Fd), which is re-oxidized by hydrogenases to form H_2.

High hydrogen yields have been observed for the hyperthermophilic archaeon *Pyrococcus furiosus* as a result of a modified Embden–Meyerhof (EM) pathway [31]. In *P. furiosus*, glyceraldehyde 3-phosphate:ferredoxin oxidoreductase (GAPOR) converts glyceraldehyde 3-phosphate directly to 3-phosphoglycerate, replacing the typical EM

enzymes (GAPDH and phosphoglycerate kinase) [32]. This mechanism utilizes ferredoxin, not NADH, as electron carrier. Therefore, a yield of 4 mol H_2/mol glucose is attainable when all reduced ferredoxin from GAPOR and PFOR is used for hydrogen production via the membrane-bound hydrogenase [31,33,34]. It should also be noted that this modified EM pathway lacks the formation of ATP by phosphoglycerate kinase. Unlike *P. furiosus*, *T. maritima* has classical EM, pentose phosphate and Entner–Doudoroff pathways [32]. The high H_2 yields attained with *T. maritima* are related to the heterotrimeric [FeFe] hydrogenase, which has 73, 68 and 19 kDa subunits (α, β, and γ, respectively). Sequence analysis suggests that the β subunit is a flavoprotein that accepts electrons from NADH, and the γ subunit transfers electrons from the β subunit to the catalytic α subunit. Initial work was unable to demonstrate the use of ferredoxin or NADH as an electron donor for hydrogen production [35,36]. However, it has since been demonstrated that the *T. maritima* hydrogenase requires both reduced ferredoxin and NADH. This enzyme has been called a 'bifurcating' hydrogenase, and it has been proposed that energy from the oxidation of ferredoxin drives the unfavorable oxidation of NADH [37]. Thus, *T. maritima* has the ability to use the NADH from glycolysis for hydrogen production, allowing yields to approach the Thauer limit. Because this hydrogenase uses both NADH and reduced ferredoxin as electron donors, H_2 production may become unfavorable at lower H_2 partial pressures than if reduced ferredoxin alone was used [29]. Genome sequence analysis indicates that such 'bifurcating' hydrogenases may be present in several *Clostridia* species. The relatively high H_2 yields and low biomass yields observed for *Thermotoga* may be related to the apparently limited options for pyruvate metabolism, discussed below.

When the H_2 partial pressure builds up, electrons from glycolysis must be dissipated in other ways to sustain growth. Fermentation products detected during growth of *Thermotoga*

species include acetate, CO_2, H_2, ethanol, lactate and alanine (Figure 1.2). Acetate production involves the action of PFOR, phosphate acetyl-transferase (TM1130), and acetate kinase (TM0274). These reactions are motivated by the formation of ATP by acetate kinase, which allows cells to yield 4 ATP/mol glucose. Hydrogen yield is optimized when all glucose is converted to acetate, because the steps from pyruvate to acetate do not involve oxidation of NADH. When H_2 accumulates and the hydrogenase no longer oxidizes NADH, pyruvate would be diverted away from acetate production, possibly towards lactate production. Lactate is produced from pyruvate by lactate dehydrogenase (TM1867) with the concomitant reoxidation of NADH (Figure 1.2). Lactate levels reported during fermentation by *Thermotoga* species have varied from trace amounts up to levels rivalling that of acetate [1,21,23,24]. Although butyrate production has been reported for *T. neapolitana* and *T. lettingae*, the genes involved in this pathway are unclear [11,22,25]. Of the sequenced *Thermotoga* species, only *T. lettingae* has a putative acetyl-CoA acetyltransferase (Tlet_1557) and butyryl-CoA dehydrogenase (Tlet_1831) [38]. On first glance, members of the genus lack the pyruvate decarboxylase activity necessary for ethanol formation. However, this reaction has been catalysed in vitro by the PFOR from *P. furiosus* [39], suggesting that the *T. maritima* homolog (TM0015-18) may have the same ability. Traces of ethanol have only been reported for *T. lettingae* and *T. hypogea* [11,40], but the gene responsible for NADPH-dependent ethanol dehydrogenase activity in *T. hypogea* appears to be present in all sequenced strains except for *T. lettingae* [41]. Low levels of alanine have been reported for many *Thermotoga* species, although this metabolite appears to be more significant for *T. elfii*, *T. lettingae*, *T. hypogea* and other members of the order Thermotogales [11,40,42]. Alanine production by *P. furiosus* has been attributed to alanine aminotransferase and glutamate dehydrogenase [43,44]. Assuming alanine is produced by

T. maritima in a similar fashion, possible genes are given in Figure 1.2. If it is not the presence or absence of these various pathways, perhaps it is their regulation which results in high hydrogen yields and distinguishes *Thermotoga* species from other fermentative anaerobes.

RESPONSE TO OXYGEN EXPOSURE

Although hydrogenases are inactivated by oxygen, increased hydrogen production in the presence of oxygen has been reported for *T. neapolitana*, with yields of 8.5 ± 2.9 mol H_2/mol glucose [20,22]. These high yields and the depletion of headspace oxygen suggested that *T. neapolitana* utilizes oxygen for a more energy-efficient catabolic process. Extra-cytoplasmic oxygen-resistant hydrogenase activity has been detected in *T. neapolitana* [45], but the physiological role is unknown and H_2-evolving hydrogenases are generally cytoplasmic [46]. Structural modelling of the [FeFe] hydrogenase α subunits suggests that the catalytic site of *T. neapolitana*'s enzyme may be less accessible to oxygen than *T. maritima*'s enzyme, accounting for increased resistance to inactivation [47]. However, subsequent studies on *T. neapolitana* have provided evidence that growth and hydrogen production are inhibited by oxygen [23,24]. In the initial experiments [20,22], the hydrogen yield may have been overestimated if any carbon sources in other media components (2 g/L yeast extract and 2 g/L trypticase) were utilized in addition to glucose. Also, Eriksen et al. [23] demonstrated that oxygen is depleted from the headspace of cell-free medium and that a lack of agitation may allow anaerobic conditions to be established in the bottom of the culture where the cells settle. Thus, the level of oxygen actually encountered and consumed by *T. neapolitana* may be much less than it initially appears,

especially in unshaken cultures.

Anaerobic microbes have been found to consume small amounts of oxygen, although they do so to protect themselves from its deleterious effects. NADH oxidases that can convert O_2 to H_2O_2 have been detected in *T. hypogea*, *T. neapolitana* and *T. maritima* [48,49]. Such enzymes are useless for oxygen detoxification unless they are accompanied by a peroxidase to convert H_2O_2 into H_2O. This peroxidase activity is presumed to be the work of rubrerythrin (TM0657), which is upregulated in *T. maritima* in response to oxidative stress [50] and has been shown to function as a NADH peroxidase in *P. furiosus* [51]. In the presence of oxygen, microbes also have to cope with reactive oxygen species like superoxide, which results from the reduction of oxygen. In *P. furiosus*, a system for superoxide reduction has been identified which involves transfer of electrons from NADPH with the involvement of NADPH:rubredoxin oxidoreductase, rubredoxin and superoxide reductase [52,53]. Homologs of these proteins have been identified in *T. maritima* (TM0754, TM0659 and TM0658, respectively), and the putative superoxide reductase is upregulated in response to oxidative stress [50]. Other genes upregulated by oxygen exposure include TM1368, which encodes a protein involved in iron–sulfur cluster assembly and repair, as well as TM0755. The recombinant protein encoded by TM0755 was characterized as a rubredoxin oxygen oxidoreductase. This enzyme is capable of converting O_2 to H_2O without the formation of H_2O_2, a favourable property for efficient oxygen detoxification.

MICROBIAL ECOLOGY OF *THERMOTOGA* SPECIES

Growth in mixed cell communities is the predominant mode in natural microbial environments [54]. Functional genomics and proteomics studies on microorganisms, to date,

have largely focused on metabolism and physiology based in the cytosol. Evidence from molecular microbial ecology studies over the past 15 years has established that bacteria utilize a chemical language based on a variety of small signalling molecules [55,56]. It is also clear that microbial communities are not merely a collection of unicellular and isolated indi- viduals, but rather a cooperative society capable of acting on a multi-cellular level [57]. A wide spectrum of molecules are known to be released into the surroundings by the bacteria, and their ability to sense the concentration of these molecules through receptors in the cell membrane forms the basis of this chemical language. Quorum sensing is defined as the ability of a single cell to sense the number of bacteria in its proximity based upon the accumulation of signalling molecules [58]. Although not extensively studied, *Thermotoga* species also appear to participate in community behavior, with some differences from less thermophilic bacteria [59,60].

Functional genomics approaches have been used to understand associated community behavior responses in *T. maritima*, including cell-density-dependent processes, such as biofilm formation during growth in mono- and co-culture [61]. Previous studies revealed significant wall growth when *T. maritima* cells were grown alone in continuous culture [62], suggesting the formation of exopolysaccharide (EPS)-associated sessile cell communities. Mixed culture experiments have also shed light on syntrophic interactions between hydrogen-producing *T. maritima* and the methanogen *Methanocaldococcus jannaschii* [63]. Transcriptomic tools, in association with statistical bioinformatics analyses, have given new insights into the ecological interactions of *Thermotoga* species under experimental conditions mimicking those that might arise in natural habitats.

Transcriptional differences were measured between sessile and planktonic *T. maritima* cells from a continuous culture reactor. The sessile state of cells in the biofilms was found to

be associated within rope-like structures, which formed on the reactor walls, polycarbonate filters and nylon mesh [59] (Figure 1.3). The volcano plot in Figure 1.3 underscores the differential transcriptional response of planktonic and sessile *T. maritima* cells obtained from cDNA microarray analysis, in which 114 genes were found to be differentially expressed two-fold or higher at this significance level ($-\log_{10}$ P value ≥4.5). Transcriptional analysis of *T. maritima* biofilms agrees with studies of mesophiles, in that heat shock genes (e.g. dnaK, smHSP) were up-regulated in biofilm cells, whereas transcripts of a cold shock gene [64] were downregulated [65–67].

Pure culture studies can provide information about the growth physiology; however, mixed culture studies can offer further understanding of intricate mechanisms relevant to growth of cells in communities, conditions more akin to the natural state of microorganisms. *T. maritima* grew to extremely high densities in a co-culture with the methanogenic archaeon *M. jannaschii*, which uses the growth-inhibiting hydrogen produced by *T. maritima* to produce methane [60] (Figure 1.4a). *T. maritima* grew to a maximum cell density of over 10^9 cells/mL in the co-culture, compared with its typical pure culture densities of 2–3 × 10^8 cells/mL. Full genome transcriptional comparisons between *T. maritima* cells in pure culture and in high density co-culture with *M. jannaschii* showed changes in sugar utilization and transport genes, including glycosyl transferases and genes encoding known glucomannan (TM1752) and α-glucan (TM1834) hydrolases, which correlated with the appearance of EPS in the culture [60].

A possible involvement of cyclic-di-GMP in regulating EPS formation and maintenance was suggested by the transcriptional patterns observed. Cyclic-di-GMP, a second messenger, is known to regulate cellulose synthesis in *Acetobacter xylinium* [68] and *S. enterica* serovar Typhimurium [69], as well as biofilm formation in *V. cholerae* [70],

Staphylococcus aureus [71], *Yersinia pestis* [72] and *Pseudomonas aeruginosa* [73], among others. Several proteins containing GGDEF domains displayed different expression patterns between the pure culture and co-culture, including a characterized diguanylate cyclase (TM1163) [74] and a putative cyclic-di-GMP phosphodiesterase (TM1184) [60]. Expression data indicated a slight upregulation for TM1163 (+1.7-fold) and a downregulation for TM1184 (−2.6-fold) in the high density co-culture compared with the pure culture. Even though *M. jannaschii* is present in the co-culture, *T. maritima* is most likely responsible for EPS production, because *T. maritima* dominates the culture (95% of the cells are *T. maritima*) and has more carbohydrate active enzymes [75]. Transcriptional response data identified a small, unknown open reading frame (ORF) (TM0504) [60]. This gene's co-localization with an ABC transporter led to the proposed secretion mechanism depicted in Figure 1.4b, c. The permeases (TM0503, TM0502) and ATP-binding subunits (TM0501, TM0500) (Figure 1.4c) are located immediately upstream of TM0504, and anti-microbial peptide transporter domains exist in both TM0500 and TM0501 (COG4167 and 4170) [60]. The absence of a periplasmic-binding protein suggests that this cluster of genes may export a small peptide derived from the protein product of TM0504. In addition, TM0504 contained a GG motif (Figure 1.4b), similar to the cleavage point for the active form of auto-inducing peptides found in *S. pneumonia* and lactic acid bacteria [76]. Due to the lack of a genetic system for *T. maritima*, TM0504 knockout mutants could not be constructed and so, to understand the function of this polypeptide, a truncated version of the TM0504 peptide was dosed into pure low-density *T. maritima* cultures. Formation of EPS was triggered in the pure cultures dosed with the synthetic peptide, whereas no EPS was formed in the undosed control cultures (Figure 1.4d) [60]. This report was the first indication of the importance of peptide-based quorum sensing in hyperthermophilic habitats.

Further analysis of the region around TM0504 revealed that it was co-located on the opposite strand with the gene encoding ssrA, a hybrid of tRNA and mRNA (tmRNA). tmRNA is widely prevailing in bacteria and is involved in a trans-translational process related to ribosome rescue [77]. Specific DNA probes were designed and used in real-time PCR assays to follow separate transcriptional responses of the co- located ORFs during transition from exponential into stationary phase, chloramphenicol challenge and syntrophic co-culture with *M. jannaschii* [77]. No significant change in either TM0504 or tmRNA transcription was observed under normal growth conditions in both pure cultures. However, an eight-fold increase in transcription of the tmRNA gene and a 30-fold decrease in TM0504 transcription levels was noted for high density co-culture with *M. jannaschii*. The down-regulation of the TM0504 gene was thought to be a result of the cells no longer participating in quorum sensing behaviour as a result of aggregation with *M. jannaschii*. Chloramphenicol challenge resulted in a 40-fold increase after five minutes and a 23-fold increase after 30 minutes in tmRNA gene transcription. A two-fold increase was noted for TM0504 transcription. The effects were attributed to disruptions of translational processes caused by the addition of antibiotic. The possibility that a biologically active peptide was encoded and independently transcribed from the strand opposite tmRNA in most bacterial genomes was found to be intriguing. Although the motivation for co-located, overlapping genes is not clear, it has been suggested that they may compress genome size in species subject to reductive evolution [78].

The genome of *T. maritima* was hypothesized to have undergone extensive gene transfers [13] and this was further investigated by comparing nine *Thermotoga* species using comparative genomic hybridization [5]. To date, the genome of *T. maritima* has been found to contain eight distinct clustered regularly interspaced short palindromic repeats

(CRISPR) regions [79]. CRISPRs represent a family of DNA repeats that have been found in most archaeal (90%) and some bacterial genomes (40%) [79,80]. A variety of putative roles were proposed for these sequences including chromosomal rearrangement, modulation of expression of neighboring genes, target for DNA binding, and DNA repair [81]. More recent studies have hypothesized that CRISPR sequences provide adaptive immunity against foreign genetic elements [82]. A whole-genome alignment of *T. maritima* and *T. neapolitana* revealed numerous large scale DNA rearrangements, most of which are associated with CRISPR sequences and/or tRNA genes [83]. This study proposed several inversion/translocation sequences consistent with the genome arrangements observed, and suggested that *Thermotoga* species favor inversion/translocation events within a replichore. Although no specific role or biological significance for such rearrangements was proven, the authors discussed the possibility of these inversions being involved in some form of gene regulation, conferring metabolic advantage through physiologically important pathways.

BIOCATALYSIS USING THERMOTOGA ENZYMES

T. maritima's xylanolytic abilities have attracted interest for food, paper and biofuel-related applications. Xylo-oligosaccharides resulting from xylan degradation have been shown to increase the number of beneficial gut bacteria, such as *Bifidobacterium* species [84]. Production of xylo-oligosaccharides, especially xylobiose, has been demonstrated using immobilized xylanase B (XynB) from *T. maritima* (TM0070), as well as a combination of arabinofuranosidase (TM0281) and XynB [85,86]. The combination of both enzymes resulted in twice as much xylan degradation than with XynB alone. The rationale for this is that arabinofuranosidase cleaves modified groups of the xylan backbone, allowing XynB

more access [86]. Xylanase B has also been used as an additive for bread-making, as it improves oven spring, volume, shape and texture. In comparison with lower temperature xylanases, the thermostable version from *T. maritima* performs better because of extended activity during the baking process. Xylanases may improve bread quality by solubilizing arabinoxylans and changing the water distribution in the dough [87]. Xylanases can also be useful to the pulp and paper industry, as they can reduce the amount of chlorine necessary for pulp bleaching. Xylanase B from *T. maritima* was shown to be effective under conditions appropriate for this application, namely high temperature and alkaline pH [88].

T. maritima XynB is a 40 kDa single domain protein that is located in the periplasm. *T. maritima* also has a larger 120 kDa multi-domain protein, xylanase A (TM0061), which is located primarily in the outer membrane [89]. Xylanase A (XynA) has two carbohydrate-binding modules (CBMs) on each side of the catalytic domain. One of the N-terminal domains binds xylan, while one of the C-terminal domains binds cellulose [90,91]. The thermostability of XynA decreases when the N-terminal modules are removed. The thermostability is not affected when the C-terminal modules are removed, suggesting that the N-terminal modules are more important for the thermostability of the enzyme [91]. A fusion protein combining one of these N-terminal domains with the xylanase 2 from *Trichoderma reesei* exhibited greater thermostability and substrate binding ability than the wild-type xylanase 2, providing further evidence that these N-terminal domains are involved in thermostability [92]. Xylanase A hydrolyses 90–95% of xylohexaose (X6) to xylobiose (X2)/xylotetraose (X4) and 5–10% to xylotriose (X3). When the C-terminal domains of XynA are removed, X2/X4 production decreases significantly, whereas X3 production appears relatively unaffected. In this case, about 50% of X6 is hydrolysed to X2/X4 while the other 50% is hydrolysed to X3. This variant also has about 60% higher activity on insoluble wheat

arabinoxylan compared with the wild-type enzyme. These modulations in substrate specificity and hydrolysis products may be useful for optimization of a biomass conversion process [93].

The thermostability of carbohydrate-active enzymes from *T. maritima* makes them attractive for transgenic expression in plants. Heterologous expression in plants would be useful for large-scale production of enzymes for biomass conversion [94], as well as for development of self-processing biomass. In either case, yields can suffer if the transgenic protein stunts plant growth. This can be avoided if the expressed protein is inactive at ambient temperatures, which is true for hyperthermophilic proteins. The endoglucanase Cel5A (TM1751) from *T. maritima* has been expressed in the chloroplasts of transgenic tobacco at levels up to 5.2% of total soluble protein. The transgenic plants exhibited normal growth and development, which may be attributed to a combination of the thermophilic nature of the protein and the chloroplast localization [95]. In planta protein production could also benefit industrial starch liquefaction, where thermostable α-amylase is used for initial solubilization. Calcium is added to these processes, as it is a requisite cofactor for α-amylase function. The α-amylase (TM1840) from *T. maritima* has been expressed in tobacco cell cultures with a significant increase in thermostability compared with the version produced by *E. coli*, as a result of the intrinsic calcium levels in the tobacco cells [96]. Below 40°C, no activity was detected for this enzyme. Thus, self-processing plants can be envisioned. Such plants could grow normally while expressing thermophilic enzymes. After harvest, the plants would be heat-treated to activate the enzymes, which would initiate degradation of the plant polysaccharides from within.

CONCLUSIONS

The genus *Thermotoga* will likely continue to be of interest to scientists and technologists because of the intriguing insights that can be obtained into life at extremely high temperatures, the fundamentals of biomolecular function in thermal environments, and biotechnological applications, especially related to bioenergy. The recent construction of a metabolic model for *T. maritima*, which incorporates 478 genes, represents a significant step towards a more systems-level understanding of this organism [97]. Pending the availability of a directed genetic system for these bacteria, such metabolic models could be combined with gene knockouts to test specific hypotheses related to physiological and ecological features. In addition, the structural genomics efforts with T. maritima have provided an important set of three-dimensional crystal structures of many proteins produced by this bacterium [98]. This information will be valuable in providing functional annotations for many proteins now listed as ‘hypothetical’ and for relating biophysical data to phys- iological roles. Ultimately, Thermotoga species could be metabolically engineered to serve as recombinant hosts for synthetic biology applications that make strategic use of elevated temperatures, especially those that seek high microbial hydrogen-production rates with yields at the theoretical maximum.

ACKNOWLEDGEMENTS

This work was supported in part by grants from the US National Science Foundation (CBET-0730091) and the US Department of Energy (DG-FG02-08ER64687). A.D. Frock acknowledges support from a US NIH T32 Biotechnology Traineeship (T32 GM008776).

REFERENCES

1. Huber R., Langworthy T.A., Köning H., Thomm M., Woese C.R., Sleytr U.B. and Stetter K.O., *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90° C. *Arch Microbiol* **144**, 324-333 (1986).

2. Belkin S., Wirsen C.O. and Jannasch H.W., A New Sulfur-Reducing, Extremely Thermophilic Eubacterium from a Submarine Thermal Vent. *Appl Environ Microbiol*, **51**, 1180-1185 (1986).

3. Jannasch H.W., Huber R.B., Belkin S. and Stetter K.O., *Thermotoga neapolitana* sp. nov. of the extremely thermophilic, eubacterial genus *Thermotoga*. *Arch Microbiol* **150**, 103-104 (1988).

4. Conners S.B., Mongodin E.F., Johnson M.R., Montero C.I., Nelson K.E. and Kelly R.M., Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. *FEMS Microbiol Rev*, **30**, 872-905 (2006).

5. Mongodin E.F., Hance I.R., Deboy R.T., Gill S.R., Daugherty S., Huber R., Fraser C.M., Stetter K. and Nelson K.E., Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. *J Bacteriol*, **187**, 4935-4944 (2005).

6. Nesbo C.L., Dlutek M. and Doolittle W.F., Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics*, **172**, 759-769 (2006).

7. Zhaxybayeva O., Swithers K., Lapierre P., Fournier G., Bickhart D., Deboy R., Nelson K., Nesbo C., Doolittle W., Gogarten J. and Noll K., On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc Natl Acad Sci USA*, **106**, 5865-5870 (2009).

8. Conners S.B., Montero C.I., Comfort D.A., Shockley K.R., Johnson M.R., Chhabra S.R. and Kelly R.M., An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. *J Bacteriol*, **187**, 7267-7282 (2005).

9. Chhabra S.R., Shockley K.R., Ward D.E. and Kelly R.M., Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown

on glucan- and mannan-based polysaccharides. *Appl Environ Microbiol*, **68**, 545-554 (2002).

10. Takahata Y., Nishijima M., Hoaki T. and Maruyama T., *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. *Int J Syst Evol Micr*, **51**, 1901-1909 (2001).

11. Balk M., Weijma J. and Stams A.J., *Thermotoga lettingae* sp. nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic anaerobic reactor. *Int J Syst Evol Micr* **52**, 1361-1368 (2002).

12. Vargas M., Kashefi K., Blunt-Harris E.L. and Lovley D.R., Microbiological evidence for Fe(III) reduction on early Earth. *Nature*, **395**, 65-67 (1998).

13. Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A., McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M., Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D., Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O., Salzberg S.L., Smith H.O., Venter J.C. and Fraser C.M., Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323-329 (1999).

14. Nesbo C.L., Nelson K.E. and Doolittle W.F., Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*. *J Bacteriol*, **184**, 4475-4488 (2002).

15. Ravot G., Magot M., Fardeau M.L., Patel B.K., Prensier G., Egan A., Garcia J.L. and Ollivier B., *Thermotoga elfii* sp. nov., a novel thermophilic bacterium from an African oil-producing well. *Int J Syst Bacteriol*, **45**, 308-314 (1995).

16. Jeanthon C., Reysenbach A.L., L'Haridon S., Gambacorta A., Pace N.R., Glenat P. and Prieur D., *Thermotoga subterranea* sp. nov., a new thermophilic bacterium isolated from a continental oil reservoir. *Arch Microbiol*, **164**, 91-97 (1995).

17. Windberger E., Huber R., Trincone A., Fricke H. and Stetter K.O., *Thermotoga thermarum* sp. nov. and *Thermotoga neapolitana* occuring in African continental solfataric springs. *Arch Microbiol* **151**, 506-512 (1989).

18. Schink B. and Zeikus J.G., Microbial methanol formation - A major end product of pectin metabolism. *Curr Microbiol*, **4**, 387-389 (1980).

19. Kluskens L.D., Molecular studies on protein- and carbohydrate-converting enzymes from thermophilic bacteria. Wageningen Universiteit, Wageningen, Netherlands p. 138 (2004).

20. Van Ooteghem S.A., Jones A., Van Der Lelie D., Dong B. and Mahajan D., H_2 production and carbon utilization by *Thermotoga neapolitana* under anaerobic and microaerobic growth conditions. *Biotechnol Lett*, **26**, 1223-1232 (2004).

21. Schroder C., Selig M. and Schonheit P., Glucose fermentation to acetate, CO2 and H2 in the anaerobic hyperthermophilic eubacterium *Thermotoga maritima* - involvement of the Embden-Meyerhof pathway. *Arch Micobiol*, **161**, 460-470 (1994).

22. Van Ooteghem S.A., Beer S.K. and Yue P.C., Hydrogen production by the thermophilic bacterium *Thermotoga neapolitana*. *Appl Biochem Biotechnol*, **98-100**, 177-189 (2002).

23. Eriksen N.T., Nielsen T.M. and Iversen N., Hydrogen production in anaerobic and microaerobic *Thermotoga neapolitana*. *Biotechnol Lett*, **30**, 103-109 (2008).

24. Munro S.A., Zinder S.H. and Walker L.P., The fermentation stoichiometry of *Thermotoga neapolitana* and influence of temperature, oxygen, and pH on hydrogen production. *Biotechnol Prog*, **25**, 1035-1042 (2009).

25. Nguyen T.A., Han S.J., Kim J.P., Kim M.S. and Sim S.J., Hydrogen production of the hyperthermophilic eubacterium, *Thermotoga neapolitana* under N2 sparging condition. *Bioresour Technol*, **101 Suppl 1**, S38-41 (2009).

26. de Vrije T., Bakker R.R., Budde M.A., Lai M.H., Mars A.E. and Claassen P.A., Efficient hydrogen production from the lignocellulosic energy crop Miscanthus by the extreme thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. *Biotechnol Biofuels*, **2**, 12 (2009).

27. van Niel E., Budde M., de Haas G., van der Wal F., Claasen P. and Stams A., Distinctive properties of high hydrogen producing extreme thermophiles, *Caldicellulosiruptor saccharolyticus* and *Thermotoga elfii*. *Int J Hydrogen Energ*, **27**, 1391-1398 (2002).

28. Thauer R.K., Jungermann K. and Decker K., Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol Rev*, **41**, 100-180 (1977).

29. Angenent L.T., Karim K., Al-Dahhan M.H., Wrenn B.A. and Domiguez-Espinosa R., Production of bioenergy and biochemicals from industrial and agricultural wastewater. *Trends Biotechnol*, **22**, 477-485 (2004).

30. Chin H., Chen Z. and Chou C., Fedbatch operation using *Clostridium acetobutylicum* suspension culture as biocatalyst for enhancing hydrogen production. *Biotechnol Progr*, **19**, 383-388 (2003).

31. Chou C.J., Shockley K.R., Conners S.B., Lewis D.L., Comfort D.A., Adams M.W. and Kelly R.M., Impact of substrate glycoside linkage and elemental sulfur on bioenergetics of and hydrogen production by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Appl Environ Microbiol*, **73**, 6842-6853 (2007).

32. Selig M., Xavier K.B., Santos H. and Schonheit P., Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. *Arch Microbiol*, **167**, 217-232 (1997).

33. Sapra R., Verhagen M.F. and Adams M.W., Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J Bacteriol*, **182**, 3423-3428 (2000).

34. Sapra R., Bagramyan K. and Adams M.W., A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc Natl Acad Sci USA*, **100**, 7545-7550 (2003).

35. Chou C.J., Jenney F.E., Jr., Adams M.W. and Kelly R.M., Hydrogenesis in hyperthermophilic microorganisms: implications for biofuels. *Metab Eng*, **10**, 394-404 (2008).

36. Verhagen M.F., O'Rourke T. and Adams M.W., The hyperthermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization. *Biochim Biophys Acta*, **1412**, 212-229 (1999).

37. Schut G.J. and Adams M.W., The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol*, **191**, 4451-4457 (2009).

38. Kanehisa M. and Goto S., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, **28**, 27-30 (2000).

39. Ma K., Hutchins A., Sung S.J. and Adams M.W., Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. *Proc Natl Acad Sci USA*, **94**, 9608-9613 (1997).

40. Fardeau M.L., Ollivier B., Patel B.K., Magot M., Thomas P., Rimbault A., Rocchiccioli F. and Garcia J.L., *Thermotoga hypogea* sp. nov., a xylanolytic, thermophilic bacterium from an oil-producing well. *Int J Syst Bacteriol*, **47**, 1013-1019 (1997).

41. Ying X., Wang Y., Badiei H.R., Karanassios V. and Ma K., Purification and characterization of an iron-containing alcohol dehydrogenase in extremely thermophilic bacterium *Thermotoga hypogea*. *Arch Microbiol*, **187**, 499-510 (2007).

42. Ravot G., Ollivier B., Fardeau M.L., Patel B.K.C., Andrews K.T., Magot M. and Garcia J.L., L-Alanine production from glucose fermentation by hyperthermophilic members of the domains Bacteria and Archaea: A remnant of an ancestral metabolism? *Appl Environ Microbiol*, **62**, 2657-2659 (1996).

43. Kengen S.W.M. and Stams A.J.M., Formation of L-Alanine as a Reduced End-Product in Carbohydrate Fermentation by the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *Arch Microbiol*, **161**, 168-175 (1994).

44. Ward D.E., Kengen S.W., van Der Oost J. and de Vos W.M., Purification and characterization of the alanine aminotransferase from the hyperthermophilic Archaeon *Pyrococcus furiosus* and its role in alanine production. *J Bacteriol*, **182**, 2559-2566 (2000).

45. Kaslin S.A., Childers S.E. and Noll K.M., Membrane-associated redox activities in *Thermotoga neapolitana*. *Arch Microbiol*, **170**, 297-303 (1998).

46. Jenney F.E., Jr. and Adams M.W., Hydrogenases of the model hyperthermophiles. *Ann N Y Acad Sci*, **1125**, 252-266 (2008).

47. Tosatto S., Toppo S., Carbonera D., Giacometti G. and Costantini P., Comparative analysis of [FeFe] hydrogenase from Thermotogales indicates the molecular basis of resistance to oxygen inactivation. *Int J Hydrogen Energ*, **33**, 570-578 (2008).

48. Yang X. and Ma K., Purification and characterization of an NADH oxidase from extremely thermophilic anaerobic bacterium *Thermotoga hypogea*. *Arch Microbiol*, **183**, 331-337 (2005).

49. Yang X. and Ma K., Characterization of an exceedingly active NADH oxidase from the anaerobic hyperthermophilic bacterium *Thermotoga maritima*. *J Bacteriol*, **189**, 3312-3317 (2007).

50. Le Fourn C., Fardeau M.L., Ollivier B., Lojou E. and Dolla A., The hyperthermophilic anaerobe *Thermotoga maritima* is able to cope with limited amount of oxygen: insights into its defence strategies. *Environ Microbiol*, **10**, 1877-1887 (2008).

51. Weinberg M.V., Jenney F.E., Jr., Cui X. and Adams M.W., Rubrerythrin from the hyperthermophilic archaeon *Pyrococcus furiosus* is a rubredoxin-dependent, iron-containing peroxidase. *J Bacteriol*, **186**, 7888-7895 (2004).

52. Jenney F.E., Jr., Verhagen M.F., Cui X. and Adams M.W., Anaerobic microbes: oxygen detoxification without superoxide dismutase. *Science*, **286**, 306-309 (1999).

53. Grunden A.M., Jenney F.E., Jr., Ma K., Ji M., Weinberg M.V. and Adams M.W., In vitro reconstitution of an NADPH-dependent superoxide reduction pathway from *Pyrococcus furiosus*. *Appl Environ Microbiol*, **71**, 1522-1530 (2005).

54. Costerton J.W., Lewandowski Z., Caldwell D.E., Korber D.R. and Lappin-Scott H.M., Microbial biofilms. *Annu Rev Microbiol*, **49**, 711-745 (1995).

55. Antunes L.C. and Ferreira R.B., Intercellular communication in bacteria. *Crit Rev Microbiol*, **35**, 69-80 (2009).

56. Gonzalez J.E. and Keshavan N.D., Messing with bacterial quorum sensing. *Microbiol Mol Biol Rev*, **70**, 859-875 (2006).

57. Bassler B.L., Small talk: Cell-to-cell communication in bacteria. *Cell*, **109**, 421-424 (2002).

58. Swift S., Throup J.P., Williams P., Salmond G.P. and Stewart G.S., Quorum sensing: a population-density component in the determination of bacterial phenotype. *Trends Biochem Sci*, **21**, 214-219 (1996).

59. Pysz M.A., Conners S.B., Montero C.I., Shockley K.R., Johnson M.R., Ward D.E. and Kelly R.M., Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*. *Appl Environ Microbiol*, **70**, 6098-6112 (2004).

60. Johnson M.R., Montero C.I., Conners S.B., Shockley K.R., Bridger S.L. and Kelly R.M., Population density-dependent regulation of exopolysaccharide formation in the hyperthermophilic bacterium *Thermotoga maritima*. *Mol Microbiol*, **55**, 664-674 (2005).

61. Johnson M.R., Montero C.I., Conners S.B., Shockley K.R., Pysz M.A. and Kelly R.M., Functional genomics-based studies of the microbial ecology of hyperthermophilic micro-organisms. *Biochem Soc Trans*, **32**, 188-192 (2004).

62. Rinker K., Han C. and Kelly R., Continuous culture as a tool for investigating the growth physiology of heterotrophic hyperthermophiles and extreme thermoacidophiles. *J Appl Microbiol*, **85**, 118S-127S (1999).

63. Muralidharan V., Rinker K.D., Hirsh I.S., Bouwer E.J. and Kelly R.M., Hydrogen transfer between methanogens and fermentative heterotrophs in hyperthermophilic cocultures. *Biotechnol Bioeng*, **56**, 268-278 (1997).

64. Phadtare S., Hwang J., Severinov K. and Inouye M., CspB and CspL, thermostable cold-shock proteins from *Thermotoga maritima*. *Genes Cells*, **8**, 801-810 (2003).

65. Whiteley M., Bangera M.G., Bumgarner R.E., Parsek M.R., Teitzel G.M., Lory S. and Greenberg E.P., Gene expression in *Pseudomonas aeruginosa* biofilms. *Nature*, **413**, 860-864 (2001).

66. Svensater G., Welin J., Wilkins J.C., Beighton D. and Hamilton I.R., Protein expression by planktonic and biofilm cells of *Streptococcus mutans*. *FEMS Microbiol Lett*, **205**, 139-146 (2001).

67. Schembri M.A., Kjaergaard K. and Klemm P., Global gene expression in *Escherichia coli* biofilms. *Mol Microbiol*, **48**, 253-267 (2003).

68. Chang A.L., Tuckerman J.R., Gonzalez G., Mayer R., Weinhouse H., Volman G., Amikam D., Benziman M. and Gilles-Gonzalez M.A., Phosphodiesterase A1, a regulator of cellulose synthesis in *Acetobacter xylinum*, is a heme-based sensor. *Biochemistry*, **40**, 3420-3426 (2001).

69. Garcia B., Latasa C., Solano C., Garcia-del Portillo F., Gamazo C. and Lasa I., Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. *Mol Microbiol*, **54**, 264-277 (2004).

70. Tischler A.D. and Camilli A., Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. *Mol Microbiol*, **53**, 857-869 (2004).

71. Karaolis D.K., Rashid M.H., Chythanya R., Luo W., Hyodo M. and Hayakawa Y., c-di-GMP (3'-5'-cyclic diguanylic acid) inhibits *Staphylococcus aureus* cell-cell interactions and biofilm formation. *Antimicrob Agents Ch*, **49**, 1029-1038 (2005).

72. Kirillina O., Fetherston J.D., Bobrov A.G., Abney J. and Perry R.D., HmsP, a putative phosphodiesterase, and HmsT, a putative diguanylate cyclase, control Hms-dependent biofilm formation in *Yersinia pestis*. *Mol Microbiol*, **54**, 75-88 (2004).

73. Hickman J.W., Tifrea D.F. and Harwood C.S., A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proc Natl Acad Sci USA*, **102**, 14422-14427 (2005).

74. Ryjenkov D.A., Tarutina M., Moskvin O.V. and Gomelsky M., Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J Bacteriol*, **187**, 1792-1798 (2005).

75. Chhabra S.R., Shockley K.R., Conners S.B., Scott K.L., Wolfinger R.D. and Kelly R.M., Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J Biol Chem*, **278**, 7540-7552 (2003).

76. Havarstein L.S., Diep D.B. and Nes I.F., A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export. *Mol Microbiol*, **16**, 229-240 (1995).

77. Montero C.I., Lewis D.L., Johnson M.R., Conners S.B., Nance E.A., Nichols J.D. and Kelly R.M., Colocation of genes encoding a tRNA-mRNA hybrid and a putative signaling peptide on complementary strands in the genome of the hyperthermophilic bacterium *Thermotoga maritima*. *J Bacteriol*, **188**, 6802-6807 (2006).

78. Sakharkar K.R. and Chow V.T., Strategies for genome reduction in microbial genomes. *Genome Inform*, **16**, 69-75 (2005).

79. Jansen R., Embden J.D., Gaastra W. and Schouls L.M., Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*, **43**, 1565-1575 (2002).

80. van der Oost J., Jore M.M., Westra E.R., Lundgren M. and Brouns S.J., CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci*, **34**, 401-407 (2009).

81. Makarova K.S., Aravind L., Grishin N.V., Rogozin I.B. and Koonin E.V., A DNA repair system specific for thermophilic Archaea and Bacteria predicted by genomic context analysis. *Nucleic Acids Res*, **30**, 482-496 (2002).

82. Makarova K.S., Grishin N.V., Shabalina S.A., Wolf Y.I. and Koonin E.V., A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*, **1**, 7 (2006).

83. DeBoy R.T., Mongodin E.F., Emerson J.B. and Nelson K.E., Chromosome evolution in the Thermotogales: large-scale inversions and strain diversification of CRISPR sequences. *J Bacteriol*, **188**, 2364-2374 (2006).

84. Va´zquez M.J., Alonso J.L., Domı´nguez H. and Parajo´ J.C., Xylooligosaccharides: manufacture and applications. *Trends Food Sci Tech*, **11**, 387-393 (2000).

85. Tan S.S., Li D.Y., Jiang Z.Q., Zhu Y.P., Shi B. and Li L.T., Production of xylobiose from the autohydrolysis explosion liquor of corncob using *Thermotoga maritima* xylanase B (XynB) immobilized on nickel-chelated Eupergit C. *Bioresour Technol*, **99**, 200-204 (2008).

86. Xue Y., Wu A., Zeng H. and Shao W., High-level expression of an alpha-L-arabinofuranosidase from *Thermotoga maritima* in *Escherichia coli* for the production of xylobiose from xylan. *Biotechnol Lett*, **28**, 351-356 (2006).

87. Jiang Z., Li X., Yang S., Li L. and Tan S., Improvement of the breadmaking quality of wheat flour by the hyperthermophilic xylanase B from *Thermotoga maritima*. *Food Res Int*, **38**, 37-43 (2005).

88. Jiang Z.Q., Li X.T., Yang S.Q., Li L.T., Li Y. and Feng W.Y., Biobleach boosting effect of recombinant xylanase B from the hyperthermophilic *Thermotoga maritima* on wheat straw pulp. *Appl Microbiol Biotechnol*, **70**, 65-71 (2006).

89. Liebl W., Winterhalter C., Baumeister W., Armbrecht M. and Valdez M., Xylanase attachment to the cell wall of the hyperthermophilic bacterium *Thermotoga maritima*. *J Bacteriol*, **190**, 1350-1358 (2008).

90. Kleine J. and Liebl W., Comparative characterization of deletion derivatives of the modular xylanase XynA of *Thermotoga maritima*. *Extremophiles*, **10**, 373-381 (2006).

91. Winterhalter C., Heinrich P., Candussio A., Wich G. and Liebl W., Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*. *Mol Microbiol*, **15**, 431-444 (1995).

92. Jun H., Bing Y., Keying Z., Xuemei D. and Daiwen C., Thermostable carbohydrate binding module increases the thermostability and substrate-binding capacity of *Trichoderma reesei* xylanase 2. *N Biotechnol*, **26**, 53-59 (2009).

93. Verjans P., Dornez E., Segers M., Van Campenhout S., Bernaerts K., Belien T., Delcour J.A. and Courtin C.M., Truncated derivatives of a multidomain thermophilic glycosyl hydrolase family 10 xylanase from *Thermotoga maritima* reveal structure related activity profiles and substrate hydrolysis patterns. *J Biotechnol*, **145**, 160-167.

94. Ma J.K., Drake P.M. and Christou P., The production of recombinant pharmaceutical proteins in plants. *Nat Rev Genet*, **4**, 794-805 (2003).

95. Kim S., Lee D.S., Choi I.S., Ahn S.J., Kim Y.H. and Bae H.J., *Arabidopsis thaliana* Rubisco small subunit transit peptide increases the accumulation of *Thermotoga maritima* endoglucanase Cel5A in chloroplasts of transgenic tobacco plants. *Transgenic Res* (2009).

96. Santa-Maria M.C., Chou C.J., Yencho G.C., Haigler C.H., Thompson W.F., Kelly R.M. and Sosinski B., Plant cell calcium-rich environment enhances thermostability of recombinantly produced alpha-amylase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biotechnol Bioeng*, **104**, 947-956 (2009).

97. Zhang Y., Thiele I., Weekes D., Li Z., Jaroszewski L., Ginalski K., Deacon A.M., Wooley J., Lesley S.A., Wilson I.A., Palsson B., Osterman A. and Godzik A., Three-dimensional structural view of the central metabolic network of *Thermotoga maritima*. *Science*, **325**, 1544-1549 (2009).

98. Lesley S.A., Kuhn P., Godzik A., Deacon A.M., Mathews I., Kreusch A., Spraggon G., Klock H.E., McMullan D., Shin T., Vincent J., Robb A., Brinen L.S., Miller M.D., McPhillips T.M., Miller M.A., Scheibe D., Canaves J.M., Guda C., Jaroszewski L., Selby T.L., Elsliger M.A., Wooley J., Taylor S.S., Hodgson K.O., Wilson I.A., Schultz P.G. and Stevens R.C., Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci U S A*, **99**, 11664-11669 (2002).

Table 1.1. *Thermotoga* species.

Species	T_{opt} (°C)	Genome size (Mbp)	GenBank Accession	Isolation site	Country	Reference
T. lettingae TMO	65	2.14	CP000812	sulfate-reducing bioreactor	Netherlands	[11]
T. elfii SEBR 6459	66			oil field	Sudan	[15]
T. hypogea SEBR 7054	70			oil-producing well	Cameroon	[40]
T. subterranea SL1	70			continental oil reservoir	France	[16]
T. thermarum LA3	70			continental solfataric springs	Djibouti	[17]
T. neapolitana NS-E	77	1.88	CP000916	shallow submarine hot springs	Italy	[3]
T. petrophila RKU-1	80	1.82	CP000702	oil reservoir	Japan	[10]
T. naphthophila RKU-10	80	1.81	CP001839	oil reservoir	Japan	[10]
T. maritima MSB8	80	1.86	AE000512	geothermally heated sea floors	Italy	[1]
T. sp. strain RQ2	76-82	1.88	CP000969	geothermally heated sea floors	Azores	[1]

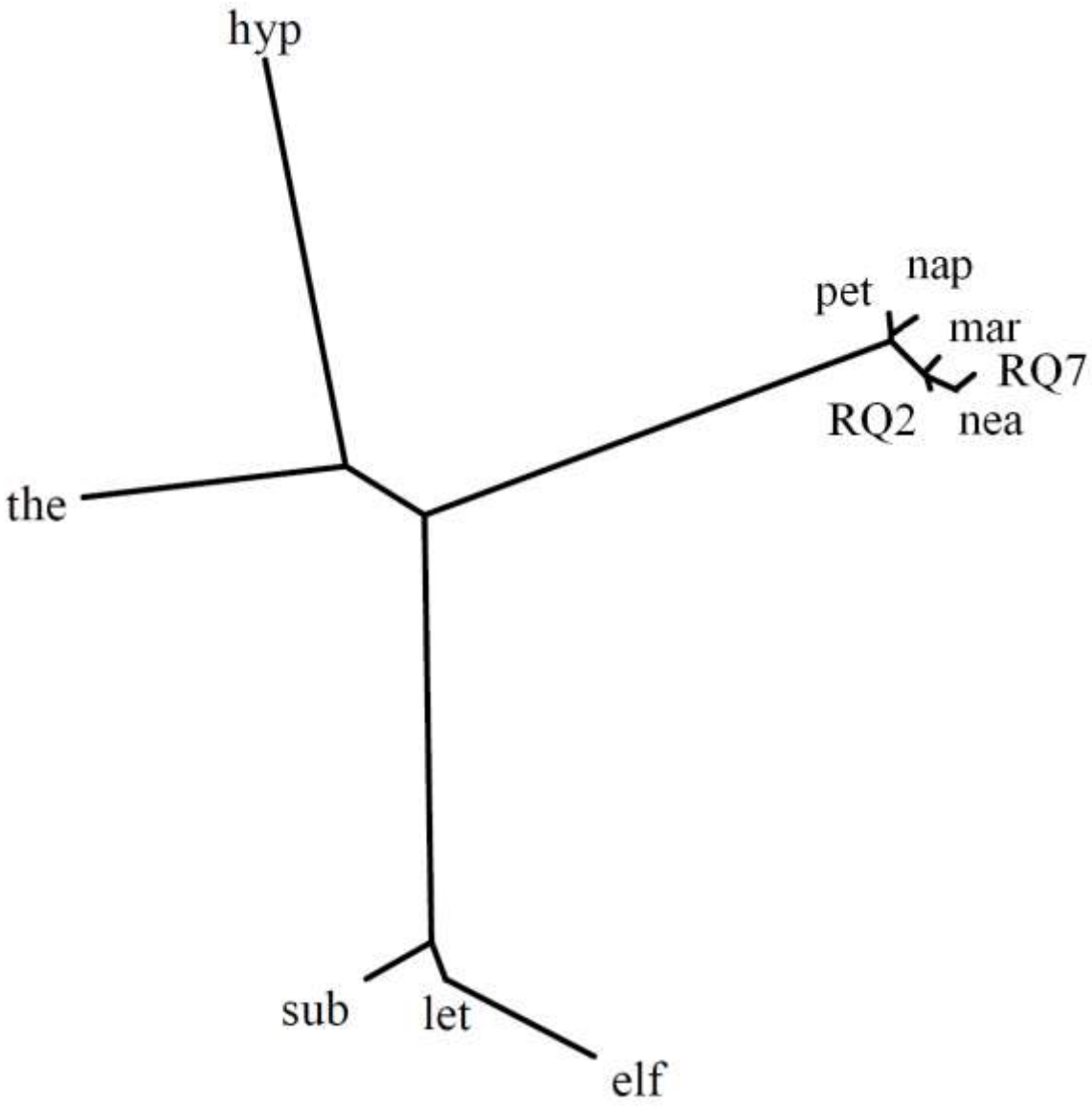

Figure 1.1. Unrooted 16S rRNA tree of *Thermotoga* species created using the Mobyle portal (http://mobyle.pasteur.fr/). hyp: *T. hypogea*; sub: *T. subterranea*; let: *T. lettingae*; elf: *T. elfii*; pet: *T. petrophila*; nap: *T. naphthophila*; RQ2: *T.* sp. RQ2; mar: *T. maritima*; nea: *T. neapolitana*; RQ7: *T.* sp. RQ7.

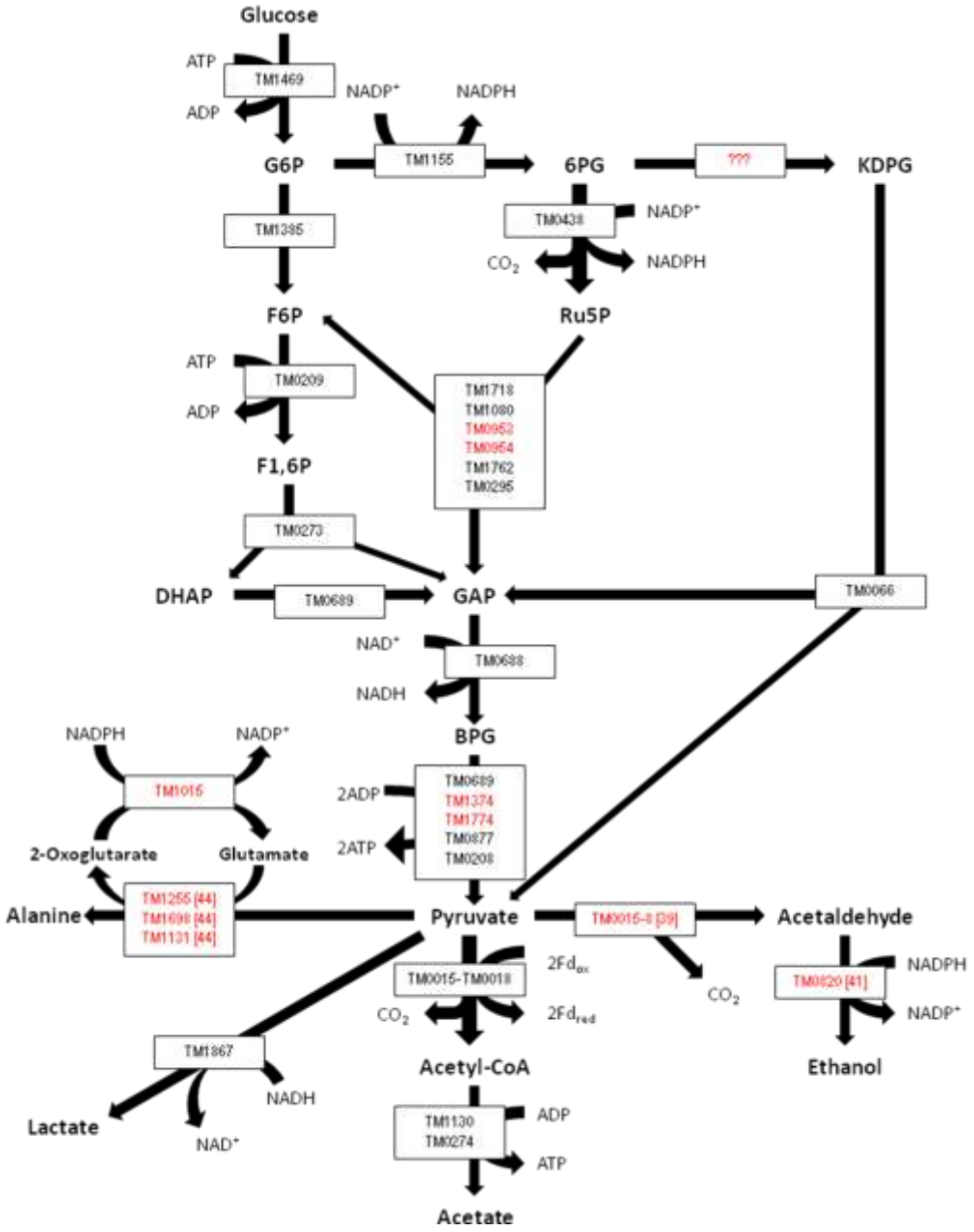

Figure 1.2. Central metabolism of *T. maritima* indicating open reading frames encoding enzymes involved in catabolic pathways. The indicated function has been experimentally validated for genes in black. TM1255, TM1698, and TM1131 are given as alanine aminotransferase candidates, based on protein BLAST of the characterized enzyme from *P. furiosus* against the *T. maritima* genome. G6P: glucose 6-phosphate; 6PG: 6-phosphogluconate; KDPG: 2-keto-3-deoxy-6-phosphogluconate; F6P: fructose 6-phosphate;

Ru5P: ribulose 5-phosphate; F1,6P: fructose 1,6-*bis*phosphate; DHAP: dihydroxyacetone phosphate; GAP: glyceraldehyde 3-phosphate; BPG: 1,3-*bis*phosphoglycerate.

A.

B.

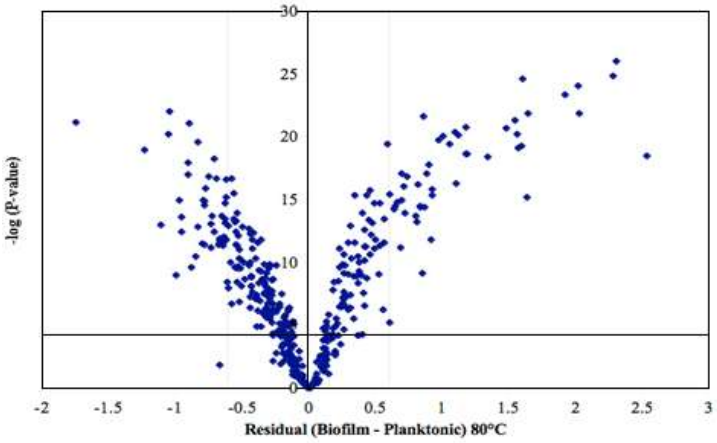

Figure 1.3. Transcriptional analysis of biofilm formation in *T. maritima* [59]; (A) biofilm cell samples were collected from a high-temperature continuous anaerobic bioreactor where the encricled region shows biofilm formed on nylon mesh and reactor walls, (B) volcano plot showing differential gene expression in planktonic and biofilm *T. maritima* cells grown in chemostat culture at 80 °C (horizontal line indicates Bonferroni correction, $-\log_{10} P$ value ≥ 4.5).

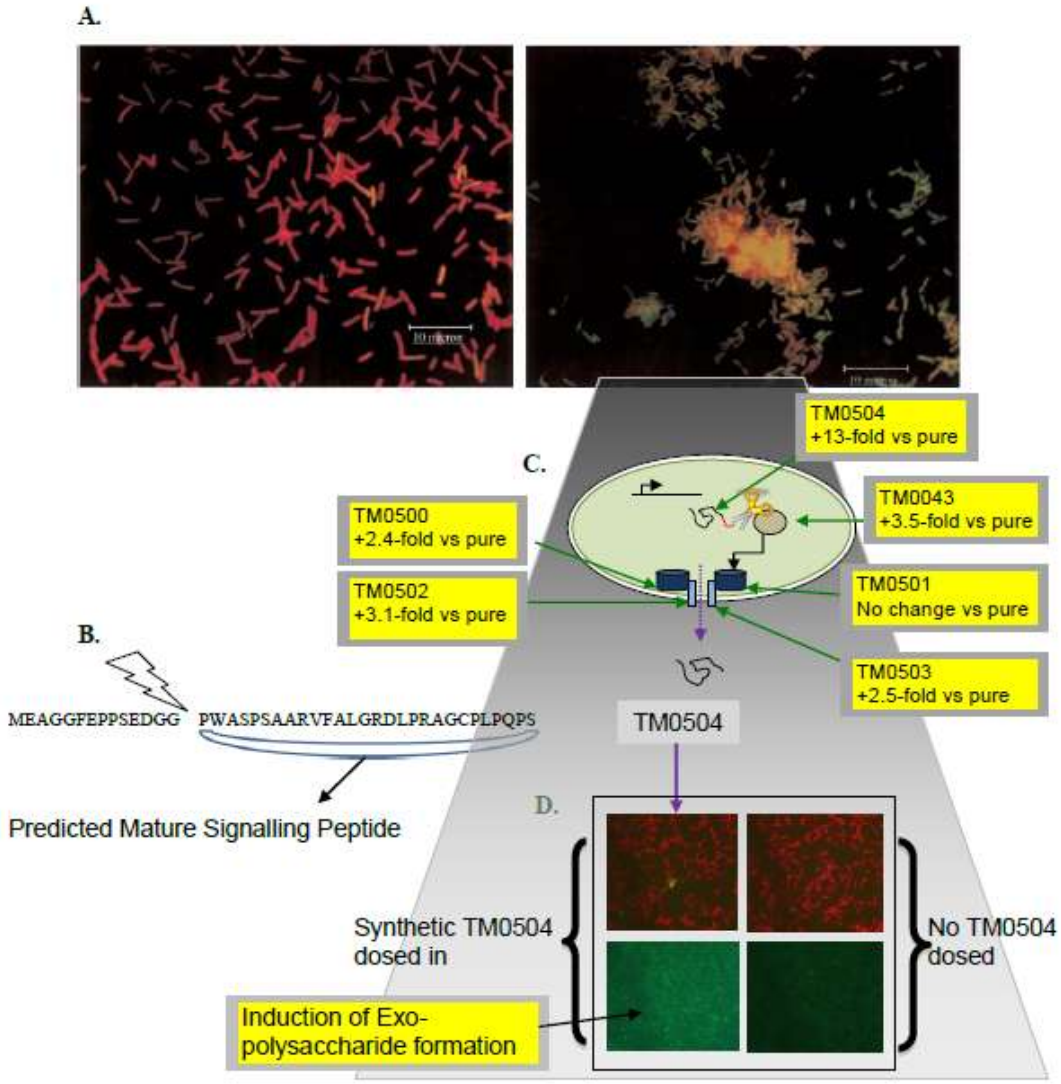

Figure 1.4. Overview of the proposed peptide-based (TM0504) signaling mechanism in *T. maritima* [60, 61]. (A) Epifluorescence images of *T. maritima* mono-culture (left) and *T. maritima* and *M. jannaschii* co-culture (right), where *T. maritima* displays rod morphology while *M. jannaschii* is the cocci bound up in the middle of the aggregate; (B) TM0504 putative signalling peptide with its predicted cleavage site after the GG motif; (C) schematic of the *T. maritima* cells in the co-culture summarizing the predicted export pathway for TM0504; (D) Acridine orange (top) and Calcofluor (bottom) stained cells of *T. maritima* 30 min after dosing with either buffer PBS (right), or synthetic TM0504 peptide (left). Reproduced with permission from M.R. Johnson, C.I. Montero, S.B. Conners, K.R.

Shockley, M.A. Pysz and R.M. Kelly, (2004), (*Biochemical Society Transactions*), (32), (188-192). © the Biochemical Society (http://www.biochemsoctrans.org).

Chapter 2

Extreme thermophiles: Moving beyond single-enzyme biocatalysis

Andrew D. Frock and Robert M. Kelly

Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

Published in *Current Opinion in Chemical Engineering* 1(4): 363-372 (2012)

ABSTRACT

Extremely thermophilic microorganisms have been sources of thermostable and thermoactive enzymes for over 30 years. However, information and insights gained from genome sequences, in conjunction with new tools for molecular genetics, have opened up exciting new possibilities for biotechnological opportunities based on extreme thermophiles that go beyond single-step biotransformations. Although the pace for discovering novel microorganisms has slowed over the past two decades, genome sequence data have provided clues to novel biomolecules and metabolic pathways, which can be mined for a range of new applications. Furthermore, recent advances in molecular genetics for extreme thermophiles have made metabolic engineering for high temperature applications a reality.

INTRODUCTION

Microbial life in terrestrial hot springs, such as those present in Yellowstone National Park (USA), has been observed and studied in earnest since at least as early as the 1960s [1–3]. However, the discovery of microbial life in deep sea thermal vents and shallow marine seeps in volcanic regions of the world in the late 1970s and early 1980s [4–6] led to the realization that the so called extreme thermophiles ($T_{opt} \geq 70°C$) were more phylogenetically, physiologically and geographically diverse than first thought. Many novel genera and species, both archaea and bacteria, were isolated and described, including several that became model microorganisms because of their interesting metabolic features and relative ease of cultivation in laboratory settings (see Table 2.1). Three archaea, *Pyrococcus furiosus*, a marine fermentative anaerobe (T_{opt} 98–100°C) [5,7], *Thermococcus*

kodakaerensis (T_{opt} 85°C) also a marine fermentative anaerobe [8,9], and *Sulfolobus solfataricus*, a terrestrial heterotrophic acidophilic aerobe (T_{opt} 80°C, pH_{opt} 3.5) [10,11], and one bacterium, *Thermotoga maritima*, a marine fermentative anaerobic bacterium (T_{opt} 80°C) [6,12,13], became the focus of most fundamental and biotechnological studies addressing life at high temperatures. In fact, much of what is known about extreme thermophile physiology and enzymology to date has been based on the study of these 'model' microorganisms (subclassified as hyperthermophiles because of their $T_{opt} \geq 80°C$). Recently, as discussed below, genetic systems have been established for the three archaea, further enhancing their value as model systems.

Difficulties in isolating extreme thermophiles because of the harsh and often inaccessible environments from which they come, and subsequently cultivating these microorganisms in laboratory settings, initially presented significant challenges to their study and, consequently, to associated biotechnological applications. In the early 1990s, however, successful attempts to clone and express genes from hyperthermophiles in mesophilic recombinant hosts (e.g. *Escherichia coli*) facilitated efforts to produce specific enzymes for characterization and application [14,15]. Furthermore, since 1995, when the genome sequence of the hyperthermophile *Methano(caldo)coccus jannaschii* was reported [16], related efforts for many high temperature microorganisms have enabled and accelerated projects to not only identify promising biocatalysts for single-step biotransformations but also to discover metabolic pathways, cellular features, and biological phenomena that are relevant to biotechnology. Now, armed with virtually unlimited access to genome sequence data, and aided by molecular genetics and 'omics' tools, the prospects for biotechnology at elevated temperatures have never been more promising, and poised to go beyond single-step biocatalysts (i.e. the use of a single enzyme for a single biotransformation).

ISOLATION OF NOVEL EXTREME THERMOPHILES

For the most part, the isolation of many currently known and studied extremely thermophilic microorganisms happened in the late 1970s through the early '90s [17,18], although isolation of several Sulfolobus species was reported in the early 1970s [19]. Description of new isolates continues to appear in the literature, but most often reside within genera of previously studied extreme thermophiles. As such, discovery of significantly different extremely thermophilic genera and species is becoming rare. Coupled to this is the fact that the criteria for designating a new isolate as 'novel' have become more stringent, given the availability of genome sequences and related quantitative measures for differentiating among microorganisms. Thus, new reports of truly novel extreme thermophiles based on more than marginal differences in 16S rRNA phylogeny and subtle variations in growth physiology are infrequent.

Nonetheless, interesting isolates continue to be reported. In the last several years, several new fermentative anaerobes have been isolated (Table 2.2). For example, *Acidilobus saccharovorans* [20,21], a terrestrial thermoacidophilic crenarchaeon (T_{opt} 80–85°C, pH_{opt} 3.5–4) contributes to closure of the anaerobic carbon cycle in terrestrial hot springs by complete oxidation of organic compounds (acetate, ethanol, and lactate). Also, *Pyrococcus yayanosii* is one of only a handful of microbes (and the only *Pyrococcus* species) shown to be an obligate piezophile (T_{opt} 98°C, P_{opt} 52 MPa); growth was not observed at atmospheric pressure, but rather between 20 and 120 MPa [22].

One promising approach to the isolation of novel microorganisms from high temperature environments is to consider mixed cultures and consortia. For example, a three species archaeal consortium, capable of deconstructing crystalline cellulose, was recently

described [23]. This consortium, with members related to archaea from the genera *Ignisphaera*, *Thermofilum*, and *Pyrobaculum*, was able to partially dissolve filter paper after incubating for 30 days. In other cases, inter-species interactions are critical for isolation of new extreme thermophiles. Along these lines, the discovery of nanoarchaea was reported, occurring as parasitic partners with the extremely thermophilic archaeon, *Ignicoccus hospitalis* [24]. Evidence from genome sequence data of lateral gene transfer between *Nanoarchaeum equitans* and *I. hospitalis* was linked to how these two microorganisms adapted to growth by sulfur–H_2 respiration coupled to inorganic carbon and nitrogen fixation. As DNA sequencing rates increase and costs continue to decrease, metagenomes from thermal environments will be examined for hints to unusual physiologies [25], as well as clues to new enzymes and metabolic pathways, the components of which can be produced recombinantly (perhaps, in extremely thermophilic hosts) for further analysis.

GENOMICS OF EXTREME THERMOPHILES

As mentioned above, access to genome sequence information for extreme thermophiles (which has facilitated functional genomics, proteomics and other 'omics-based' approaches) has enabled rapid advances in our understanding of these microbes' physiology over the past 15 years, despite the lack of genetic systems. In some cases, genome sequence information has been used to ask global questions about how thermophilic proteins fold and function at high temperatures [26]. Along these lines, lower levels of structural disorder and functional simplification determined at the level of individual genes and proteins, as well as of whole genomes, was proposed as the basis for prokaryotic thermophily [27]. Genome sequence information led to the provocative proposal that high

temperature bacteria and archaea have lower spontaneous mutation rates than mesophiles [28]. The relationship between mesophiles and thermophiles has been given new perspective through comparative genomics. For example, reverse gyrase, an enzyme involved in thermophilic transcriptional processes to deal with uncoiling DNA, was once thought to be a defining feature of extreme thermophiles [29,30]. However, the genome sequence of *Nautilla profoundicola* (T_{opt} 40–45°C) encodes the gene for this enzyme, probably acquired through lateral gene transfer within the submarine hydrothermal environment that it inhabits [31]. The phylogenetic lines between thermophily and mesophily may be more blurred than expected. Mesophilic members of the Order Thermotogales (or 'mesotogas') were recently identified [32], as well as related species that have very broad growth temperature ranges [33], raising questions about the thermal direction of microbial evolution. It is also clear that extreme thermophiles from the same genus can have differences in their genome sequences that map to subtle but significant differences in their growth physiology [34]. Finally, examination of the *P. furiosus* and *S. solfataricus* proteomes, with respect to metal content, revealed a far more extensive set of metals implicated in protein structure and function, including 'non-biological' metals, such as uranium and vanadium [35]. Since similar results were found for *E. coli*, it seems that the microbial world employs more of the periodic table for biological function than previously thought.

In certain cases, genome sequences of extreme thermophiles have revealed physiological insights not previously known, despite years of microbiological study. The genome sequence of *Metallosphaera sedula*, an extremely thermoacidophilic archaeon (T_{opt} 73°C, pH_{opt} 2.0) [36], originally isolated and studied for its ability to mobilize metals from sulfidic ores [37], confirmed the presence of a novel CO2 fixation pathway (3-

hydroxypropionate/ 4-hydroxybutyrate cycle) [38]. Consequently, this archaeon possesses a much more versatile growth physiology than previously thought [39]. In fact, *M. sedula* bioenergetics can be fueled by CO_2/H_2 autotrophy, heterotrophy, and metal/sulfur oxidation, separately or in combination (mixotrophy) [39].

Genome sequencing has also brought renewed interest to extreme thermophiles that had been isolated 20 years ago. For example, the genome sequences of *Caldicellulosiruptor saccharolyticus* [40,41] and *Caldicellulosiruptor bescii* [42] (formerly *Anaerocellum thermophilum*) shed new light on microbial mechanisms [43,44] and enzymology [45– 47] of lignocellulose deconstruction. Furthermore, comparative genomics of eight *Caldicellulosiruptor* species revealed key determinants of lignocellulose degradation, based on the core and pan genomes of this genus [48].

MOLECULAR GENETICS TOOLS FOR EXTREME THERMOPHILES

As mentioned above, studies of extreme thermophiles, and hence biotechnological applications, have been hampered to a certain extent by the limited availability of tools for genetic manipulation and metabolic engineering. Selection of thermostable markers, high temperature solid media, need for anaerobic conditions, and lack of defined media are among the challenges faced. Despite these obstacles, in recent years new molecular genetics systems have been developed for several extreme thermophiles, as well as being refined and expanded in cases where these had previously existed [49,50].

Initial success with molecular genetics in extreme thermophiles was achieved with the extremely thermoacidophilic archaeon S. solfataricus. The fact that this archaeon grows aerobically and can be cultivated on solid media no doubt contributed to progress in this

regard. Most genetic manipulation in this species has utilized lacS mutants [51,52]. DNA is introduced to the cell via electroporation and growth on lactose (which requires lacS complementation) is used to select for successful transformation. This strategy has been used as the basis to generate deletion mutants for the study of copper response [53,54], toxin–antitoxin pairs [55], and antimicrobial proteins [56]. In addition, virus-based vectors encoding genes under the control of arabinose-inducible and heat-inducible promoters have been used to overexpress proteins in S. solfataricus [57,58], and reporter systems for monitoring gene expression have been developed based on b-galactosidase (LacS) [58] and b-glucuronidase (GusB) [59].

In recent years, success with molecular genetics has gone beyond *Sulfolobus* species. Development of genetic techniques for euryarchaeal Thermococcales has been facilitated by isolation of naturally competent strains of *T. kodakaraensis* and *P. furiosus* [60,61]. Manipulation of both species initially hinged upon *pyrF*, the gene encoding orotidine-5'-monophosphate (OMP) decarboxylase; 5-fluoroorotic acid (5-FOA) can be used to select for *pyrF* mutants, because cells with the functional gene are sensitive to this compound. On the contrary, media lacking uracil can be used to select for complementation of *pyrF*, as the gene is required for uracil biosynthesis. Once *pyrF* mutants of *P. furiosus* and T. *kodakaraensis* were generated [60,62], methods were developed for performing selections in complex media [63,64], overexpressing proteins, secreting proteins [65], and generating 'markerless' deletions that allow multiple manipulations [62,66].

These tools have facilitated more targeted investigation of fermentative H_2 metabolism, a defining feature of *Pyrococcus* and *Thermococcus* species. One problem, central to understanding hydrogenase function, is producing sufficient amounts of active forms of the enzyme to study and evaluate. This issue was addressed for soluble

hydrogenase I (SHI) from *P. furiosus*, which was recombinantly produced in *E. coli* by using an anaerobically driven promoter native to this bacterium [67]. Furthermore, an engineered hydrogenase, consisting of two of the four native subunits, was overexpressed heterologously in the native host, *P. furiosus*, and found to utilize electrons directly from pyruvate ferredoxin oxidoreductase without the involvement of an intermediate electron carrier (NADPH or ferredoxin) [68]. There are biotechnological implications of this work. Hydrogen production is normally growth-associated, but an electron carrier-independent hydrogenase might partially decouple these processes, resulting in higher yields of hydrogen.

Directed gene knockouts have been pivotal in probing metabolic mechanisms related to H_2 generation in *P. furiosus* [69] and *T. kodakaerensis* [70,71]. Deletion of *surR*, which encodes a transcriptional regulator, revealed that this gene is required for expression of the membrane bound hydrogenase [71], the primary H_2-evolving enzyme [70]. Deletion of cytosolic hydrogenases limited growth of the microbe, but increased specific H_2 production rates, suggesting that re-oxidation of H_2 by these enzymes is an important energy conservation mechanism in this species [70,71]. Deletion of the membrane-bound oxidoreductase complex, which is required for sulfide production [69,70], also resulted in increased specific hydrogen production [69,70,71].

BIOCATALYSIS AT ELEVATED TEMPERATURES

To date, the largest impact enzymes from extreme thermophiles have had on science and technology relates to their use in catalyzing the polymerase chain reaction (PCR) [72]. Beyond that, the first thoughts for biotechnological applications for enzymes

from these extremophiles turned to thermostable and thermoactive ‘drop in’ replacements for industrial enzymes already in use (e.g. proteases, amylases, glucose isomerases). Several insightful reviews have appeared over the past 25 years that examine scientific and biotechnological aspects of biocatalysis at elevated temperatures (for a recent excellent review see [73]). One of the opportunities afforded by biocatalysts that function at temperatures approaching and exceeding 100°C is the exploitation of the intrinsic features that underlie their unprecedented thermostability and thermoactivity. In this sense, the key strategic questions to consider are whether the biocatalytic process requires high temperatures, and, as a consequence, if high temperatures confer any strategic advantage.

The structural stability of enzymes from extreme thermophiles makes them attractive candidates for protein engineering, based on the premise that they are less susceptible to undesired consequences from genetic manipulations. The extensive protein engineering of β-glycosidases from *S. solfataricus* to catalyze chemoenzymatic synthesis of oligosaccharides serves as an excellent example in this regard [74]. In another case, cofactor specificity of an alcohol dehydrogenase from *P. furiosus* could be modified by site-directed mutations in the co-factor binding pocket [75]. Building on the theme of enzymes from extreme thermophiles as robust protein engineering targets, the transcription factor Sso7d from *S. solfataricus* was used as a scaffold for creating binding partners to a variety of biomolecules [76]. The thermostability of these enzymes was also a key factor in examining the effect of microwaves on biocatalysis. Under certain conditions, the biocatalytic rates of enzymes from *P. furiosus*, *S. solfataricus* and *T. maritima* could be enhanced under microwave irradiation, a prospect not possible for mesophilic enzymes whose stability was significantly impacted by microwaves [77].

Key to many enzyme applications is the capability to immobilize the biocatalyst for stabilization or for a specific process strategy (e.g. re-use or localization). Enzymes from extreme thermophiles present some opportunities, as well as challenges, when it comes to immobilization [78]. For example, entrapment in a porous gel as a means of immobilization, commonly employed for mesophilic enzymes, is problematic because of the thermolability of the matrix. Alternatively, carbohydrate binding domains from extreme thermophiles can be employed in fusions with extremely thermophilic enzymes for immobilization, as was demonstrated with the chitin-binding domain from a *P. furiosus* chitinase and the xylose isomerase from *Thermotoga neapolitana* [79].

Not surprisingly, enzymes from extreme thermophiles have been examined closely for applications related to the emergence of biofuels. It has been argued that the deconstruction of lignocellulose is best done at elevated temperatures, either because thermal factors facilitate this process, or because biofuels bioprocessing already involves thermal steps for pretreatment [45]. To this point, computational studies suggest that thermal contributions to enzyme plasticity and molecular motion at high temperatures play a role in enhancing cellulose-binding domain and catalytic domain synergy in cellulose [80]. The genomes of species within the extremely thermophilic genus *Caldicellulosiruptor* encode a host of multidomain glycoside hydrolases that contribute to the breakdown of crystalline cellulose and hemicellulose [43,48]. Recent work has looked at the contributions of the various domains within these enzymes to complex carbohydrate hydrolysis [46,81] and the potential role of certain multidomain glycoside hydrolases, which also use S-layer homology domains to anchor to the cell envelope [82]. The high temperature, cellulose-degrading consortium of archaea, described above, also gave rise to the discovery of a

novel hyperthermophilic cellulase, a multi-domain enzyme exhibiting optimal activity at 109°C [23].

Biocatalysis based on whole cells has also been the objective of efforts with extreme thermophiles. For example, the degradation of toxic pollutants has been demonstrated; *S. solfataricus* 98/2 could utilize phenol for growth in a fed-batch bioreactor [83]. The recovery of base, precious and strategic metals through whole cell bio-oxidation processes has been a long-term goal for extreme thermoacidophiles and recent efforts have focused on identifying process bottlenecks and improved processing strategies. It is becoming clear, not surprisingly, that mixed cultures will be the most effective approach to biohydrometallurgy and could be a way to overcome problems with surface passivation by jarosite and elemental sulfur by-products [84].

METABOLIC ENGINEERING APPLICATIONS AND OPPORTUNITIES AT HIGH TEMPERATURES

Until very recently, metabolic engineering involving enzyme and pathways from extreme thermophiles were carried out in mesophilic hosts. For example, the archaeal isoprenoid ether lipid biosynthesis pathway was reconstructed in *E. coli* to produce digeranylgeranylglyceryl phosphate (DGGGP) [85]. The strategic use of temperature for bioprocessing was demonstrated by cloning a hyperthermophilic α-amylase from *T. maritima* into transgenic sweet potato [86]. Below 40°C, the normal growth temperature range of the plant, no significant amylase activity could be detected nor was plant growth and development impacted. But upon switching to 80°C, the hyperthermophilic enzyme was activated such that the plant storage root was rapidly hydrolyzed. This created a mechanism

for first producing the sweet potato starch then using a thermal switch to convert the starch into a fermentable sugar for biofuels applications.

Of course, the next frontier is metabolic engineering at high temperatures (Table 2.3). In fact, the rapid progress with molecular genetics for extreme thermophiles has given rise to the prospect of using these microorganisms as recombinant hosts for metabolic engineering. The three archaea shown in Table 2.1 offer the most near-term prospects in this regard. For example, *S. solfatarcius* P2 differs from *S. solfataricus* 98/2 in that the latter strain is less able grow on surfaces. However, the insertion of two genes from P2 into 98/2, encoding α-mannosidase and β-galactosidase, enabled 98/2 to mimic P2 by attaching to glass and forming static biofilms [87]. The use of thermally driven gene regulation has been elegantly demonstrated using *P. furiosus* as a recombinant host. When microorganisms are employed to generate a desired product, the production pathway often competes with the microbe's natural biosynthesis pathways for key intermediates or cofactors. In this situation, one might envision an ideal two-stage process in which biomass is generated during the first stage with minimal product formation, while cellular activity ceases and the desired product is generated during the second stage (see Figure 2.1). The recent proof-of-concept work of Basen et al. [63] uses temperature as the switch to halt cell growth and initiate product formation. Lactate dehydrogenase from *Caldicellulosiruptor bescii* (T_{opt} = 78°C) was cloned into *P. furiosus* (T_{opt} = 98°C) under the control of a *P. furiosus* 'cold shock' promoter that is turned on at 70–75°C. In this situation *P. furiosus* can be cultured at 98°C until it reaches a high cell density, at which point it can be transferred to 72°C, resulting in expression of the heterologous lactate dehydrogenase and formation of lactate (a product that *P. furiosus* is unable to produce naturally).

In the near future, we will probably see the first demonstrations of high temperature strains used as hosts for biotechnological applications. For example, in biofuels production, the ultimate goal is to create metabolically engineered extreme thermophiles that breakdown lignocellulose and convert fermentable sugars to liquid biofuels (so-called 'consolidated bioprocessing'). At high temperatures, there is the possibility that biofuels can be recovered directly through direct evaporation and distillation (see Figure 2.2). Other strategic uses of high temperatures will probably emerge as tools for molecular genetics in extreme thermophiles become more firmly established.

CONCLUSIONS

Although discovered more than 40 years ago, in many ways extremely thermophilic microorganisms are just at the beginning when it comes to biotechnological applications. Virtually any enzyme that is identified in a mesophile has a homologous version in an extreme thermophile, typically with significantly higher levels of thermostability, if not thermoactivity. As genetic systems for extreme thermophiles become more widely used and more tractable, the challenge will be to exploit elevated temperatures to improve upon existing bioprocessing strategies, or even better, make possible novel multi-step, biotransformations. Ideas along these lines have already been proposed [88], and it is only a matter of time before these process concepts are demonstrated and put into practice.

ACKNOWLEDGEMENTS

The authors acknowledge support from the U.S. Department of Energy ARPA-E Program (DE-AR0000081), U.S. Department of Energy GTL Program (DG-FG02-08ER64687), the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy Bioenergy Research Center funded by the Office of Biological and Environmental Research in the DOE Office of Science [contract no. DE-PS02-06ER64304] [DOE 4000063512], U.S. National Institute of Health (R01GM90209), and the U.S. Defense Threat Reduction Agency (HDTRA1-09-1-0030). AD Frock acknowledges support from an U.S. National Institutes of Health Biotechnology Traineeship (T32 GM008776-06).

REFERENCES

Papers of particular interest, published within the period of review, have been highlighted as:

● of special interest

●● of outstanding interest

1. Brock TD: **Life at high temperatures. Evolutionary, ecological, and biochemical significance of organisms living in hot springs is discussed**. *Science* 1967, **158**:1012-1019.

2. Brock TD: **Micro-organisms adapted to high temperatures**. *Nature* 1967, **214**:882-885.

3. Bott TL, Brock TD: **Bacterial growth rates above 90 degrees C in Yellowstone hot springs**. *Science* 1969, **164**:1411-1412.

4. Jones WJ, Leigh JA, Mayer F, Woese CR, Wolfe RS: ***Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent**. *Arch Microbiol* 1983, **136**:254-261.

5. Fiala G, Stetter KO: ***Pyrococcus furiosus*** **sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100C**. *Arch Microbiol* 1986, **145**:56-61.

6. Huber R, Langworthy TA, Konig H, Thomm M, Woese CR, Sleytr UB, Stetter KO: ***Thermotoga maritima*** **sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90C**. *Arch Microbiol* 1986, **144**:324-333.

7. Robb FT, Maeder DL, Brown JR, DiRuggiero J, Stump MD, Yeh RK, Weiss RB, Dunn DM: **Genomic sequence of hyperthermophile,** ***Pyrococcus furiosus*****: implications for physiology and enzymology**. *Methods Enzymol* 2001, **330**:134-157.

8. Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T: **Complete genome sequence of the hyperthermophilic archaeon** ***Thermococcus kodakaraensis*** **KOD1 and comparison with** ***Pyrococcus*** **genomes**. *Genome Res* 2005, **15**:352-363.

9. Morikawa M, Izawa Y, Rashid N, Hoaki T, Imanaka T: **Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic** ***Pyrococcus*** **sp**. *Appl Environ Microbiol* 1994, **60**:4559-4566.

10. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, et al.: **The complete genome of the crenarchaeon** ***Sulfolobus solfataricus*** **P2**. *Proc Natl Acad Sci U S A* 2001, **98**:7835-7840.

11. Zillig W, Stetter KO, Wunderl S, Schulz W, Priess H, Scholz I: **The** ***Sulfolobus-"Caldariella"*** **group: Taxonomy on the basis of the structure of DNA-dependent RNA polymerases**. *Arch Microbiol* 1980, **125**:269-269.

12. Conners SB, Mongodin EF, Johnson MR, Montero CI, Nelson KE, Kelly RM: **Microbial biochemistry, physiology, and biotechnology of hyperthermophilic** ***Thermotoga*** **species**. *FEMS Microbiol Rev* 2006, **30**:872-905.

13. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, et al.: **Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of** ***Thermotoga maritima***. *Nature* 1999, **399**:323-329.

14. Fabry S, Lehmacher A, Bode W, Hensel R: **Expression of the glyceraldehyde-3-phosphate dehydrogenase gene from the extremely thermophilic archaebacterium** ***Methanothermus fervidus*** **in** ***E. coli*****. Enzyme purification, crystallization, and preliminary crystal data**. *FEBS Lett* 1988, **237**:213-217.

15. Moracci M, La Volpe A, Pulitzer JF, Rossi M, Ciaramella M: **Expression of the thermostable beta-galactosidase gene from the archaebacterium** ***Sulfolobus solfataricus*** **in** ***Saccharomyces cerevisiae*** **and characterization of a new inducible promoter for heterologous expression**. *J Bacteriol* 1992, **174**:873-882.

16. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, et al.: **Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii***. *Science* 1996, **273**:1058-1073.

17. Stetter KO: **Hyperthermophiles in the history of life**. *Philos Trans R Soc Lond B Biol Sci* 2006, **361**:1837-1842; discussion 1842-1833.

18. Stetter KO: **History of discovery of the first hyperthermophiles**. *Extremophiles* 2006, **10**:357-362.

19. Brock TD, Brock KM, Belly RT, Weiss RL: ***Sulfolobus*: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature**. *Arch Microbiol* 1972, **84**:54-68.

20. Mardanov AV, Svetlitchnyi VA, Beletsky AV, Prokofeva MI, Bonch-Osmolovskaya EA, Ravin NV, Skryabin KG: **The genome sequence of the crenarchaeon *Acidilobus saccharovorans* supports a new order, Acidilobales, and suggests an important ecological role in terrestrial acidic hot springs**. *Appl Environ Microbiol* 2010, **76**:5652-5657.

21. Prokofeva MI, Kostrikina NA, Kolganova TV, Tourova TP, Lysenko AM, Lebedinsky AV, Bonch-Osmolovskaya EA: **Isolation of the anaerobic thermoacidophilic crenarchaeote *Acidilobus saccharovorans* sp. nov. and proposal of *Acidilobales* ord. nov., including *Acidilobaceae* fam. nov. and *Caldisphaeraceae* fam. nov**. *Int J Syst Evol Microbiol* 2009, **59**:3116-3122.

22. Birrien JL, Zeng X, Jebbar M, Cambon-Bonavita MA, Querellou J, Oger P, Bienvenu N, Xiao X, Prieur D: ***Pyrococcus yayanosii* sp. nov., an obligate piezophilic hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent**. *Int J Syst Evol Microbiol* 2011, **61**:2827-2831.

23. Graham JE, Clark ME, Nadler DC, Huffer S, Chokhawala HA, Rowland SE, Blanch HW, Clark DS, Robb FT: **Identification and characterization of a multidomain hyperthermophilic cellulase from an archaeal enrichment**. *Nat Commun* 2011, **2**:375.

●● Deconstruction of crystalline cellulose at 90°C, the highest temperature to date, was observed by a consortium of three hyperthermophilic archaea. A multi-domain cellulase was identified with optimal activity at 109°C.

24. Podar M, Anderson I, Makarova KS, Elkins JG, Ivanova N, Wall MA, Lykidis A, Mavromatis K, Sun H, Hudson ME, et al.: **A genomic analysis of the archaeal system *Ignicoccus hospitalis-Nanoarchaeum equitans***. *Genome Biol* 2008, **9**:R158.

25. Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings R, Fouke BW, et al.: **Metagenomes from high-**

temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* **5**:e9773.

26. Sawle L, Ghosh K: **How do thermophilic proteins and proteomes withstand high temperature?** *Biophys J* 2011, **101**:217-227.

27. Burra PV, Kalmar L, Tompa P: **Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes**. *PLoS One* 2010, **5**:e12069.

28. Drake JW: **Avoiding dangerous missense: thermophiles display especially low mutation rates**. *PLoS Genet* 2009, **5**:e1000520.

29. Bouthier de la Tour C, Portemer C, Nadal M, Stetter KO, Forterre P, Duguet M: **Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria**. *J Bacteriol* 1990, **172**:6803-6808.

30. Forterre P: **A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein**. *Trends Genet* 2002, **18**:236-237.

31. Campbell BJ, Smith JL, Hanson TE, Klotz MG, Stein LY, Lee CK, Wu D, Robinson JM, Khouri HM, Eisen JA, et al.: **Adaptations to submarine hydrothermal environments exemplified by the genome of *Nautilia profundicola***. *PLoS Genet* 2009, **5**:e1000362.

32. Nesbo CL, Kumaraswamy R, Dlutek M, Doolittle WF, Foght J: **Searching for mesophilic Thermotogales bacteria: "mesotogas" in the wild**. *Appl Environ Microbiol* 2010, **76**:4896-4900.

● Mesophilic Thermotogales have been detected in mesophilic, hydrocarbon-rich environments across the world, suggesting that they originated in thermophilic oil reservoirs, adapted to lower temperatures as the reservoirs cooled, and may have spread relatively recently.

33. Swithers KS, DiPippo JL, Bruce DC, Detter C, Tapia R, Han S, Goodwin LA, Han J, Woyke T, Pitluck S, et al.: **Genome sequence of *Kosmotoga olearia* strain TBF 19.5.1, a thermophilic bacterium with a wide growth temperature range, isolated from the Troll B oil platform in the North Sea**. *J Bacteriol* 2011, **193**:5566-5567.

34. Frock AD, Gray SR, Kelly RM: **Hyperthermophilic *Thermotoga* species differ with respect to specific carbohydrate transporters and glycoside hydrolases**. *Appl Environ Microbiol* 2012, **78**:1978-1986.

35. Cvetkovic A, Menon AL, Thorgersen MP, Scott JW, Poole FL, 2nd, Jenney FE, Jr., Lancaster WA, Praissman JL, Shanmukh S, Vaccaro BJ, et al.: **Microbial metalloproteomes are largely uncharacterized**. *Nature* 2010, **466**:779-782.

●● Novel metalloproteins were identified, in some cases incorporating novel metals (i.e. molybdenum), by a metals-based approach using liquid chromatography and mass

spectrometry, suggesting that metalloproteomes are much more extensive than previously realized.

36. Auernik KS, Maezato Y, Blum PH, Kelly RM: **The genome sequence of the metal-mobilizing, extremely thermoacidophilic archaeon *Metallosphaera sedula* provides insights into bioleaching-associated metabolism**. *Appl Environ Microbiol* 2008, **74**:682-692.

37. Huber G, Spinnler C, Gambacorta A, Stetter KO: ***Metallosphaera sedula* gen. and sp. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic archaebacteria**. *Syst Appl Microbiol* 1989, **12**:38-47.

38. Berg IA, Kockelkorn D, Buckel W, Fuchs G: **A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea**. *Science* 2007, **318**:1782-1786.

39. Auernik KS, Kelly RM: **Physiological versatility of the extremely thermoacidophilic archaeon *Metallosphaera sedula* supported by transcriptomic analysis of heterotrophic, autotrophic, and mixotrophic growth**. *Appl Environ Microbiol* 2010, **76**:931-935.

- Candidate genes of the 3-hydroxypropionate/4-hydroxybutyrate pathway were identified by analyzing transcription during autotrophic, heterotrophic, and mixotrophic growth.

40. Rainey FA, Donnison AM, Janssen PH, Saul D, Rodrigo A, Bergquist PL, Daniel RM, Stackebrandt E, Morgan HW: **Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium**. *FEMS Microbiol Lett* 1994, **120**:263-266.

41. van de Werken HJ, Verhaart MR, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE, van Niel EW, et al.: **Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus***. *Appl Environ Microbiol* 2008, **74**:6720-6729.

42. Dam P, Kataeva I, Yang SJ, Zhou F, Yin Y, Chou W, Poole FL, 2nd, Westpheling J, Hettich R, Giannone R, et al.: **Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725**. *Nucleic Acids Res* 2011, **39**:3240-3254.

- The genome of *Caldicellulosiruptor bescii* contains a high duplication of carbohydrate active domains and is enriched with genes encoding multi-domain, multi-functional carbohydrate active proteins.

43. Blumer-Schuette SE, Lewis DL, Kelly RM: **Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor***. *Appl Environ Microbiol* 2010, **76**:8084-8092.

● Of seven *Caldicellulosiruptor* species examined, four hydrolyzed crystalline cellulose and secreted a prominent S-layer protein, which is likely involved in cell-substrate interactions.

44. Yang SJ, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MW: **Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725**. *Appl Environ Microbiol* 2009, **75**:4762-4769.

45. Blumer-Schuette SE, Kataeva I, Westpheling J, Adams MW, Kelly RM: **Extremely thermophilic microorganisms for biomass conversion: status and prospects**. *Curr Opin Biotechnol* 2008, **19**:210-217.

46. VanFossen AL, Ozdemir I, Zelin SL, Kelly RM: **Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus***. *Biotechnol Bioeng* 2011, **108**:1559-1569.

47. VanFossen AL, Verhaart MR, Kengen SM, Kelly RM: **Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences**. *Appl Environ Microbiol* 2009, **75**:7718-7724.

48. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL, 2nd, Adams MW, et al.: ***Caldicellulosiruptor* core and pan genomes reveal determinants for non-cellulosomal thermophilic deconstruction of plant biomass**. *J Bacteriol* 2012.

●● A comprehensive analysis of eight *Caldicellulosiruptor* species involving microbiological, genomic, and proteomic methods revealed GH 9 and GH 48 containing multi-domain cellulases and a novel adhesin that are key for deconstruction of plant biomass in these species.

49. Taylor MP, van Zyl L, Tuffin IM, Leak DJ, Cowan DA: **Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts**. *Microb Biotechnol* 2011, **4**:438-448.

50. Leigh JA, Albers SV, Atomi H, Allers T: **Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales**. *FEMS Microbiol Rev* 2011, **35**:577-608.

●● The authors review the significant progress that has been made in the development of genetic tools for manipulation of various groups of archaea.

51. Schelert J, Dixit V, Hoang V, Simbahan J, Drozda M, Blum P: **Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption**. *J Bacteriol* 2004, **186**:427-437.

52. Worthington P, Hoang V, Perez-Pomares F, Blum P: **Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus***. *J Bacteriol* 2003, **185**:482-488.

53. Villafane A, Voskoboynik Y, Ruhl I, Sannino D, Maezato Y, Blum P, Bini E: **CopR of *Sulfolobus solfataricus* represents a novel class of archaeal-specific copper-responsive activators of transcription**. *Microbiology* 2011, **157**:2808-2817.

54. Vollmecke C, Drees SL, Reimann J, Albers SV, Lubben M: **Both ATPases CopA and CopB contribute to copper resistance of the thermoacidophilic archaeon *Sulfolobus solfataricus***. *Microbiology* 2012.

55. Maezato Y, Daugherty A, Dana K, Soo E, Cooper C, Tachdjian S, Kelly RM, Blum P: **VapC6, a ribonucleolytic toxin regulates thermophilicity in the crenarchaeote *Sulfolobus solfataricus***. *RNA* 2011, **17**:1381-1392.

- Gene disruption of the vapBC6 toxin-antitoxin pair resulted in increased sensitivity to heat shock. Evidence was presented that vapC6 regulates gene expression by exhibiting ribonucleolytic activity against specific transcripts.

56. Ellen AF, Rohulya OV, Fusetti F, Wagner M, Albers SV, Driessen AJ: **The sulfolobicin genes of *Sulfolobus acidocaldarius* encode novel antimicrobial proteins**. *J Bacteriol* 2011, **193**:4380-4387.

- Antimicrobial proteins were identified in *Sulfolobus* species which require two proteins for their membrane-vesicle associated activity against closely related species.

57. Albers SV, Jonuscheit M, Dinkelaker S, Urich T, Kletzin A, Tampe R, Driessen AJ, Schleper C: **Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus***. *Appl Environ Microbiol* 2006, **72**:102-111.

58. Jonuscheit M, Martusewitsch E, Stedman KM, Schleper C: **A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector**. *Mol Microbiol* 2003, **48**:1241-1252.

59. Honarbakhsh M, Villafane AA, Ruhl I, Sannino D, Bini E: **Development of a thermostable beta-glucuronidase-based reporter system for monitoring gene expression in hyperthermophiles**. *Biotechnol Bioeng* 2012.

60. Lipscomb GL, Stirrett K, Schut GJ, Yang F, Jenney FE, Jr., Scott RA, Adams MW, Westpheling J: **Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases**. *Appl Environ Microbiol* 2011, **77**:2232-2238.

- A naturally competent mutant was used to generate markerless single- and double-deletion mutants in *P. furiosus* for the first time.

61. Sato T, Fukui T, Atomi H, Imanaka T: **Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1**. *J Bacteriol* 2003, **185**:210-220.

62. Sato T, Fukui T, Atomi H, Imanaka T: **Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis***. *Appl Environ Microbiol* 2005, **71**:3889-3899.

63. Basen M, Sun J, Adams MW: **Engineering a hyperthermophilic archaeon for temperature-dependent product formation**. *MBio* 2012, **3**.

●● *P. furiosus* was engineered to produce a non-native product (lactate) under conditions (low temperature) where cell growth is neglible. This was done by inserting the lactate hydrogenase from *C. bescii* into *P. furiosus* under the control of a *P. furiosus* "cold shock" promoter.

64. Santangelo TJ, Cubonova L, Reeve JN: ***Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology**. *Appl Environ Microbiol* 2010, **76**:1044-1052.

● Genetic tools for *T. kodakaraensis* were expanded to allow for multiple deletions, selection on complex media, and improvement of a reporter gene system based on a beta-glycosidase.

65. Takemasa R, Yokooji Y, Yamatsu A, Atomi H, Imanaka T: ***Thermococcus kodakarensis* as a host for gene expression and protein secretion**. *Appl Environ Microbiol* 2011, **77**:2392-2398.

● *T. kodakaraensis* was engineered to efficiently secrete overexpressed proteins by fusing the signal sequence from a protease to a chitinase and placing expression of the fusion protein under the control of a strong constitutive promoter.

66. Farkas J, Stirrett K, Lipscomb GL, Nixon W, Scott RA, Adams MW, Westpheling J: **The recombinogenic properties of the *Pyrococcus furiosus* COM1 strain enable rapid selection of targeted mutants**. *Appl Environ Microbiol* 2012.

● A method was developed for generation of markerless deletion mutants in *P. furiosus*.

67. Sun J, Hopkins RC, Jenney FE, McTernan PM, Adams MW: **Heterologous expression and maturation of an NADP-dependent [NiFe]-hydrogenase: a key enzyme in biofuel production**. *PLoS One* 2010, **5**:e10526.

●● The challenge of producing significant amounts of oxygen-sensitive *P. furiosus* hydrogenase was overcome by expressing all the structural and maturation proteins necessary under the control of an anaerobic-inducible promoter in *E. coli.* It was found that some of *E. coli* hydrogenase maturation proteins could substitute for the *P. furiosus* versions.

68. Hopkins RC, Sun J, Jenney FE, Jr., Chandrayan SK, McTernan PM, Adams MW: **Homologous expression of a subcomplex of *Pyrococcus furiosus* hydrogenase that interacts with pyruvate ferredoxin oxidoreductase**. *PLoS One* 2011, **6**:e26569.

69. Bridger SL, Clarkson SM, Stirrett K, DeBarry MB, Lipscomb GL, Schut GJ, Westpheling J, Scott RA, Adams MW: **Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus***. *J Bacteriol* 2011, **193**:6498-6504.

70. Kanai T, Matsuoka R, Beppu H, Nakajima A, Okada Y, Atomi H, Imanaka T: **Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon *Thermococcus kodakarensis***. *J Bacteriol* 2011, **193**:3109-3116.

71. Santangelo TJ, Cubonova L, Reeve JN: **Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production**. *Mol Microbiol* 2011, **81**:897-911.

● Targeted gene disruptions in *T. kodakaraensis* demonstrated that the membrane-bound hydrogenase (MBH) is primarily responsible for hydrogen production, the SurR regulator is required for MBH expression, and specific hydrogen production could be increased by removing the cytosolic hydrogenases (which primarily consume hydrogen).

72. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA: **Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase**. *Science* 1988, **239**:487-491.

73. Atomi H, Sato T, Kanai T: **Application of hyperthermophiles and their enzymes**. *Curr Opin Biotechnol* 2011, **22**:618-626.

●● Review of hyperthermophiles as sources of thermostable enzymes and potential hosts for high-temperature cell engineering and biofuel production.

74. Cobucci-Ponzano B, Perugino G, Rossi M, Moracci M: **Engineering the stability and the activity of a glycoside hydrolase**. *Protein Eng Des Sel* 2011, **24**:21-26.

75. Campbell E, Wheeldon IR, Banta S: **Broadening the cofactor specificity of a thermostable alcohol dehydrogenase using rational protein design introduces novel kinetic transient behavior**. *Biotechnol Bioeng* 2010, **107**:763-774.

● Analysis of mesophilic and thermophilic hydrogenases was used to guide directed mutations of *P. furiosus* alcohol dehydrogenase, resulting in improved activity, broadened cofactor specifity, and a cofactor concentration-dependent binding mechanism.

76. Gera N, Hussain M, Wright RC, Rao BM: **Highly stable binding proteins derived from the hyperthermophilic Sso7d scaffold**. *J Mol Biol* 2011, **409**:601-616.

● Random mutagenesis of a small hyperthermophilic protein was used to generate stable, high affinity, highly specific binding proteins for a wide variety of model targets.

77. Young DD, Nichols J, Kelly RM, Deiters A: **Microwave activation of enzymatic catalysis**. *J Am Chem Soc* 2008, **130**:10048-10049.

78. Cowan DA, Fernandez-Lafuente R: **Enhancing the functional properties of thermophilic enzymes by chemical modification and immobilization**. *Enzyme Microb Technol* 2011, **49**:326-346.

79. Harris JM, Epting KL, Kelly RM: **N-terminal fusion of a hyperthermophilic chitin-binding domain to xylose isomerase from *Thermotoga neapolitana* enhances kinetics and thermostability of both free and immobilized enzymes**. *Biotechnol Prog* 2010, **26**:993-1000.

80. Batista PR, Costa MG, Pascutti PG, Bisch PM, de Souza W: **High temperatures enhance cooperative motions between CBM and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics**. *Phys Chem Chem Phys* 2011, **13**:13709-13720.

● According to computational simulations, the substrate binding and catalytic domains of a thermostable cellulase were observed to move cooperatively at high temperatures, and two key “hinge” amino acid residues were identified.

81. Su X, Mackie RI, Cann IK: **Biochemical and mutational analyses of a multidomain cellulase/mannanase from *Caldicellulosiruptor bescii***. *Appl Environ Microbiol* 2012, **78**:2230-2240.

82. Ozdemir I, Blumer-Schuette SE, Kelly RM: **S-layer homology domain proteins Csac_0678 and Csac_2722 are implicated in plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus***. *Appl Environ Microbiol* 2012, **78**:768-777.

83. Christen P, Davidson S, Combet-Blanc Y, Auria R: **Phenol biodegradation by the thermoacidophilic archaeon *Sulfolobus solfataricus* 98/2 in a fed-batch bioreactor**. *Biodegradation* 2011, **22**:475-484.

84. Zhu W, Xia JL, Yang Y, Nie ZY, Zheng L, Ma CY, Zhang RY, Peng AA, Tang L, Qiu GZ: **Sulfur oxidation activities of pure and mixed thermophiles and sulfur speciation in bioleaching of chalcopyrite**. *Bioresour Technol* 2011, **102**:3877-3882.

85. Lai D, Lluncor B, Schroder I, Gunsalus RP, Liao JC, Monbouquette HG: **Reconstruction of the archaeal isoprenoid ether lipid biosynthesis pathway in *Escherichia coli* through digeranylgeranylglyceryl phosphate**. *Metab Eng* 2009, **11**:184-191.

86. Santa-Maria MC, Yencho CG, Haigler CH, Thompson WF, Kelly RM, Sosinski B: **Starch self-processing in transgenic sweet potato roots expressing a hyperthermophilic alpha-amylase**. *Biotechnol Prog* 2011, **27**:351-359.

● A hyperthermophilic alpha-amylase was inserted into sweet potato without affecting growth and development of the plant so that the plant could be "self-processed" by incubating at an elevated temperature that activates the thermostable enzyme.

87. Koerdt A, Jachlewski S, Ghosh A, Wingender J, Siebers B, Albers SV: **Complementation of *Sulfolobus solfataricus* PBL2025 with an alpha-mannosidase: effects on surface attachment and biofilm formation**. *Extremophiles* 2012, **16**:115-125.

88. Hawkins AS, Han Y, Lian H, Loder AJ, Menon AL, Iwuchukwu IJ, Keller M, Leuko TT, Adams MW, Kelly RM: **Extremely thermophilic routes to microbial electrofuels**. *ACS Catalysis* 2011, **1**:1043-1050.

89. Reysenbach AL, Liu Y, Banta AB, Beveridge TJ, Kirshtein JD, Schouten S, Tivey MK, Von Damm KL, Voytek MA: **A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents**. *Nature* 2006, **442**:444-447.

90. Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO: **A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont**. *Nature* 2002, **417**:63-67.

91. Bellack A, Huber H, Rachel R, Wanner G, Wirth R: ***Methanocaldococcus villosus* sp. nov., a heavily flagellated archaeon that adheres to surfaces and forms cell-cell contacts**. *Int J Syst Evol Microbiol* 2011, **61**:1239-1245.

92. Slobodkina GB, Kolganova TV, Querellou J, Bonch-Osmolovskaya EA, Slobodkin AI: ***Geoglobus acetivorans* sp. nov., an iron(III)-reducing archaeon from a deep-sea hydrothermal vent**. *Int J Syst Evol Microbiol* 2009, **59**:2880-2883.

93. Osburn MR, Amend JP: ***Thermogladius shockii* gen. nov., sp. nov., a hyperthermophilic crenarchaeote from Yellowstone National Park, USA**. *Arch Microbiol* 2011, **193**:45-52.

94. Kublanov IV, Bidjieva S, Mardanov AV, Bonch-Osmolovskaya EA: ***Desulfurococcus kamchatkensis* sp. nov., a novel hyperthermophilic protein-degrading archaeon isolated from a Kamchatka hot spring**. *Int J Syst Evol Microbiol* 2009, **59**:1743-1747.

Table 2.1. Model Extreme Thermophiles						
Organism	**Isolation site**	**T_{opt} (°C)**	**Domain**	**Genome size (bp)**	**Growth physiology**	**PubMed Citations** (May, 2012)
Sulfolobus solfataricus P2	solfatara, Naples, Italy	80°C	Crenarchaea	2,992,245	Aerobic, extreme thermoacidophile	1252
Thermotoga maritima MSB8	shallow marine sediments, Vulcano Island, Italy	80°C	Bacteria	1,860,725	Fermentative anaerobe, facultative S° reducer	1148
Thermococcus kodakaerensis KOD1	solfatara, Kodakara Island, Japan	85°C	Euryarchaea	2,088,737	Fermentative anaerobe, facultative S° reducer	156
Pyrococcus furiosus	shallow marine sediments, Vulcano Island, Italy	98°C	Euryarchaea	1,908,255	Fermentative anaerobe, facultative S° reducer	1047

Table 2.2. Recently Described Extreme Thermophiles

Organism	T_{opt} (°C)	16S rRNA identity to closest relative	Isolation site	Metabolism	Notes	Reference
Aciduliprofundum boonei	70	83	Deep sea hydrothermal vent	Anaerobic fermentation	First obligate thermoacidophile from deep sea vents	[89]
Nanoarchaeum equitans	70-98	81	Submarine hydrothermal vent	Parasitism	Small (0.5 Mb) genome, parasite of *Ignicoccus hospitalis*	[90]
Methanocaldococcus villosus	80	95	Submarine hydrothermal vent	Chemolithoautotrophy	Unique striated cell surface pattern observed	[91]
Geoglobus acetivorans	81	97	Deep sea hydrothermal vent	Chemolithoautotrophy	1st hyperthermophile enriched on acetate as e^- donor	[92]
Thermogladius shockii	84	96	Hot spring	Anaerobic fermentation	Unlike close relatives, growth unaffected by sulfur	[93]
Desulfurococcus kamchatkensis	85	98.1	Hot spring	Anaerobic fermentation	Can use keratin as sole carbon and energy source	[94]
Pyrococcus yayanosii	98	99.4	Deep sea hydrothermal vent	Anaerobic fermentation	Obligate piezophile	[22]
Acidilobus saccharovorans	80-85	98.1	Hot spring	Anaerobic fermentation	Unlike *A. aceticus*, can use mono- and di-saccharides	[21]

Table 2.3. Considerations for Metabolic Engineering Bioprocesses at High Temperatures	
Advantages	**Disadvantages**
Reduced recalcitrance of plant biomass for biofuels applications	Possible energy burden of heating reactor contents
Reduced risk of contamination	Lower yields of cellular biomass
Use of temperature regulation to optimize product formation	Genetic stability of thermophilic recombinant hosts unknown
Facilitated recovery of volatile products	Lower gas solubilities
Lower risk of release of viable genetically modified organisms	Substrate, product lability at elevated temperatures
Higher temperatures more consistent with chemical processes	Limited to genes encoding themostable proteins/enzymes
Higher mass transfer rates	Less known about microbial physiology
Improved solubility of carbohydrates, amino acids	Genetics tools are in infancy

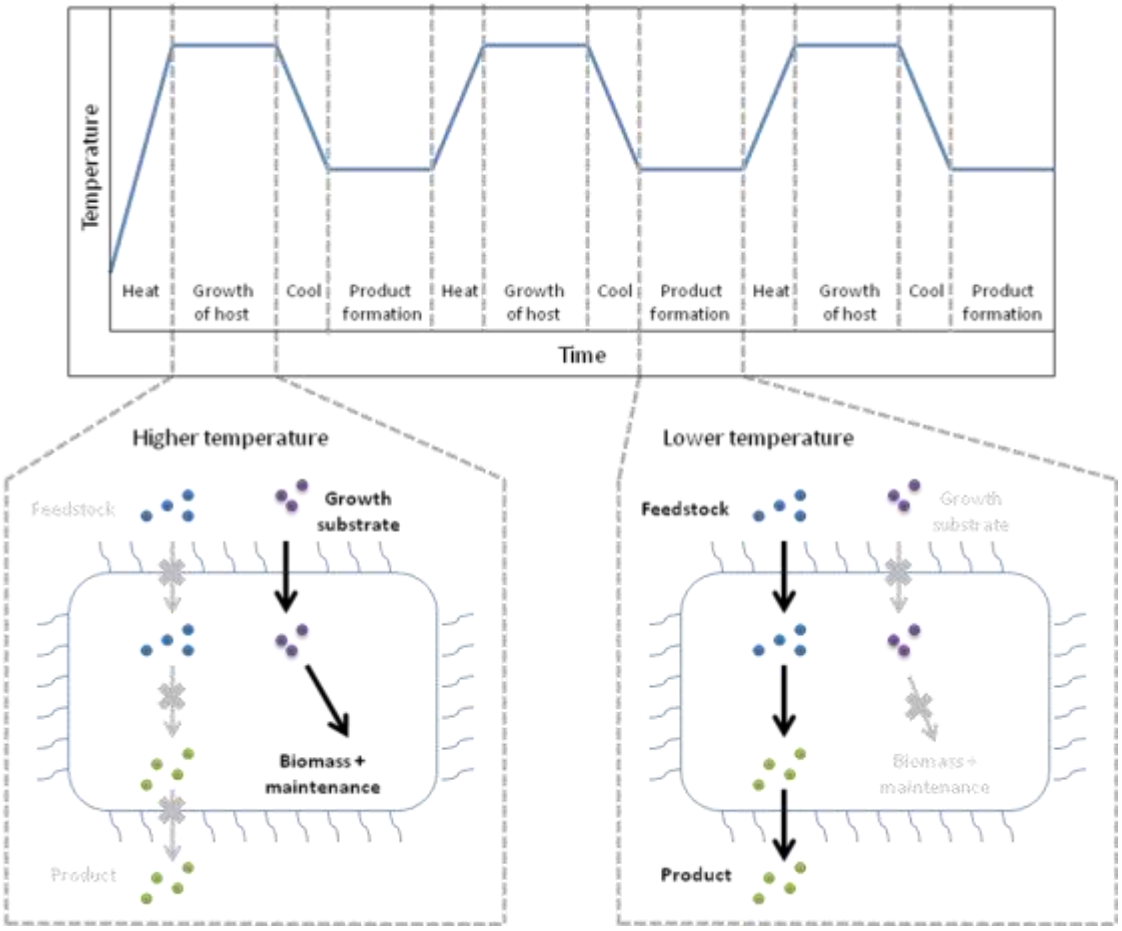

Figure 2.1. Temperature-dependent regulation of product formation. Two substrates are provided initially: one that can be catabolized by the extremely thermophilic host's endogenous metabolism for growth, another that cannot and is used for product formation. At higher temperature, the host grows but product formation is silenced. After sufficient cell generation has occurred, the temperature is lowered, inhibiting the host's metabolism and inducing a cold shock promoter that controls expression of heterologous enzymes active at the lower temperature. The heterologous enzymes catalyze all reactions necessary to generate the desired product from the substrate that is provided specifically for product formation. When product formation deteriorates, the temperature can be raised to rehabilitate the extremely thermophilic host and regenerate cofactors and intermediates to prepare for another round of product formation. Additional growth substrate can be provided

if necessary, or the extreme thermophile may be able to subsist by catabolizing the heterologous enzymes.

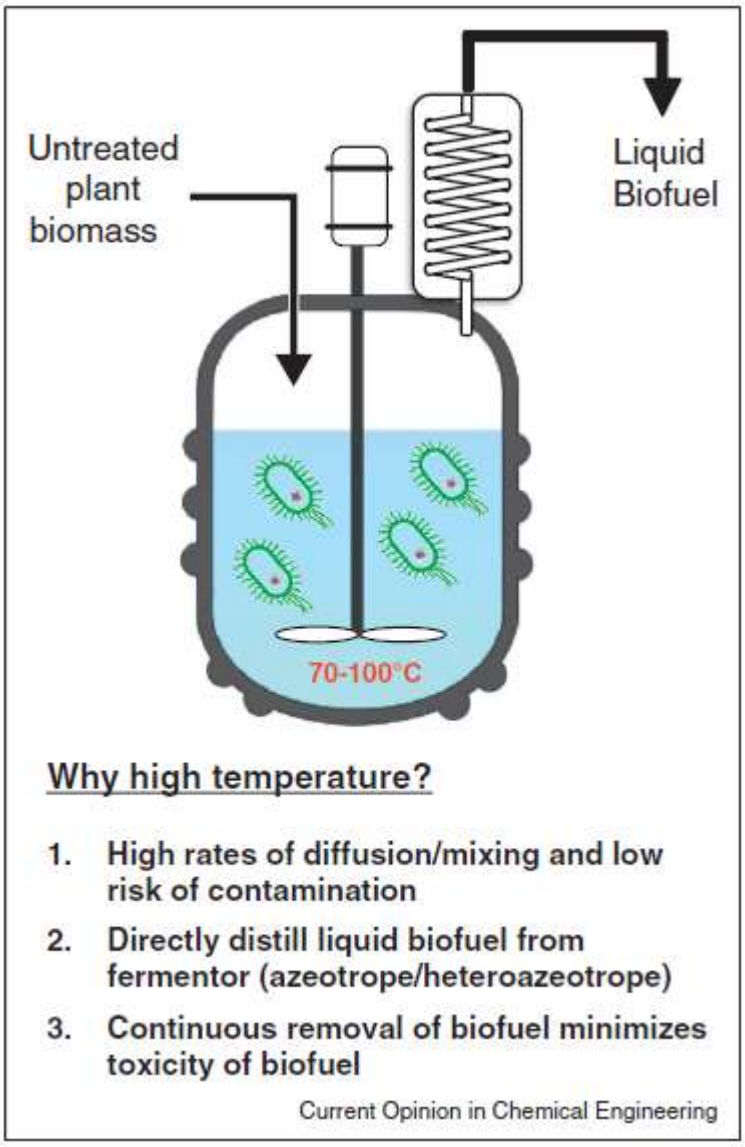

Figure 2.2. Consolidated bioprocessing at high temperatures for biofuels production. Metabolic engineering tools for extreme thermophiles could be used to create trains capable of deconstruction of plant biomass and conversion to volatile liquid biofuels that can be recovered directly by distillation from fermentation broths.

Chapter 3

Hyperthermophilic *Thermotoga* species differ with respect to specific carbohydrate transporters and glycoside hydrolases

Andrew D. Frock, Steven R. Gray, and Robert M. Kelly

Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

Published in *Applied and Environmental Microbiology* 78(6): 1978-1986 (2012)

ABSTRACT

Four hyperthermophilic members of the bacterial genus *Thermotoga* (*T. maritima*, *T. neapolitana*, *T. petrophila*, and *T.* sp. strain RQ2) share a core genome of 1,470 open reading frames (ORFs), or about 75% of their genomes. Nonetheless, each species exhibited certain distinguishing features during growth on simple and complex carbohydrates that correlated with genomic inventories of specific ABC sugar transporters and glycoside hydrolases. These differences were consistent with transcriptomic analysis based on a multi-species cDNA microarray. Growth on a mixture of six pentoses and hexoses showed no significant utilization of galactose or mannose by any of the four species. *T. maritima* and *T. neapolitana* exhibited similar monosaccharide utilization profiles, with a strong preference for glucose and xylose over fructose and arabinose. *T.* sp. strain RQ2 also used glucose and xylose, but was the only species to utilize fructose to any extent, consistent with a phosphotransferase system (PTS) specific for this sugar encoded in its genome. *T. petrophila* used glucose to a significantly lesser extent than the other species. In fact, the XylR regulon was triggered by growth on glucose for *T. petrophila*, which was attributed to the absence of a glucose transporter (XylE2F2K2), otherwise present in the other *Thermotoga* species. This suggested that *T. petrophila* acquires glucose through the XylE1F1K1 transporter, which primarily serves to transport xylose in the other three species. The results here show that subtle differences exist among the hyperthermophilic *Thermotoga* species with respect to carbohydrate utilization, which supports their designation as separate species.

INTRODUCTION

Metabolic engineering efforts to develop microorganisms capable of producing biofuels through the fermentation of sugars have largely focused on manipulating central and intermediary metabolism to direct electrons to the desired bioenergy product (1, 54, 55). In most cases, this presumes that the carbon source is homogenous and readily utilized by the microorganism. However, if lignocellulosic feedstocks are to be processed to biofuels efficiently, microbial systems capable of simultaneously processing mixtures of simple and complex sugars are highly desirable since such systems could circumvent or minimize biomass pretreatment requirements. The strategies used by microorganisms capable of utilizing heterogeneous mixtures of complex carbohydrates rely heavily on the regulation of glycoside hydrolase inventories and the presence of ATP-binding cassette (ABC) transporters to move sugars across cell boundaries (51, 52). Understanding these strategies can ultimately lead to metabolically engineered microorganisms that coordinate the uptake and processing of carbohydrates with the synthesis of biofuel molecules.

Members of the genus *Thermotoga* have been isolated from locations throughout the world, including Italy, France, The Netherlands, the Azores, Japan, and Africa (4, 12, 19, 21, 25, 28, 39, 53). These bacteria possess a large number of glycoside hydrolases and ABC transporters relative to their genome size (7, 41), which allow them to ferment a wide variety of simple and complex carbohydrates, forming hydrogen (or hydrogen sulfide), acetate, and CO_2 as the primary metabolic products (10). Polysaccharides that can support growth of *Thermotoga* include xylan, laminarin, carboxymethyl cellulose, starch, glucomannan, galactomannan, and chitin (7, 11). However, *Thermotoga* species lack multifunctional enzymes with both endoglucanase and carbohydrate binding module (CBM) domains, at

least partially explaining their inability to grow on crystalline cellulose (3). Nevertheless, *Thermotoga* strains have been extensively investigated as sources of hyperthermophilic mannanases, xylanases, and amylases (2, 40, 45). These bacteria are also noteworthy for high-yield hydrogen production (3 to 4 mol/mol glucose), which may be related to the synergistic utilization of NADH and reduced ferredoxin by a bifurcating hydrogenase (9, 46).

Regulatory networks related to carbohydrate utilization in *Thermotoga* have been predicted by transcriptomics and comparative genomics (11, 38) and, in some cases, validated by characterization of transcriptional regulators (D. Rodionov, personal communication). Transcription of genes in a particular regulon is generally repressed by the transcriptional regulator and induced by binding of the regulator to an effector (often a mono- or di-saccharide). For example, genes in the CelR regulon are involved in catabolism and transport of β-linked glucans and glucomannan. Binding of the CelR regulator (TM1218) to the predicted effector (cellobiose) induces transcription of two endoglucanases (TM1524 and TM1525), a mannobiose transporter (TM1219 to TM1223), a transporter of oligosaccharides derived from xyloglucans (TM0300 to TM0304), and a cellobiose phosphorylase (TM1848) (33, 38). The BglR regulon is also induced by cellobiose and was found to be upregulated during growth on various β-linked glucans, including laminarin, barley glucan, and pustulan (11). This regulon's single operon includes the regulator (TM0032), a laminarinase (TM0024), a β-glucosidase (TM0025), and a cellobiose/ laminaribiose transporter (TM0027 to TM0031) (33, 38).

The XylR regulon is composed of xylan utilization genes and is organized into 7 or 8 operons, depending on the *Thermotoga* species. Binding of the XylR repressor (TM0110) to DNA is disrupted by the inducers xylose and glucose *in vitro* (D. Rodionov, personal communication). This large regulon includes two xylanases (TM0061 and TM0070), a

xylobiose/xylotriose transporter (TM0071 to TM0075), a transporter of unknown xylan degradation products (TM0056 to TM0060), the xylose/glucose transporter XylE1F1K1 (TM0112, TM0114, and TM0115), a β-xylosidase (TM0076), a xylose isomerase (TM1667), and a xylulokinase (TM0116) (11, 33, 38, 50).

Another regulon of interest here (GluR) consists of two transporters organized in one operon that were, until recently, uncharacterized. This operon shows variability from one species to another, as *T. petrophila* is lacking one of the two transporters and *T.* sp. strain RQ2 is missing an α-amylase. Nearly the entire operon is missing from the *T. maritima* strain used to generate the whole-genome sequence published in 1999 (34), but a recent study has shown that the strain of *T. maritima* deposited in DSMZ has all of these genes, except for the gene coding for α-amylase (5). The transporters in this operon from *T. maritima* were characterized; TreE was shown to bind trehalose, sucrose, and glucose, while XylE2 was shown to bind xylose, glucose, and fucose (5). The regulator encoded in TM1847 is predicted to bind glucose (38).

Given the wide range of geographic locations inhabited by *Thermotoga* species, it is interesting to consider the underlying factors that make them so adaptive as well as what, if anything, differentiates one species from another with respect to growth physiology on carbohydrates. Genome sequences are available for several hyperthermophilic *Thermotoga* species, including *T. maritima* (34), *T. petrophila* (56), *T. neapolitana*, and *T.* sp. strain RQ2 (GenBank accession no. AE000512, CP000702, CP000916, and CP000969, respectively). An understanding of the physiological characteristics encoded in the “core” genome, as well as the features that are associated with the “non-core” open reading frames (ORFs), could provide new insights into the intricacies of microbial carbohydrate utilization strategies. To this end, growth of four hyperthermophilic *Thermotoga* species on glucose, a mixture of six

monosaccharides, and a polysaccharide mixture was examined to probe the relationship between genotype and phenotype. Even among these closely related hyperthermophiles, sugar utilization patterns and corresponding transcriptomes revealed distinguishing features that could be important in selecting a platform for biofuel production.

MATERIALS AND METHODS

Design of the multi-species cDNA microarray. A *Thermotoga* multi-species cDNA microarray was developed by using a combination of previously constructed cDNA probes from *T. maritima* (7, 23), along with new probes representing ORFs specific to three other species (*T. neapolitana*, *T.* sp. strain RQ2, and *T. petrophila*). At the time that the multi-species array was constructed, a complete genome sequence was available for *T. petrophila*, a draft genome sequence was available for *T. neapolitana*, and suppressive subtractive hydbridization had been used to compare *T. maritima* and *T.* sp. strain RQ2 (36). Due to the incomplete genome information that was available for *T.* sp. strain RQ2 at that time, in some cases, there are multiple probes for the same gene. Note that the *T.* sp. strain RQ2 genome is now available (GenBank accession no. CP000969).

The genomes of *T. neapolitana* and *T. petrophila* were compared to that of *T. maritima* using GenomeBlast (27). ORFs with amino acid sequences sequences that were not 70% identical with 80% coverage compared to the other three species were considered to be unique to that bacterium. Vector NTI Advance 10 (Invitrogen, Carlsbad, CA) was used to design probes for these unique ORFs. Probes were produced by PCR using genomic DNA as the template and primers obtained from Integrated DNA Technologies (Coralville, IA). PCR products were spotted onto UltraGAPS slides (Corning), as previously described

(7). A few of the smaller probes were ordered directly as oligonucleotides from Integrated DNA Technologies (Coralville, IA).

Growth of *Thermotoga* species on simple and complex carbohydrates. Cell densities were determined by epifluorescence microscopy using acridine orange stain, as described previously (48). For monosaccharide utilization experiments, the four species were grown on modified sea salts medium to minimize the effect on the refractive index baseline during high-performance liquid chromatography (HPLC) analysis (with the concentration of sea salts lowered to 10 g/liter and resazurin omitted) (17). The monosaccharide mixture (containing glucose, xylose, arabinose, mannose, galactose, and fructose) was prepared in solution, sterilized by filtration, and added to autoclaved medium at a final concentration of 0.25 g/L of each sugar. Samples taken were passed through a 0.22-μm-pore filter. Monosaccharide concentrations were determined by HPLC with a Shodex SP0810 column at 80°C (52). Water was used as the mobile phase at a flow rate of 0.2 mL/min.

For transcriptional analysis experiments, four *Thermotoga* species (*T. maritima* MSB8, *T. neapolitana* DSM 4359, *T. petrophila* RKU-1, and *T.* sp. strain RQ2) were each grown in pure culture at 80°C on artificial seawater (ASW) medium (6), supplemented with either glucose (0.25% [wt/vol]) (Sigma-Aldrich, St. Louis, MO) or a polysaccharide mixture (0.083% [wt/vol]) consisting of equal parts by mass of lichenan (~85% purity) (Megazyme, Wicklow, Ireland), konjac glucomannan (90% purity) (Jarrow Formulas, Los Angeles, CA), pectin (~85% purity) (Sigma-Aldrich), galactomannan (95% purity) (Sigma-Aldrich), carboxymethyl cellulose (~99% purity) (Sigma-Aldrich), xylan from oat spelts (90% purity) (Sigma-Aldrich), and xylan from birchwood (90% purity) (Sigma-Aldrich). Polysaccharides were added to autoclaved medium without sterilization. Prior to extracting total RNA,

cultures were passed at least six times on either glucose or the polysaccharide mixture, using a 1% (vol/vol) inoculum. Polysaccharide- or glucose-grown cultures were harvested during mid-exponential phase (3-7 × 10^7 cells/mL), quickly cooled in a dry ice-ethanol bath, and cells were collected by centrifugation. Cultures grown on polysaccharides were passed through a coffee filter prior to centrifugation. RNA extraction, reverse transcription, labeling, hybridization, and slide scanning were performed with modifications of previously described protocols (7, 23). Loess normalization, analysis of variance (ANOVA) normalization, and ANOVA row-by-row modeling were performed using JMP Genomics (SAS Institute, Cary, NC). Fold changes at or above 2-fold with significant values at or above the Bonferroni correction (typically equivalent to a *P* value of approximately 10^{-5}) were considered to be significant differential transcription.

Microarray data accession numbers. Full results, additional experimental details, and more information about the multispecies array have been deposited in NCBI's Gene Expression Omnibus (13) and are accessible through GEO platform accession no. GPL13655 and GEO series accession no. GSE29557.

RESULTS AND DISCUSSION

Genomic comparison of *Thermotoga* species. The 10 *Thermotoga* species currently identified can clearly be divided into one group with optimum growth temperatures of 65 to 70°C and another group with optimum growth temperatures of 75 to 80°C (Table 3.1). 16S rRNA sequences indicate that the isolates with higher optimum growth temperatures are more closely related (≥99% identical) to each other than to the less thermophilic *Thermotoga*

strains (15). Complete genome sequences are available for seven *Thermotoga* species, and all of these, except for *T. lettingae* and *T. thermarum*, belong to the group of higher-temperature species (Table 3.1). Only the 16S rRNA sequence was available for *T. naphthophila* when this study was initiated, so it was not included; note that *T. petrophila*, which is closely related to *T. naphthophila*, was included (35). To probe the significance of differences among these species, the genomes of *T. maritima*, *T. neapolitana*, *T.* sp. strain RQ2, and *T. petrophila* were compared on an ORF-by-ORF basis. Based on 70% amino acid sequence identity over 80% of the ORF, a core genome of approximately 1,470 ORFs (Fig. 3.1) was defined for the four hyperthermophilic *Thermotoga* species. In general, genes involved in central metabolism are highly conserved. All four species have complete glycolytic, pentose phosphate, and Entner-Doudoroff pathways. *T. maritima* and *T. neapolitana* have an extra transketolase that is not present in the other two species, but the physiological role and phenotypic implications of this enzyme are unknown.

These four species have notable genomic differences with respect to carbohydrate utilization and transport (Table 3.2). Non-core genes include an arabinose utilization locus, present only in *T. petrophila* and *T.* sp. strain RQ2 (TRQ2_0657 to TRQ2_0667), which consists of a predicted arabinose transporter (Table 3.2) and four enzymes that are apparently targeted to the secretome. Mannan-derived oligosaccharides are likely transported into the cell by the products of TM1746 to TM1750, before being further hydrolyzed by two intracellular glucomannanases (TM1751 and TM1752) (8), which are only present in *T. maritima* and *T.* sp. strain RQ2. A recently characterized glucose/xylose transporter (5) is missing from *T. petrophila* (Table 3.2), while a nearby α-amylase (CTN_0781) is missing from *T. maritima* and *T.* sp. strain RQ2. There are two relatively large loci which are not well characterized and are unique to *T. neapolitana*. The first

(CTN_0355 to CTN_0373) includes a glycoside hydrolase family 31 (GH31) enzyme (CTN_0355) and two putative monosaccharide transporters (Table 3.2), while the second (CTN_1540 to CTN_1555) includes another GH31 enzyme and an uncharacterized transporter (Table 3.2). The question arises as to whether these variations in glycoside hydrolases and ABC transporters relate to carbohydrate utilization patterns.

Utilization of a mixture of monosaccharides. The existence of monosaccharide transporters that are not conserved among all four species (Table 3.2) suggests that these species may show different sugar preferences. To test this, all four species were grown on equal amounts of glucose, galactose, mannose, fructose, arabinose, and xylose. Consumption of the individual monosaccharides was monitored by HPLC. *T. maritima* and *T. neapolitana* have very similar monosaccharide utilization patterns (Fig. 3.2A and B), and there are no well-annotated monosaccharide transporters that differ between them (Table 3.2). Both species clearly co-utilize glucose and xylose, while utilizing arabinose to a lesser extent. *T.* sp. strain RQ2 (Fig. 2D) has a similar pattern, except that fructose is consumed almost as quickly as glucose and xylose. This is most certainly due to the fructose-specific phosphotransferase system (PTS), which is present in *T.* sp. strain RQ2 but missing from the other three species (Table 3.2). All six genes of the PTS locus were significantly up-regulated on fructose compared to glucose in *T.* sp. strain RQ2 (A. D. Frock and R. M. Kelly, unpublished results), supporting the annotation of this PTS as fructose specific. Similar results were seen for a homologous PTS in *Caldicellulosiruptor saccharolyticus* (52). Although *T. petrophila* was initially reported to be unable to grow on xylose (49), in this experiment, xylose was depleted faster than the other monosaccharides (Fig. 3.2C). Comparison of *T. petrophila*'s utilization pattern to the other species reveals that *T.*

petrophila shows less of a preference for glucose, which can be traced to the absence of a transporter that will be discussed in further detail below.

Design of the *Thermotoga* multi-species cDNA microarray. Transcriptomic data for the four *Thermotoga* species were obtained to provide further insights into metabolic differences arising from carbohydrate processing. Previously, a whole-genome cDNA microarray for *T. maritima* was used to investigate a range of physiological and ecological issues pertaining to the growth of this bacterium (7, 10, 11, 22–24, 29, 30, 42, 43, 47). The high level of sequence identity between these four species suggested that common microarray probes could be used, expanding the transcriptomics analysis to several *Thermotoga* species. To include those non-core genes absent from the original *T. maritima* array, probes were added based on sequences from *T. neapolitana* (304 probes), *T.* sp. strain RQ2 (284 probes), and *T. petrophila* (60 probes).

Monosaccharide and polysaccharide transcriptomes for *Thermotoga* species. To compare the genes expressed for carbohydrate utilization in these species, transcriptomes for each species during growth on a mixture of polysaccharides (including glucan, mannan, xylan, and pectin) were compared to that during growth on a single monosaccharide (glucose). As discussed above, these species possess a significant number of carbohydrate active enzymes and transporters that are not conserved among all four species. However, few of these genes were differentially transcribed on the polysaccharide mix compared to glucose. The most significant differential transcription of non-core genes involved a mannan utilization locus (TM1746 to TM1752) (8) and a mannooligosaccharide binding protein (Table 3.2) (33), which are present only in *T. maritima* and *T.* sp. strain RQ2. For these two

species, these genes were up-regulated on the polysaccharide mix compared to glucose (Table 3.3). It is not known how the absence of these genes affects mannan utilization by *T. neapolitana* and *T. petrophila*. All four species do have an extracellular endo-1,4-β-mannanase (TM1227) and a mannobiose transporter (TM1219 to TM1223) (33), which may be sufficient for mannan consumption.

Assuming that conserved genes have similar functions in all four species, these genes would be expected to exhibit similar transcriptional responses. This is true in some cases, such as the genes regulated by CelR (Table 3.3), which are involved in utilizing β-glucans and mannans (11). Genes in this regulon include the previously mentioned mannobiose transporter (TM1219 to TM1223), intracellular and extracellular endoglucanases (TM1524 and TM1525), and a cellobiose phosphorylase (TM1848) with a broad substrate range *in vitro* (44). TM1223 and TM1848 were up-regulated significantly in all four species on the polysaccharide mix compared to glucose. The only genes under the control of CelR that were not upregulated on the polysaccharide mix (TM0299 to TM0307) are putatively co-regulated by GloR (38), suggesting that under these conditions their expression may be repressed by this poorly studied regulator.

In other cases, transcriptional responses were less conserved among the four species. All six genes in an operon (TM0322 to TM0327) that includes a tripartite ATP-independent periplasmic (TRAP) transporter were significantly up-regulated on the polysaccharide mix in *T.* sp. strain RQ2 (Table 3.3). Two of these genes were up-regulated in *T. neapolitana*, while none were in *T. petrophila* or *T. maritima*. This locus was not observed to be responsive in *T. maritima* grown on 10 different carbon sources (11). TRAP transporters have primarily been shown to bind organic acids (14), so this transporter's role in polysaccharide utilization is unclear. Other genes with different responses in at least one

species include those regulated by BglR and those involved in processing α-glucans (Table 3.3). BglR-regulated genes were generally more highly transcribed during growth on glucose, except for in *T. neapolitana*, where only one gene was differentially regulated (TM0025), and it was more highly transcribed on the polysaccharide mix. Genes listed in Table 3 involved in processing α-glucans were generally up-regulated on glucose, except in the case of *T. neapolitana*, where one gene (TM0752) was upregulated on the polysaccharide mix.

Glucose transport. Transcriptomes for growth on glucose and the polysaccharide mix facilitated the identification of a glucose binding protein not present in the initial *T. maritima* genome sequence. Although *Thermotoga* species have been known to grow on glucose since the initial isolation of *T. maritima* (21), the corresponding genes responsible for glucose uptake were not identified. Glucose transport in *T. neapolitana* was previously shown to be dependent on ATP and sodium ions (16), and glucose binding activity was detected in cell extracts of *T. maritima* (31), but subsequent characterization of 15 binding proteins failed to identify one with a high specificity for glucose (33). XylE1 (the product of TM0114) is present in all four species (Table 3.2) and was shown to bind glucose (50), but this protein also binds xylose with high affinity (K_d [dissociation constant], ~0.01 μM) (33). Furthermore, TM0114 was more highly transcribed in *T. maritima* during chemostat growth on xylose than glucose (C.-J. Chou and R. M. Kelly, unpublished results). Therefore, TM0114's involvement in *in vivo* glucose transport is unclear.

Here, four genes encoding sugar binding proteins were upregulated on glucose compared to the polysaccharide mix (Tables 3.3 and 3.4): TM1839, TM0031, TRQ2_0973, and TRQ2_0970 (note that TRQ2_0973 and TRQ2_0970 are homologous to CTN_0777

and CTN_0780, respectively). TM0031 was shown to bind cellobiose and laminaribiose (33), and TM1839 was shown to bind maltose, maltotriose, and trehalose (32). Both of these genes are present in all four species (Table 3.2). Homologs of TRQ2_0970 and TRQ2_0973 are not present in the published *T. maritima* genome (34), yet there was significant hybridization of *T. maritima* cDNA to the probes for these genes. Components of both transporters were significantly up-regulated in *T. maritima* grown on glucose (Table 4). The microarray probes significantly up-regulated in *T. maritima* grown on glucose correspond to a total of 2,658 bp ranging over 5,032 of the 8,862 bp that are present in *T.* sp. strain RQ2 but absent from the published genome for *T. maritima*, strongly suggesting that this entire region is actually present in the *T. maritima* strain used in this study. In fact, it was recently reported that the *T. maritima* isolate deposited in DSMZ (used in this study) contains 8,870 bp of genomic DNA that is missing from the isolate sequenced by TIGR in 1999 (5, 34). The product of the TRQ2_0970 homolog (*treE*) was found to bind trehalose, sucrose, and glucose, while the product of the TRQ2_0973 homolog (*xylE2*) was found to bind xylose, glucose, and fucose (5). Past studies that focused on *T. maritima* (33) failed to identify *xylE2* because the sequence was missing from the sequenced strain. Noll et al. (37) predicted this genomic variation by examining gene synteny in related species and a β-glucosidase sequence in this region that was published prior to the complete genome (26). In the absence of these hints, genomic variations may not be discovered until the genome is resequenced or other analyses (e.g., RNA-seq) are used. The genomic variation was revealed by the multi-species microarray developed for this study, demonstrating the utility of microarrays in detecting mutations.

Regulation of xylan utilization genes. The most striking difference observed between these species involved genes under the control of the regulator XylR (Table 3.5) that are present in all four species. As described above, this regulon includes all genes necessary for hydrolyzing xylan, transporting degradation products into the cell, and assimilating the resulting sugars into the pentose phosphate pathway. One would expect that these genes would be upregulated on the polysaccharide mix, which included xylan. For three of these species, this was the case (Table 3.5), but in *T. petrophila*, these genes were surprisingly downregulated on the polysaccharide mix compared to glucose.

One possible explanation for this unique regulation pattern is that *T. petrophila* is simply not using the xylan component of the polysaccharide mix. Initially, *T. petrophila* was reported to be unable to use xylose (49), but the cell density reached when xylose or xylan was supplemented at 2.5 g/L was approximately 3-fold higher than that on ASW base medium (data not shown), suggesting that *T. petrophila* is fully capable of growing on these substrates. Also, if *T. petrophila* was simply not using xylan, one would expect transcript levels for these genes to be similar for both conditions, not higher on glucose. A second possible explanation for the down-regulation of the XylR regulon on the polysaccharide mix in *T. petrophila* is that these genes are repressed by one of the other carbohydrates in the polysaccharide mix. To investigate this possibility, *T. petrophila* was grown with xylan as the sole carbon source, and transcript levels were compared to glucose. Again, most XylR genes were significantly down-regulated on xylan compared to glucose, while no XylR genes were significantly up-regulated (data not shown), suggesting that the other components of the polysaccharide mix did not repress the transcription of XylR genes.

Another possible explanation for *T. petrophila*'s regulation of XylR genes is that these genes were actually induced by glucose. As previously mentioned, in *T. maritima*, *T.*

sp. strain RQ2, and *T. neapolitana*, XylE2 is the only sugar binding protein whose transcript was up-regulated on glucose compared to the polysaccharide mix that has also been shown to bind glucose (5). However, the transporter is missing from *T. petrophila* (Table 3.2), which causes this species to show less of a preference for glucose (Fig. 3.2C), as discussed above. However, *T. petrophila* does utilize glucose (Fig. 3.2C), so this species must bring glucose into the cell by some other means. The only other protein shown to bind glucose in these bacteria, XylE1 (encoded by TM0114), also binds xylose and is regulated by XylR (33, 50). The XylR transcriptional regulator (TM0110) can bind both glucose and xylose, which de-represses transcription of XylR regulon genes (D. Rodionov, personal communication). Therefore, when *T. petrophila* grows on glucose, it appears that glucose is brought into the cell via XylE1, and glucose binds to the XylR regulator, resulting in increased transcription of all of the XylR regulon genes. *T. petrophila* might be considered inefficient at utilizing glucose, because it must up-regulate so many unneeded XylR genes. This is an interesting example of small differences between closely related species resulting in significantly different strategies for carbohydrate utilization and transcriptional regulation.

SUMMARY

T. maritima, *T. neapolitana*, *T. petrophila*, and *T.* sp. strain RQ2 share approximately 1,470 genes (about 75% of their genomes), yet there are differences in sugar utilization and regulatory phenotypes that relate to the genotypes of these species. The multi-species microarray revealed multiple genes that are present in the *T. maritima* strain from DSMZ, but missing from the published genome sequence, including a glucose transporter that had eluded previous research efforts. When given a mixture of monosaccharides, only *T.* sp.

strain RQ2 consumed fructose, consistent with the fact that the fructose specific phosphotransferase system is not present in the other three species. Also, the absence of *xylE2* in *T. petrophila* impairs its ability to use glucose and necessitates the up-regulation of *xylE1* along with the rest of the xylan utilization regulon during growth on glucose.

The results presented here provide further evidence that the ability of *Thermotoga* species to ferment individual carbohydrates in heterogeneous growth substrates is dictated by the available set of ABC transporters. The inability of some model biofuel producing microorganisms to co-ferment glucose and xylose has prompted metabolic engineering efforts to knock out certain native transporters and insert heterologous systems to remedy this deficiency (18, 20). *Thermotoga* species are known to naturally co-ferment glucose and xylose and merit close examination to understand the metabolic basis for this physiological trait. It is interesting that while *Thermotoga* species appear to transport glucose and xylose using separate ABC transporters, no transporters have been identified that are specific for glucose or xylose. The two sugar binding proteins known to bind either sugar (XylE1 and XylE2) in fact bind both glucose and xylose. This versatility may be important in utilizing renewable heterogeneous feedstocks, which otherwise could limit biofuel production. This underscores the importance of understanding the connection between carbohydrate active enzymes, carbohydrate transporters, and their coordinated regulation.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy GTL Program (DG-FG02-08ER64687). A.D.F. and S.R.G. acknowledge support from NIH Pre-doctoral Biotechnology Traineeships (NIH T32 GM008776-06).

Helpful discussions with K. Noll (University of Connecticut) and D. Rodionov (Sanford-Burnham Medical Research Institute, La Jolla, CA) are also acknowledged.

REFERENCES

1. **Atsumi, S., T. Hanai, and J. C. Liao.** 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature **451:**86-89.

2. **Ballschmiter, M., O. Futterer, and W. Liebl.** 2006. Identification and characterization of a novel intracellular alkaline alpha-amylase from the hyperthermophilic bacterium *Thermotoga maritima* MSB8. Appl. Environ. Microbiol. **72:**2206-11.

3. **Blumer-Schuette, S. E., I. Kataeva, J. Westpheling, M. W. Adams, and R. M. Kelly.** 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr. Opin. Biotechnol. **19:**210-7.

4. **Bonch-Osmolovskaya, E. A., M. L. Miroshnichenko, A. V. Lebedinsky, N. A. Chernyh, T. N. Nazina, V. S. Ivoilov, S. S. Belyaev, E. S. Boulygina, Y. P. Lysov, A. N. Perov, A. D. Mirzabekov, H. Hippe, E. Stackebrandt, S. L'Haridon, and C. Jeanthon.** 2003. Radioisotopic, culture-based, and oligonucleotide microchip analyses of thermophilic microbial communities in a continental high-temperature petroleum reservoir. Appl. Environ. Microbiol. **69:**6143-6151.

5. **Boucher, N., and K. M. Noll.** 2011. Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. Appl. Environ. Microbiol. **77:**6395-6399.

6. **Chhabra, S. R., and R. M. Kelly.** 2002. Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74 with and without a carbohydrate binding module (CBM). FEBS Lett. **531:**375-380.

7. **Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly.** 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J. Biol. Chem. **278:**7540-7552.

8. **Chhabra, S. R., K. R. Shockley, D. E. Ward, and R. M. Kelly.** 2002. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown on glucan- and mannan-based polysaccharides. Appl. Environ. Microbiol. **68:**545-554.

9. **Chou, C. J., F. E. Jenney, Jr., M. W. Adams, and R. M. Kelly.** 2008. Hydrogenesis in hyperthermophilic microorganisms: implications for biofuels. Metab. Eng. **10:**394-404.

10. **Conners, S. B., E. F. Mongodin, M. R. Johnson, C. I. Montero, K. E. Nelson, and R. M. Kelly.** 2006. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. FEMS Microbiol. Rev. **30:**872-905.

11. **Conners, S. B., C. I. Montero, D. A. Comfort, K. R. Shockley, M. R. Johnson, S. R. Chhabra, and R. M. Kelly.** 2005. An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **187:**7267-7282.

12. **Dahle, H., F. Garshol, M. Madsen, and N. K. Birkeland.** 2008. Microbial community structure analysis of produced water from a high-temperature North Sea oil-field. Antonie Van Leeuwenhoek **93:**37-49.

13. **Edgar, R., M. Domrachev, and A. E. Lash.** 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. **30:**207-210.

14. **Fischer, M., Q. Y. Zhang, R. E. Hubbard, and G. H. Thomas.** 2010. Caught in a TRAP: substrate-binding proteins in secondary transport. Trends Microbiol. **18:**471-478.

15. **Frock, A. D., J. S. Notey, and R. M. Kelly.** 2010. The genus *Thermotoga*: recent developments. Environ. Technol. **31:**1169-1181.

16. **Galperin, M. Y., K. M. Noll, and A. H. Romano.** 1996. The glucose transport system of the hyperthermophilic anaerobic bacterium *Thermotoga neapolitana*. Appl. Environ. Microbiol. **62:**2915-2918.

17. **Gao, J., M. W. Bauer, K. R. Shockley, M. A. Pysz, and R. M. Kelly.** 2003. Growth of hyperthermophilic archaeon *Pyrococcus furiosus* on chitin involves two family 18 chitinases. Appl. Environ. Microbiol. **69:**3119-3128.

18. **Gosset, G.** 2005. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb. Cell Fact. **4:**14.

19. **Hetzer, A., H. W. Morgan, I. R. McDonald, and C. J. Daughney.** 2007. Microbial life in Champagne Pool, a geothermal spring in Waiotapu, New Zealand. Extremophiles **11:**605-614.

20. **Ho, N. W., Z. Chen, and A. P. Brainard.** 1998. Genetically engineered *Saccharomyces* yeast capable of effective cofermentation of glucose and xylose. Appl. Environ. Microbiol. **64:**1852-1859.

21. **Huber, R., T. A. Langworthy, H. Konig, M. Thomm, C. R. Woese, U. B. Sleytr, and K. O. Stetter.** 1986. *Thermotoga maritima* sp-nov represents a new genus of unique extremely thermophilic eubacteria growing up to 90 degrees C. Arch. Microbiol. **144:**324-333.

22. **Johnson, M. R., S. B. Conners, C. I. Montero, C. J. Chou, K. R. Shockley, and R. M. Kelly.** 2006. The *Thermotoga maritima* phenotype is impacted by syntrophic interaction with *Methanococcus jannaschii* in hyperthermophilic coculture. Appl. Environ. Microbiol. **72:**811-818.

23. **Johnson, M. R., C. I. Montero, S. B. Conners, K. R. Shockley, S. L. Bridger, and R. M. Kelly.** 2005. Population density-dependent regulation of exopolysaccharide formation in the hyperthermophilic bacterium *Thermotoga maritima*. Mol. Microbiol. **55:**664-674.

24. **Johnson, M. R., C. I. Montero, S. B. Conners, K. R. Shockley, M. A. Pysz, and R. M. Kelly.** 2004. Functional genomics-based studies of the microbial ecology of hyperthermophilic micro-organisms. Biochem. Soc. Trans. **32:**188-192.

25. **Li, H., S. Z. Yang, B. Z. Mu, Z. F. Rong, and J. Zhang.** 2007. Molecular phylogenetic diversity of the microbial community associated with a high-temperature petroleum reservoir at an offshore oilfield. FEMS Microbiol. Ecol. **60:**74-84.

26. **Liebl, W., J. Gabelsberger, and K. H. Schleifer.** 1994. Comparative amino acid sequence analysis of *Thermotoga maritima* beta-glucosidase (BglA) deduced from the nucleotide sequence of the gene indicates distant relationship between beta-glucosidases of the BGA family and other families of beta-1,4-glycosyl hydrolases. Mol. Gen. Genet. **242:**111-115.

27. **Lu, G., L. Jiang, R. M. Helikar, T. W. Rowley, L. Zhang, X. Chen, and E. N. Moriyama.** 2006. GenomeBlast: a web tool for small genome comparison. BMC Bioinformatics **7 Suppl 4:**S18.

28. **Miroshnichenko, M. L., and E. A. Bonch-Osmolovskaya.** 2006. Recent developments in the thermophilic microbiology of deep-sea hydrothermal vents. Extremophiles **10:**85-96.

29. **Montero, C. I., M. R. Johnson, C. J. Chou, S. B. Conners, S. G. Geouge, S. Tachdjian, J. D. Nichols, and R. M. Kelly.** 2007. Responses of wild-type and resistant strains of the hyperthermophilic bacterium *Thermotoga maritima* to chloramphenicol challenge. Appl. Environ. Microbiol. **73:**5058-5065.

30. **Montero, C. I., D. L. Lewis, M. R. Johnson, S. B. Conners, E. A. Nance, J. D. Nichols, and R. M. Kelly.** 2006. Colocation of genes encoding a tRNA-mRNA hybrid and a putative signaling peptide on complementary strands in the genome of the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **188:**6802-6807.

31. **Nanavati, D., K. M. Noll, and A. H. Romano.** 2002. Periplasmic maltose- and glucose-binding protein activities in cell-free extracts of *Thermotoga maritima*. Microbiology **148:**3531-3537.

32. **Nanavati, D. M., T. N. Nguyen, and K. M. Noll.** 2005. Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*. J. Bacteriol. **187:**2002-2009.

33. **Nanavati, D. M., K. Thirangoon, and K. M. Noll.** 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. Appl. Environ. Microbiol. **72:**1336-1345.

34. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, L. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature **399:**323-329.

35. **Nesbo, C. L., M. Dlutek, and W. F. Doolittle.** 2006. Recombination in *Thermotoga*: implications for species concepts and biogeography. Genetics **172:**759-769.

36. **Nesbo, C. L., K. E. Nelson, and W. F. Doolittle.** 2002. Suppressive subtractive hybridization detects extensive genomic diversity in *Thermotoga maritima*. J. Bacteriol. **184:**4475-4488.

37. **Noll, K. M., P. Lapierre, J. P. Gogarten, and D. M. Nanavati.** 2008. Evolution of mal ABC transporter operons in the Thermococcales and Thermotogales. BMC Evol. Biol. **8:**7.

38. **Novichkov, P. S., O. N. Laikova, E. S. Novichkova, M. S. Gelfand, A. P. Arkin, I. Dubchak, and D. A. Rodionov.** 2010. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. Nucleic Acids Res. **38:**D111-8.

39. **Orphan, V. J., L. T. Taylor, D. Hafenbradl, and E. F. Delong.** 2000. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. Appl. Environ. Microbiol. **66:**700-711.

40. **Parker, K. N., S. R. Chhabra, D. Lam, W. Callen, G. D. Duffaud, M. A. Snead, J. M. Short, E. J. Mathur, and R. M. Kelly.** 2001. Galactomannanases Man2 and Man5 from *Thermotoga* species: growth physiology on galactomannans, gene sequence analysis, and biochemical properties of recombinant enzymes. Biotechnol. Bioeng. **75:**322-33.

41. **Paulsen, I. T., L. Nguyen, M. K. Sliwinski, R. Rabus, and M. H. Saier, Jr.** 2000. Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. J. Mol. Biol. **301:**75-100.

42. **Pysz, M. A., S. B. Conners, C. I. Montero, K. R. Shockley, M. R. Johnson, D. E. Ward, and R. M. Kelly.** 2004. Transcriptional analysis of biofilm formation processes in the anaerobic, hyperthermophilic bacterium *Thermotoga maritima*. Appl. Environ. Microbiol. **70:**6098-6112.

43. **Pysz, M. A., D. E. Ward, K. R. Shockley, C. I. Montero, S. B. Conners, M. R. Johnson, and R. M. Kelly.** 2004. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8:**209-217.

44. **Rajashekhara, E., M. Kitaoka, Y. K. Kim, and K. Hayashi.** 2002. Characterization of a cellobiose phosphorylase from a hyperthermophilic eubacterium, *Thermotoga maritima* MSB8. Biosci. Biotechnol. Biochem. **66:**2578-2586.

45. **Saul, D. J., L. C. Williams, R. A. Reeves, M. D. Gibbs, and P. L. Bergquist.** 1995. Sequence and expression of a xylanase gene from the hyperthermophile *Thermotoga* sp. strain FjSS3-B.1 and characterization of the recombinant enzyme and its activity on kraft pulp. Appl. Environ. Microbiol. **61:**4110-3.

46. **Schut, G. J., and M. W. Adams.** 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. J. Bacteriol. **191:**4451-7.

47. **Shockley, K. R., K. L. Scott, M. A. Pysz, S. B. Conners, M. R. Johnson, C. I. Montero, R. D. Wolfinger, and R. M. Kelly.** 2005. Genome-wide transcriptional variation within and between steady states for continuous growth of the hyperthermophile *Thermotoga maritima*. Appl. Environ. Microbiol. **71:**5572-5576.

48. **Shockley, K. R., D. E. Ward, S. R. Chhabra, S. B. Conners, C. I. Montero, and R. M. Kelly.** 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microbiol. **69:**2365-2371.

49. **Takahata, Y., M. Nishijima, T. Hoaki, and T. Maruyama.** 2001. *Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan. Int. J. Syst. Evol. Microbiol. **51:**1901-1909.

50. **Tian, Y., M. J. Cuneo, A. Changela, B. Hocker, L. S. Beese, and H. W. Hellinga.** 2007. Structure-based design of robust glucose biosensors using a *Thermotoga maritima* periplasmic glucose-binding protein. Protein Sci. **16:**2240-2250.

51. **VanFossen, A. L., I. Ozdemir, S. L. Zelin, and R. M. Kelly.** 2011. Glycoside hydrolase inventory drives plant polysaccharide deconstruction by the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Biotechnol. Bioeng. **108:**1559-1569.

52. **VanFossen, A. L., M. R. Verhaart, S. M. Kengen, and R. M. Kelly.** 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl. Environ. Microbiol. **75:**7718-7724.

53. **Windberger, E., R. Huber, A. Trincone, H. Fricke, and K. Stetter.** 1989. *Thermotoga thermarum* sp-nov and *Thermotoga neapolitana* occurring in African continental solfataric springs. Arch. Microbiol. **151:**506-512.

54. **Yomano, L. P., S. W. York, S. Zhou, K. T. Shanmugam, and L. O. Ingram.** 2008. Re-engineering *Escherichia coli* for ethanol production. Biotechnol. Lett. **30:**2097-2103.

55. **Yoshida, A., T. Nishimura, H. Kawaguchi, M. Inui, and H. Yukawa.** 2005. Enhanced hydrogen production from formic acid by formate hydrogen lyase-overexpressing *Escherichia coli* strains. Appl. Environ. Microbiol. **71:**6762-6768.

56. **Zhaxybayeva, O., K. S. Swithers, P. Lapierre, G. P. Fournier, D. M. Bickhart, R. T. Deboy, K. E. Nelson, C. L. Nesbo, W. F. Doolittle, J. P. Gogarten, and K. M. Noll.** 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. U.S.A. **106:**5865-5870.

Table 3.1. Genomic characteristics of *Thermotoga* species. Shading indicates species with an optimum growth temperature (T_{opt}) of 75 to 80°C.

Species	Isolation site	Location	T_{opt} (°C)	16S rRNA homology to *T. maritima*	Protein coding ORFs
T. maritima	geothermally heated sea floor	Vulcano, Italy	80	100	1858
T. sp. RQ2	geothermally heated sea floor	São Miguel, Azores	80	99.7	1819
T. neapolitana	submarine thermal vent	Naples, Italy	77	99.4	1905
T. petrophila	oil reservoir production fluid	Niigata, Japan	80	99.2	1785
T. naphthophila	oil reservoir production fluid	Niigata, Japan	80	99.0	1768
T. thermarum	continental solfataric spring	Djibouti, Africa	70	91.8	1945
T. lettingae	sulfate-reducing bioreactor	Wageningen, Netherlands	65	90.8	2040
T. elfii	oil-producing well	Africa	66	90.7	-
T. subterranea	deep continental oil reservoir	East Paris Basin, France	70	90.6	-
T. hypogea	oil-producing well	Cameroon, Africa	70	90.2	-

Table 3.2. Carbohydrate transporter binding proteins identified in genomes of hyperthermophilic *Thermotoga* species.

Predicted substrates	Species			
	T. maritima	***T. neapolitana***	***T. petrophila***	***T.*sp RQ2**
α-Arabinosides	TM0277	CTN_0408	Tpet_0646-7	TRQ2_0671
Cellobiose, laminaribiose	TM0031	CTN_0664	Tpet_0892	TRQ2_0914
Chitobiose	TM0810	CTN_1767	Tpet_0118	TRQ2_0116
Maltose	TM1204	CTN_1367	Tpet_1552	TRQ2_1614
Maltose	TM1839	CTN_0765	Tpet_0966	TRQ2_0984
Mannobiose	TM1223	CTN_1348	Tpet_1545	TRQ2_1595
Rhamnose-related	TM1067	CTN_1502-3	Tpet_1677, Tpet_1687	TRQ2_0510
Trehalose, maltose	*treE*	CTN_0780	Tpet_0954	TRQ2_0970
Xylan-related	TM0056	CTN_0638	Tpet_0868	TRQ2_0890
Xylobiose, xylotriose	TM0071	CTN_0622	Tpet_0853	TRQ2_0876
Xylose, glucose	TM0114	CTN_0576	Tpet_0810	TRQ2_0833
Unknown carbohydrate	TM0309	CTN_0378-9	Tpet_0608	TRQ2_0622
Unknown carbohydrate	TM1235	CTN_1337	Tpet_1534	TRQ2_1583
Unknown monosaccharide	TM0595	CTN_0067	Tpet_0322	TRQ2_0340
Galactoside	TM1199	CTN_1375		TRQ2_1619
Glucose, xylose	*xylE*	CTN_0777		TRQ2_0973
Mannose-glucuronic acid-related	TM1855	CTN_0790	Tpet_0942	
Myo-inositol	TM0418	CTN_0252	Tpet_0503	
Pectin/galacturonate-related	TM0432		Tpet_1679	TRQ2_0512
Ribose	TM0958	CTN_1618	Tpet_1794	
Arabinose			Tpet_0636	TRQ2_0661
Mannan-related	TM1746			TRQ2_1079
Mannooligosaccharides	TM1226			TRQ2_1592
Xyloglucooligosaccharides	TM0300			TRQ2_0631
β-Glucoside		CTN_0660		
Fructose (PTS)				TRQ2_0639
Galactoside		CTN_1372		
Pectin/galacturonate-related			Tpet_0485	
		CTN_0240		
Unknown carbohydrate		CTN_1541		
Unknown monosaccharides		CTN_0358		
Unknown monosaccharides		CTN_0364		

Table 3.3. ORFs responding in hyperthermophilic *Thermotoga* species during growth on polysaccharide mixture compared to growth on glucose. Up-regulated on polysaccharides (black), up on glucose (white), no significant change (light gray), absent from genome (dark gray).

Probe	Product	M	N	P	R
Mannan utilization					
TM1746	mannan ABC transporter, periplasmic oligopeptide-binding protein	4.6			3.7
TM1747	mannan ABC transporter, permease protein	3.7			8.3
TM1748	mannan ABC transporter, permease protein				
TM1749	mannan ABC transporter, ATP-binding protein	6.4			4.3
TM1750	mannan ABC transporter, ATP-binding protein	5.5			6.7
TM1751	endoglucanase	5.9			2.5
ManR operon					
TM1224	ManR transcriptional regulator, ROK family	3.3	2.4		8.7
TM1226	mannooligosaccharide ABC transporter, sugar-binding protein	5.4			18.3
TM1227	endo-1,4-beta-mannosidase	24.1	12.7	3.9	15.3
CelR regulon					
TM0312	predicted dehydrogenase	2.1	2.1		
TM0313	predicted aldo/keto reductase	2.0		2.6	
TM1218	CelR transcriptional regulator, LacI family	3.8		3.9	
TM1219	mannobiose ABC transporter, ATP-binding protein		2.6		2.5
TM1223	mannobiose ABC transporter, sugar-binding protein	10.0	25.4	4.9	18.8
TM1524	endoglucanase	7.7		6.8	3.9
TM1525	endoglucanase	9.5		7.3	
TM1848	cellobiose phosphorylase	10.2	4.7	19.7	9.4
TRAP transporter					
TM0322	TRAP transporter, periplasmic substrate-binding protein, putative		8.1		14.7
TM0323	TRAP transporter, small transmembrane component				7.0
TM0324	TRAP transporter, large transmembrane component				3.9
TM0325	predicted sugar dehydrogenase		2.8		12.0
TM0326	transcriptional regulator, RpiR family	-3.6			2.6
TM0327	phosphoglycerate dehydrogenase, putative				6.2
Processing of α-glucans					
TM0752	alpha-glucosidase, putative	5.3			2.2
TM0767	maltodextrin glycosyltransferase		-2.7		-6.1
TM1834	alpha-glucosidase				-3.6
TM1839	maltose ABC transporter, periplasmic maltose-binding protein		-3.8	-2.3	-6.6
TM1841	hypothetical protein				-2.5
TM1842	hypothetical protein		-2.6	-3.2	-5.9
TM1843	hypothetical protein				-3.2
TM1844	hypothetical protein				-5.1
TM1845	pullulanase			-5.1	
CTN_1407	alpha-glucan phosphorylase				-2.3
TRQ2_1647a			-2.2		-2.5
BglR operon					
TM0023	methyl-accepting chemotaxis protein			-7.8	-2.4
TM0024	laminarinase	-2.6		-11.8	
TM0025	beta-glucosidase	-2.1	2.1	-6.7	
TM0026	hypothetical protein			-3.1	
TM0027	cellobiose/laminaribiose ABC transporter, ATP-binding protein	-2.0		-4.7	-2.6
TM0028	cellobiose/laminaribiose ABC transporter, ATP-binding protein	-2.0		-8.0	-2.4
TM0029	cellobiose/laminaribiose ABC transporter, permease protein			-5.0	-2.6
TM0031	cellobiose/laminaribiose ABC transporter, sugar-binding protein	-2.7		-11.6	-2.7
TM0032	BglR transcriptional regulator, ROK family			-4.2	

Table 3.4. ORFs regulated by GluR in four hyperthermophilic *Thermotoga* species during growth on a polysaccharide mix compared to growth on glucose.

Probe	Product	M	N	P	R
TM1847	GluR transcriptional regulator, ROK family				-2.8
TRQ2_0975-0976	GluR transcriptional regulator/glucose ABC transporter, ATP-binding protein				-2.7
CTN_0775	glucose ABC transporter, ATP-binding protein	-2.4			-4.6
TRQ2_0975a		-3.2			-2.5
TRQ2_0975b		-3.1			-2.2
TRQ2_0975c		-3.0			-2.5
TRQ2_0974-0975	glucose ABC transporter, ATP-binding protein/permease protein	-3.0			-3.1
TRQ2_0974	glucose ABC transporter, permease protein				-2.4
CTN_0777	glucose ABC transporter, sugar-binding protein		-2.2		-7.0
TRQ2_0973b		-2.9	-2.0		-14.4
TRQ2_0973c		-3.5			-6.3
TRQ2_0972-0973	glucose transporter sugar-binding protein/trehalose transporter ATP-binding	-2.2			-6.4
CTN_0779b	trehalose ABC transporter, permease protein	-2.1		-2.4	
TRQ2_0971		-3.3		-3.0	
CTN_0780	trehalose ABC transporter, sugar-binding protein			-2.8	-8.4
TRQ2_0970a				-4.0	-18.2
TRQ2_0970b				-4.2	-13.1

Table 3.5. ORFs regulated by XylR in four *Thermotoga* species during growth on a polysaccharide mix compared to growth on glucose.

Probe	Product	M	N	P	R
TM0056	xylan ABC transporter, periplasmic sugar-binding protein	6.1	4.8	-5.8	9.7
TM0057	xylan ABC transporter, ATP-binding protein	5.7	3.6		6.2
TM0058	xylan ABC transporter, ATP-binding protein	5.7		-3.5	
TM0060	xylan ABC transporter, permease protein	4.7			
TM0061	endo-1,4-beta-xylanase A	8.2	11.7	-15.3	16.8
CTN_0632		2.2	5.9		
TM0062	hypothetical protein	19.4		-6.6	
TM0070	endo-1,4-beta-xylanase B	10.6	2.7		3.1
TM0071	xylooligosaccharide ABC transporter, periplasmic sugar-binding protein	13.8	2.6	-7.3	14.4
TM0072	xylooligosaccharide ABC transporter, permease protein	2.1		-4.8	
TM0073	xylooligosaccharide ABC transporter, permease protein	2.1		-3.9	2.9
TM0074	xylooligosaccharide ABC transporter, ATP-binding protein			-2.9	
TM0075	xylooligosaccharide ABC transporter, ATP-binding protein			-2.6	
TM0076	xylosidase			-2.6	
TM0077	acetyl xylan esterase		2.6	-2.2	2.6
TM0110	transcriptional regulator, XylR-related	2.5			
TM0111	alcohol dehydrogenase, iron-containing			-3.6	
TM0113	xylU-related protein			-4.0	
TM0114	xylose/glucose ABC transporter, periplasmic sugar-binding protein			-4.6	
TM0115	xylose/glucose ABC transporter, ATP-binding protein			-2.7	
TM0309	xylan ABC transporter, periplasmic sugar-binding protein	7.5	3.0	-2.6	10.3
TM0310	beta-D-galactosidase	9.2		-2.1	2.8
TM1667	xylose isomerase	7.2	2.8		6.3

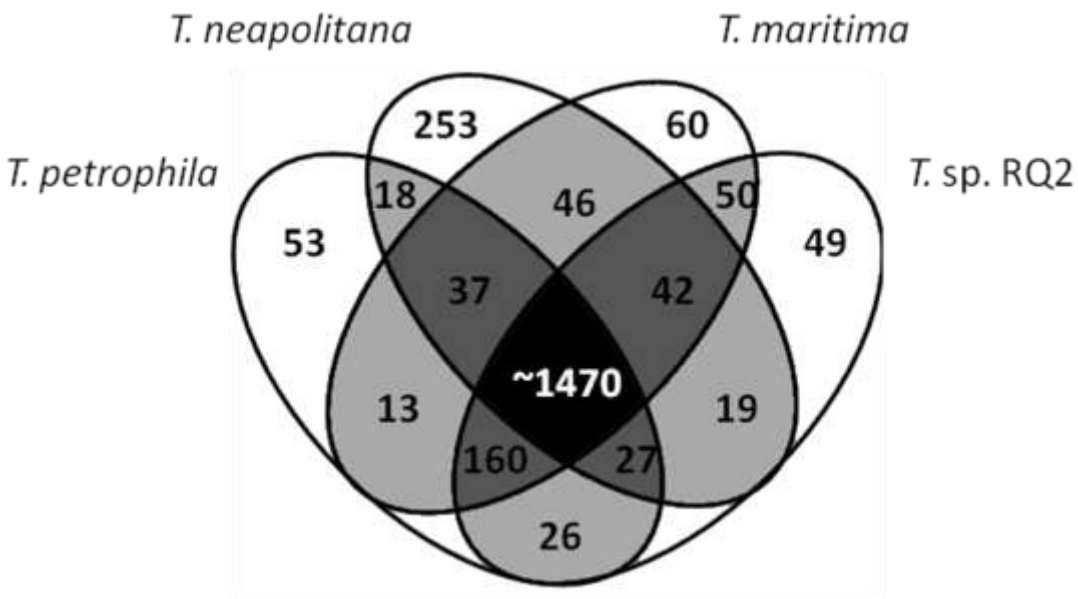

Figure 3.1. Venn diagram summarizing the shared ORFs of the selected hyperthermophilic *Thermotoga* species, based on 70% identity and 80% coverage at the amino acid level over the entire ORF. Numbers indicate the ORFs shared by subsets of species.

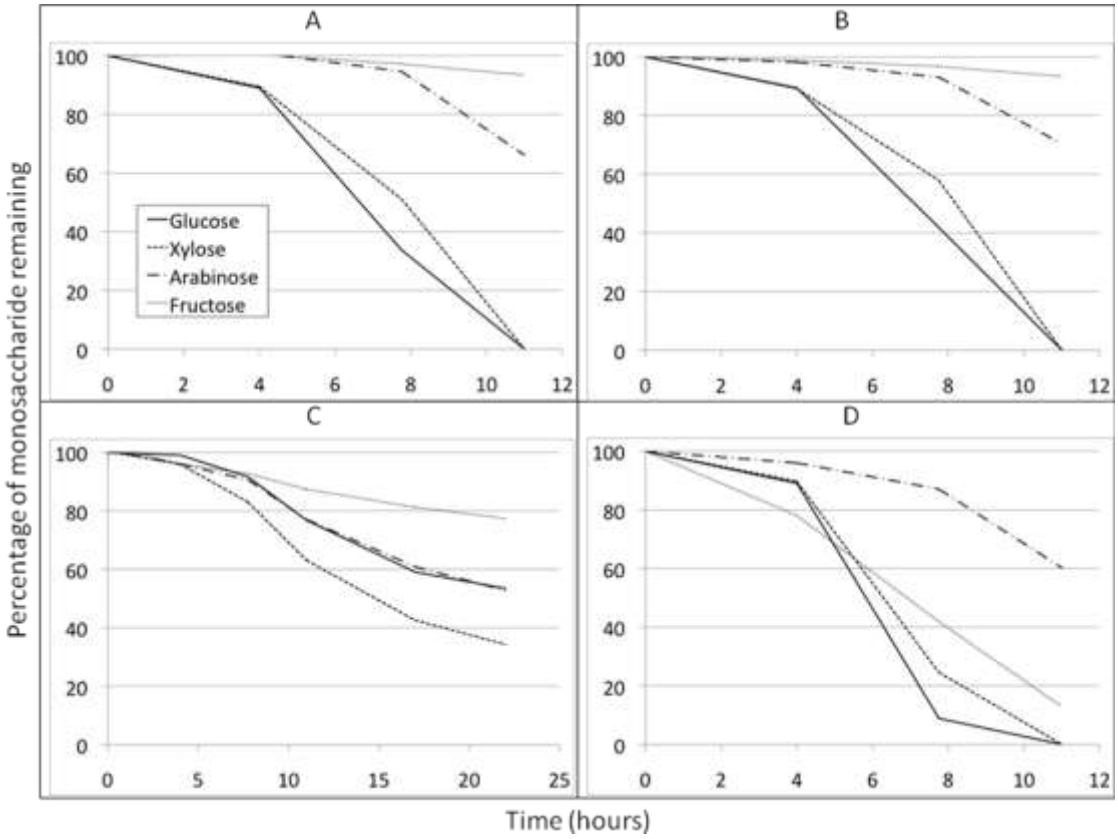

Figure 3.2. Utilization of a mixture of six monosaccharides by four *Thermotoga* species. Mannose and galactose were not significantly utilized by any of the four species, so these two sugars were omitted from the figures for clarity. (A) *T. maritima*; (B) *T. neapolitana*; (C) *T. petrophila*; (D) *Thermotoga* sp. strain RQ2.

Chapter 4

Stationary phase triggers unusual coccoid morphology and production of a peptide (TM1316) in the hyperthermophilic bacterium *Thermotoga maritima*

Andrew D. Frock, Clemente I. Montero, Sara Blumer-Schuette, and Robert M. Kelly

Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

To be submitted to: *Applied and Environmental Microbiology*

ABSTRACT

Although proteins and peptides encoded in small open reading frames (ORFs < 100 AA) in microbial genomes can play critical roles as toxins, bacteriocins, transcriptional regulators, signaling molecules, and chaperones, many such ORFs remain annotated as "hypothetical proteins". In the genome of the hyperthermophilic bacterium, *Thermotoga maritima*, nearly 2/3 of the 167 small ORFs have no known function. As a strategy for investigating the potential significance of specific small ORFs, growth conditions that could trigger the expression of genes encoding small ORFs were sought. A defined medium supplemented with either 1 or 5 g/L yeast extract was used to track transcriptional response during the transition from exponential to stationary phase. The phase transition for the richer medium resulted in differential transcription of 359 genes, many of which were involved in flagella biosynthesis, cell division, carbon metabolism, as well as the synthesis of proteins, vitamins, and amino acids. Growth on the leaner medium triggered a comparable number of genes (254) with similar functional roles, but also a ~3-fold lower cell density, with many cells becoming bloated and coccoidal upon entry into stationary phase. Of particular interest, an ORF corresponding to a putative 31-amino acid peptide (TM1316) was dramatically up-regulated in stationary phase on the lean medium, such that it was the most highly transcribed gene in the transcriptome. TM1316 is associated with the TM1298-1336 region of the genome, a locus which has a repetitive organization of putative toxin/antitoxin pairs, small hypothetical ORFs, radical SAM proteins, and ABC-2 transporters. The only functionally characterized loci with similar organization were found to be responsible for bacteriocin production in *Bacillus* species. TM1316, consistent with transcriptomic data, was produced at high levels during stationary phase and localized to the cell envelope. Possible

physiological roles and implications for the study of small ORFs are discussed in light of experimental results and bioinformatic analysis of this unique region of the *T. maritima* genome.

INTRODUCTION

The increasing availability of genome sequence information has driven the need for, and development of, improved tools for gene annotation that predict the function of encoded biomolecules. As a result, the number of open reading frames (ORFs) annotated as "hypothetical proteins" continues to shrink. However, smaller ORFs (i.e., corresponding to proteins and peptides of less than 100 amino acids) remain annotated to a disproportionate extent as "hypotheticals" **(Table 1)**, awaiting closer bioinformatic inspection and complementary biochemical evaluation. The problem is that small proteins and peptides are "needles in a haystack" of small ORFs that are present in genomes (4). Any random stretch of 50 codons has a 9% chance of lacking a stop codon, while a stretch of 100 codons has a 0.8% chance of lacking a stop codon. Therefore, even a small bacterial genome will contain an abundance of small ORFs, many of which are likely to have minimal biological significance.

For many reasons, small ORFs are often subject to inaccurate gene calls. The bioinformatic difficulties in accurately predicting small protein- and peptide-coding genes can be addressed by experimental techniques, such as tiling arrays and deep sequencing of RNA, but these methods are not yet typically employed as a part of genome annotation. However, even after these small ORFs are accurately identified, there are other challenges to annotating and characterizing the corresponding small proteins and peptides. Many have

physiological roles that depend on specific modifications, localization to the membrane, or association with other proteins, such that over-production in a heterologous host (a workhorse method for characterization of larger proteins) may not be useful, certainly the case if post-translational processing is involved. The small size also complicates one- or two-dimensional PAGE and generation of non-polar disruption mutants. Finally, some small ORFs may be located in "silent" genomic regions that are not expressed under standard laboratory conditions.

Although small proteins and peptides may not be biocatalytic, they can function in a variety of roles as toxins, bacteriocins, extracellular signaling peptides, subunits of larger protein complexes, regulators of other proteins/enzymes, and chaperones of metals and nucleic acids (1, 18). There are many such examples. Increased production of the MntS protein (42 aa) from *Escherichia coli* was observed under manganese-limited conditions, but its overproduction renders cells sensitive to manganese, suggesting that it plays a role in increasing the intracellular bioavailability of the metal (43). SafA (65 aa) from *E. coli* is induced by the EvgS/EvgA two-component system (TCS), and it activates the PhoQ/PhoP TCS by directly interacting with PhoQ (10). KdpF (29 aa) is a stabilizing component of the K^+-translocating KdpFABC complex (13). In *Bacillus subtilis,* the ComX peptide (55 aa) regulates competence and surfactin synthesis when the secreted form activates the histidine kinase ComP, which phosphorylates the transcription factor ComA (25). In the hyperthermophilic bacterium, *Thermotoga maritima*, the tmRNA gene is co-located with a gene on the opposite strand (TM0504) that encodes a 42 aa peptide involved in quorum sensing-like induction of exopolysaccharide production (21, 27). Finally, a number of Class I and Class II bacteriocins are encoded in small ORFs, including subtilosin (43 aa) and lantibiotics, such as nisin (57 aa) (22, 47).

To understand the significance of small ORFs in microbial genomes, growth conditions under which their transcription and production are triggered must be identified. Here, the transcriptional response of *T. maritima* during growth phase transitions and subject to varying nutrient levels was examined with an eye towards ORFs corresponding to fewer than 100 aa. These experiments revealed that transcripts for most small ORFs were generated under these growth conditions. In particular, this led to the identification of a genomic island that contained several putative peptides, one of which corresponded to the most highly expressed ORF in the *T. maritima* transcriptome during stationary phase.

MATERIALS AND METHODS

Growth of *Thermotoga maritima*. *T. maritima* DSM-3109 was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ). *T. maritima* was grown anaerobically in an oil bath at 80°C with shaking at 100 rpm. Culture bottles were made anaerobic by sealing, sparging with nitrogen, and adding $Na_2S•9H_2O$ at a final concentration of 0.05% (w/v). For preparation of freezer stocks, 970 µl of cooled log-phase culture was added to 30 µl of dimethylsulfoxide and stored at -80°C. Freezer stocks were thawed and used to inoculate (1% v/v) artificial sea water (ASW) media (9) with 2.5 g/l maltose. After ~12 hours of growth, this culture was used to inoculate medium with varying concentrations of yeast extract (YE). This medium was based on a defined medium developed from ASW. The base medium consisted of (per L, final concentration): 15 g NaCl, 2 g Na_2SO_4, 2 g $MgCl_2•6H_2O$, 0.25 g $NaHCO_3$, 20 mg KBr, 20 mg H_3BO_3, 20 mg KI, 3 mg Na_2WO_4, and 2 mg $NiCl_2•6H_2O$. The base medium was typically prepared at 2x concentration, adjusted to pH 7 with H_2SO_4, and sterilized by autoclaving. After autoclaving,

1.2 g KH_2PO_4, 2.5 g maltose monohydrate, 5 g NH_4Cl, 5.4 mg $FeCl_3$•$6H_2O$, 1x trace elements (see below), and 1x vitamins (see below) were added (per L, final concentration). All of these components were prepared as concentrated, sterile-filtered stock solutions. YE concentration was varied by addition of a 25 g/L stock solution that was sterilized by filtration. Vitamin solution (1000x) consisted of (per L): 10 mg pyridoxine•HCl, 5 mg calcium pantothenate, 5 mg nicotinic acid, 5 mg riboflavin, 5 mg thiamine•HCl, 5 mg thioctic acid, 2 mg biotin, 2 mg folic acid, and 0.1 mg cyanocobalamin. Trace element solution (1000x) consisted of (per L): 1.5 g nitrilotriacetic acid (NTA), 0.5 g $MnSO_4$•H_2O, 0.2 g NiCl•$6H_2O$, 0.1 g $ZnSO_4$•$7H_2O$, 0.01 g $CuSO_4$•$5H_2O$, and 0.01 g Na_2MoO_4•$2H_2O$. NTA was dissolved by adding concentrated KOH to pH 6.5. For cell enumeration, cells were fixed with 0.25% glutaraldehyde, stained with acridine orange, and viewed under an epifluorescence microscope (17).

Transcriptional response analysis via cDNA microarray. Cultures (300 mL) were harvested at different time points as indicated, rapidly cooled in a dry ice-ethanol bath, following which cells were collected by centrifugation at 7500 × g for 10 minutes. RNA extraction, reverse transcription, labeling, hybridization, and slide scanning were performed with modifications of previously described protocols (8, 21). Loess normalization, analysis of variance (ANOVA) normalization, and ANOVA row-by-row modeling were performed using JMP Genomics (SAS Institute, Cary, NC). Fold changes at or above 2-fold, with significance based on values at or above the Bonferroni correction (typically equivalent to a *P* value of approximately 10^{-5}). These experiments were performed with a multi-species microarray (12), which included probes for *T. maritima* and other Thermotoga species (*T. neapolitana*, *T. petrophila*, and *T.* sp. RQ2).

Immunocytochemistry of TM1316. Polyclonal rabbit antibodies (GeneTel Laboratories LLC, Madison, WI) were raised against a chemically synthesized version of the TM1316 peptide conjugated to carrier protein (keyhole limpet hemocyanin). For dot blots, 2 µl of sample were manually spotted onto a nitrocellulose membrane and allowed to dry. The membrane was blocked with 2.5% non-fat dry milk in TBS for 30 minutes, then incubated for 30 minutes with anti-TM1316 rabbit serum diluted 5000x in 2.5% non-fat dry milk in TBS-T (TBS + 0.1% Tween 20). The membrane was washed twice briefly with TBS-T, then washed four times for 5 minutes each, before 30 minute incubation with goat anti-rabbit IgG – HRP conjugate (Invitrogen) diluted 100,000x in 2.5% non-fat dry milk in TBS-T. The membrane was again washed as before; HRP was detected with SuperSignal West Femto chemiluminescent substrate (Pierce), and the membrane was imaged using a Carestream Gel Logic 2200 Pro system. For western blots, NuPAGE 4-12% Bis-Tris gels (Invitrogen) were run according to the manufacturer's instructions, and proteins were transferred to nitrocellulose membrane at 1 mA per cm^2 of membrane for 40 minutes. For chemiluminescent ladder detection, biotinylated protein ladder was loaded onto at least one lane of the gel, and 10,000x dilute goat anti-biotin IgG – HRP conjugate (Cell Signaling Technology, Danvers, MA) was added to the secondary antibody incubation.

RESULTS

Growth and morphology of *T. maritima* during phase transitions. Growth rates for *T. maritima* on 1 and 5 g/L yeast extract (YE), both in addition to 2.5 g/L maltose, were not significantly different (doubling time of ~80-90 minutes), although final cell densities on 5 g/L YE were significantly higher (4-5 x 10^8 cells/ml vs. 1.5-2 x 10^8 cells/mL) **(Figure 1)**.

During exponential phase of both cultures, cells exhibited the rod-shaped morphology that is generally observed for *T. maritima* (19). However, during stationary phase of the 1 g/L YE culture, a significant proportion of cells (roughly 20-50%) became bloated and coccoidal, as shown in **Figure 2**. This was not observed in cultures grown on 5 g/L YE.

Exponential and stationary phase transcriptomes (rich medium – 5 g/L yeast extract). When *T. maritima* was grown on medium containing 5 g/L YE and 2.5 g/L maltose, the genes differentially regulated between exponential and stationary phase were consistent with a significant phase transition and modification of a variety of cellular functions upon entry into stationary phase. Of the 187 ORFs significantly up-regulated in log phase compared to stationary phase **(Tables 2, 3, S1)**, 55 were involved in carbohydrate transport and metabolism, including glucose-inducible components of transporters for cellobiose (31), glucose (7), and trehalose (7). Two maltose transporters (30) were transcribed at slightly higher levels during exponential phase, but these differences generally did not meet the criteria for statistical significance. The cell division genes *minC*, *minD,* and *ftsI* were also more highly transcribed during log phase, along with genes involved in fatty acid biosynthesis (2), nucleic acid replication and repair (11), translation (35), and protein processing and secretion (10).

The stationary phase transcriptome for the rich medium suggested that various vitamins and amino acids were depleted during growth. Assimilation of ammonium as a nitrogen source apparently increased during stationary phase. An ammonium transporter (TM0402) and the nitrogen regulatory protein P-II (TM0403) were up-regulated, along with 12 flagella-related genes and genes involved in the biosynthesis of tyrosine (TM0344), leucine (TM0548-TM0555), cysteine (TM0882), and lysine (TM1519-TM1520), among

others. Genes involved in the biosynthesis of the vitamins riboflavin (TM1826-TM1827), thiamine (TM0787-8, TM0888, TM1244-51), and pantothenate (TM1077) were also more highly transcribed during stationary phase.

The transcriptional analysis also revealed a shift in carbon catabolism during the transition from exponential to stationary phase. During exponential phase, much of the glycolytic pathway from glucose to acetate is up-regulated, as well as a pathway (TM0064-TM0069) for the catabolism of glucuronate to glyceraldehyde-3-phosphate and pyruvate (38). Up-regulation of glycolysis is consistent with maltose utilization, but up-regulation of the glucuronate pathway suggests utilization of an unknown carbohydrate that is presumably present in the YE. A few transporter components of unknown function (TM0103, TM0104, TM0309) were up-regulated during exponential phase and may be involved in uptake of this unknown carbohydrate. During stationary phase, a glycerol uptake facilitator (TM1429) and a glycerol-3-phosphate transporter (TM1120-TM1122) were up-regulated, as well as a glycerol kinase (TM1430) and a glycerol dehydrogenase (TM0423) that might be involved in converting glycerol to glycerone phosphate (a glycolytic intermediate). The gene responsible for the downstream glycolytic glyceraldehyde-3-phosphate dehydrogenase reaction (TM0688) was also up-regulated. However, the glycerol kinase gene has a mutation that results in drastic loss of activity (39) and *T. maritima* is reportedly incapable of growth on glycerol (5). This might be simply explained by an inflexible metabolism that is energetically unable to utilize substrates that are more reduced than glucose. A transporter for a more oxidized substrate was also up-regulated during stationary phase, suggesting that its catabolism might balance that of glycerol. The transporter (encoded by TM0430-TM0432) was shown to bind α-1,4-digalacturonate *in vitro* (31). Its catabolism could intersect with glucuronate catabolism (which was up-regulated in exponential phase), but in

this case there is evidence that galacturonate might feed into the pentose phosphate pathway during stationary phase. TM0441-TM0443 were up-regulated during stationary phase, and these genes encode proteins that convert fructuronate (an isomer of galacturonate) to 6-phosphogluconate (38). Another gene of the oxidative pentose phosphate pathway (glucose-6-phosphate dehydrogenase, TM1155) was also up-regulated, indicating that *T. maritima* required increased reduction of $NADP^+$ during stationary phase.

The transcriptional response of a number of other genes provided further evidence that balancing redox reactions was a key issue for *T. maritima* under these conditions, although genes implicated in oxidative stress response (23) were not differentially regulated. The Rex transcriptional regulator (TM0169) responds to the cellular NADH/NAD^+ ratio and is predicted to control expression of 39 genes in 12 operons (32, 35). Ten of these genes were up-regulated in stationary phase compared to exponential phase, and half of those use NAD(P)H as a cofactor. These Rex-regulated genes include rubrerythrin (TM0009), a NADP-reducing hydrogenase (TM0011 and TM0012), a serine-pyruvate aminotransferase and hydroxypyruvate reductase (TM1400 and TM1401) (45), as well as the previously mentioned glycerol dehydrogenase (TM0423) and glyceraldehyde-3-phosphate dehydrogenase (TM0688).

Exponential and stationary phase transcriptomes (lean medium – 1 g/L yeast extract). The exponential to stationary phase transition on 1 g/L YE shared many features with that of growth on 5 g/L YE. During growth on 5 g/L YE, 187 probes were at least 2-fold higher in exponential phase compared to stationary phase; 80 of those 187 met the same criterion for the same comparison on 1 g/L YE. Nearly all probes responded in the same direction when exponential and stationary phase were compared for 1 and 5 g/L YE,

although not all met the criteria for significance. In general, the 1 g/L YE transcriptome exhibited a less pronounced transition from exponential phase to stationary phase. For example, 11 flagella genes were up-regulated in stationary phase compared to exponential phase on 5 g/L YE, but only 2 flagella genes were up-regulated for the same comparison on 1 g/L YE **(Table 3)**. During the exponential to stationary phase transition, many growth-associated genes (including ribosomal proteins and the cellobiose, glucose, and trehalose transporters) decreased significantly on 5 g/L YE, while they decreased to a much lesser extent (or not at all) on 1 g/L YE. Many of these genes are among the 70 that were "up-regulated" in 1 g/L YE stationary phase vs. 5 g/L YE stationary phase **(Table 2)**. So, 46 of those 70 genes did not respond to the same extent during the phase transition on 1 g/L YE as they did on 5 g/L YE.

The number of genes that responded to the change in medium composition (bottom half of **Table 2**) was clearly much smaller than those that responded to phase transition (top half of **Table 2**). During exponential phase, two glucose-inducible sugar-binding transporter components (TM0031 and *xylE2*) were up-regulated on 1 g/L YE compared to 5 g/L YE. This suggests that *T. maritima* was more dependent upon utilization of the defined carbon source (maltose) during growth on the leaner media, presumably because the increased levels of YE in the richer media provide additional alternative carbon sources. CRISPR-associated adaptive immunity proteins (TM1792-TM1802) (3) and the previously mentioned glycerol uptake operon (TM1429-TM1431) were also up-regulated in 1 g/L vs. 5 g/L YE during exponential phase. Only four genes were up-regulated in 5 g/L vs. 1 g/L YE during exponential phase: two hypothetical proteins (TM1266 and TM1268) in the same operon as putative thiamine and biotin synthesis proteins (TM1267 and TM1269, respectively), a prismane protein (TM1172), and a ferrous iron transport protein (TM0051). The majority of

iron in the media comes from the supplemented ferric chloride (1.1 mg/L iron), instead of the Bacto yeast extract (~0.06 mg/L iron for 1 g/L YE and 0.3 mg/L iron for 5 g/L YE), so the up-regulation of the ferrous iron transport protein may indicate that YE either affects iron solubility or provides iron in a more accessible form. A component of a ferric iron transporter (TM0190) was up-regulated in 5 g/L YE vs. 1 g/L YE during stationary phase. Other genes up-regulated in 5 g/l YE vs. 1 g/L YE during stationary phase include flagella-related genes, the primary sigma factor RpoD (TM1451), and the heat-shock inducible sigma factor RpoE (TM1598) (34). The rod shape-determining MreB protein was also up-regulated in 5 g/L YE stationary phase, which may be related to the coccoidal morphology observed in 1 g/L YE stationary phase **(Figure 2)**.

As discussed above, the transcription of a number of growth-associated genes did not decrease during the 1 g/L YE phase transition as was the case for the 5 g/L YE phase transition. Other genes up-regulated during stationary phase on the leaner medium included a putative sulfur relay operon (TM0979-TM0983). TM0981, TM0980, and TM0979 are homologous to DsrEFH and TusBCD, which are involved in sulfur relay for sulfide oxidation in *Allochromatium vinosum* (42) and tRNA thiolation in *Escherichia coli* (20), respectively. TM0982 appears to be a sulfur transport membrane protein (46). TM0816 (a MarR homolog) and TM0817 (an AcrB homolog) were also up-regulated in the leaner medium during stationary phase. In *E. coli*, MarR (multiple antibiotic resistance regulator) is a transcriptional repressor of the *acrAB* locus, among other genes. AcrAB is an efflux pump involved in resistance to many drugs, including tetracycline, chloramphenicol, and ampicillin (33). In this case, it is unclear whether TM0816 and TM0817 were induced in response to a general stress or the presence of a particular compound.

Response of small ORFs annotated as hypothetical proteins. As noted in **Table 1**, there are 113 ORFs of 100 or fewer amino acids that are annotated as “hypothetical proteins”. **Table 4** shows that many of these putative small ORFs generated transcripts for the growth conditions studied here, the levels of which varied significantly, but some at very high levels. Thirteen of these ORFs were differentially transcribed during phase transition on 1 or 5 g/L YE. Only one (TM1275) was more highly transcribed during exponential phase, while the other twelve were more highly transcribed during stationary phase. Three of these twelve (TM1309, TM1315, and TM1316) occur in a genomic locus (TM1298-TM1336) which was chosen for further analysis.

Transcription of the ‘TM1300s’ region. During growth on 1 g/L YE, as mentioned above, the transition to stationary phase was associated with a large fraction of the cell population becoming bloated and coccoidal **(Figure 2)**; this was not observed during growth on 5 g/L YE. Only 12 ORFs were both up-regulated in 1 g/L YE stationary phase vs. 5 g/L stationary phase *and* 1 g/L stationary phase vs. 1 g/L exponential phase. Three of these ORFs are the xylobiose-binding protein of an ABC transporter (TM0071) (31), TM0816, and TM0817 (discussed above). The remaining nine are in a particular genomic locus (TM1298-TM1336 or ‘TM1300s’) **(Table 5)** that was previously found to be responsive when *T. maritima* was grown on spent medium from another hyperthermophile, *Pyrococcus furiosus* (26). All nine TM1300s ORFs that were induced are found in four predicted operons, each of which has a similar organization consisting of one or more small ORFs (31-79 amino acids), followed by one or two radical SAM proteins, followed by multiple ABC-2-type transporter proteins or hypothetical proteins **(Figure 3)**. Each of these four predicted operons is separated by putative transcriptional regulators and toxin/antitoxin pairs. When the ABC-2

transporter family was first described (37), known members included drug resistance transporters and capsular polysaccharide transporters. All five of the radical SAM proteins in the TM1300s region (TM1301, TM1317, TM1324, TM1325, and TM1334) have a Cx_3Cx_2C motif in the N-terminal portion, which is very highly conserved among nearly all radical SAM proteins (41) **(Figure 4)**. This motif coordinates a [4Fe-4S] center, which donates an electron to form a 5'-deoxyadenosyl radical from cleavage of *S*-adenosylmethionine (SAM) (28). One of the small ORFs in this region (TM1316) was the most highly transcribed gene in the genome during stationary phase on 1 g/L YE, the gene with the highest response (27-fold) on 1 g/L vs. 5 g/L YE during stationary phase, and the gene with the highest response (35-fold) in stationary phase vs. exponential phase during growth on 1 g/L YE.

Production and localization of TM1316. To gain further insight into TM1316 and track its production in *T. maritima*, rabbit antibodies were raised against a synthetic version of the peptide. Using anti-TM1316 serum, a single band well below 20 kDa was detected by western blot, consistent with the predicted size of TM1316 (3.2 kDa) **(Figure 5A)**. TM1316 was not strongly detected during growth on 1 g/L YE until stationary phase, and the amount of TM1316 increased significantly as stationary phase progressed **(Figure 1)**. During growth on 5 g/L YE, much lower levels of TM1316 were detected in early stationary phase, and TM1316 signal disappeared as stationary phase progressed. Further investigation showed that TM1316 production progressively decreased as YE concentration was increased **(Figure 6)**, and TM1316 production was also repressed on media containing 1 g/l YE + 5 g/l tryptone.

To further investigate the regulation of TM1316 and the rest of the TM1300s region, two proteins in this region with identified helix-turn-helix domains (TM1314 and TM1330)

were cloned and expressed in *E. coli*. Expression of TM1330, but not TM1314, resulted in soluble protein. Un-tagged TM1330 could be co-purified with 6x His-tagged TM1331 by affinity chromatography (results not shown), suggesting that these proteins interact. This observation is consistent with the prediction that these two proteins comprise a toxin-antitoxin pair (40). However, sequence-specific DNA-binding activity of TM1330 was not observed by electrophoretic mobility shift assay, and the mechanisms of transcriptional regulation in this region remain unknown.

For initial attempts to detect TM1316 in *T. maritima*, cultures were fractionated and concentrated to prepare extracellular supernatant, soluble lysate, and insoluble lysate samples. It was subsequently discovered that TM1316 could be readily detected in *T. maritima* culture without any concentration, lysis, or processing; culture samples could be simply spotted onto nitrocellulose for dot blot analysis. If the cells were removed by centrifugation or filtration, most (if not all) TM1316 signal disappeared **(Figure 5B)**. TM1316 was not released by a freeze-thaw method previously used for examination of maltose binding activity (29), suggesting that it is not located in the periplasm or the outer membrane. When cells were lysed, TM1316 was distributed approximately equally between the soluble and insoluble fractions **(Figure 5B)**. Each fraction clearly exhibited less TM1316 signal than the whole cell lysate or the intact cells. All of this evidence taken together suggests that TM1316 is either associated with the cytoplasmic membrane or an extracellular matrix/biofilm, and it is partially released from this membrane/matrix when the cells are lysed. This hypothesis is supported by **Figure 5C**, which shows that TM1316 antibodies revealed that the peptide was closely associated with cells that had formed aggregates during stationary phase growth on 1 g/L YE.

DISCUSSION

The small ORFs in the TM1300s region have no detectable homology to proteins outside of the Thermotogales. The radical SAM enzymes, which are adjacent to the small ORFs, do have conserved motifs identifiable in similar types of enzymes encoded in other genomes. In addition to the N-terminal Cx_3Cx_2C motif that is present in nearly all members of the radical SAM superfamily, four of the proteins in the TM1300s exhibit seven cysteines with fairly conserved spacing near the C-terminus that presumably coordinate another Fe-S cluster **(Figure 4)**. The fifth protein (TM1334) has five of these seven cysteines with slightly different spacing. The subset of radical SAM enzymes with such a C-terminal domain is described by the hidden Markov model (HMM) TIGR04085 (16). These proteins usually occur in genomes near small peptides, suggesting that they are involved in peptide modification.

Unfortunately, very few of these radical SAM proteins have been characterized biochemically. PqqE is a radical SAM enzyme that is involved in pyrroloquinoline quinone (PQQ) biosynthesis (44). Rv0693 from *Mycobacterium tuberculosis*, another such enzyme, is involved in the maturation of mycofactocin, a putative electron carrier (15). In *B. subtilis*, the radical SAM protein AlbA has been shown to contribute to the maturation of the anti-listerial bacteriocin subtilosin A by catalyzing the formation of three linkages between cysteine sulfurs and amino acid α-carbons (11). Similar radical SAM enzymes are presumed to be involved in the production of the sporulation killing factor in *B. subtilis* and the bacteriocins thuricin CD and thurincin H from *B. thuringiensis* (24, 36). To further speculate on the function of radical SAM proteins in the TM1300s, it is important to consider the neighboring ORFs. PqqE and Rv0693-like radical SAM enzymes often co-occur with co-

factor dependent enzymes, while genes involved in bacteriocin synthesis generally co-occur with transporters. From this perspective, the genomic organization of the TM1300s is clearly more consistent with that of known bacteriocin production genes. However, because these types of radical SAM enzymes and their neighboring genes are widespread and not well characterized, there are likely more non-bacteriocin roles to be found (16). It is also worth noting that the TM1300s region does not include an identifiable protease for leader peptide cleavage that is common among bacteriocins.

The small peptide examined here (TM1316) does have similar spacing of cysteines and similar length compared to subtilosin A **(Figure 7)**. However, TM1316 is apparently primarily localized to the cell membrane **(Figure 5C)**, and it is unclear whether this supports or detracts from the hypothesis that this peptide is a bacteriocin. Most characterized bacteriocins have been purified from cell-free supernatants, but it does seem plausible that a bacterium might direct a bacteriocin to its membrane, where that bacteriocin either awaits a signal or is slowly released from the cell. To test for bactericidal activity, concentrated cell-free supernatants from *T. maritima* were spiked into log phase cultures of *T. maritima*, *T. neapolitana*, and the hyperthermophilic archaeon *Pyrococcus furiosus*, but no effects on growth were observed. Because TM1316 was generally not detected in cell-free supernatants, we also spiked concentrated cell lysate containing TM1316 into log phase *T. maritima* cultures, but again no effect was observed. Alternatively, is it possible that under these conditions, TM1316 was primarily detected in the membranes of cells targeted for cannibalization, in a manner similar to the sporulation killing factor from *B. subtilis* (14). If this is the case, it would be interesting to determine whether or not TM1316 is primarily detected in the coccoidal cells that were observed during stationary phase on 1 g/L YE. However, there is no evidence at present to support this theory and it is more likely that the

coccoid morphology can be attributed to nutritional limitations. TM1316 might also be an immunity protein for a bacteriocin encoded elsewhere in the TM1300s region, or serve as a modifier of other membrane proteins or signal transduction systems.

Nevertheless, the TM1300s region is intriguing because of its repetitive organization, including four loci made up of small hypotheticals, radical SAM proteins, and transporters. We have been unable to locate a similar genomic region consisting of more than one or two loci in other species outside of the Thermotogales. There are number of interesting small ORFs in the TM1300s region, and it seems highly possible that at least one of them (if not TM1316) is a bacteriocin. More detailed study of the region would benefit from more accurate identification of small ORFs that are transcribed into protein. Nine ORFs less than 100 amino acids were originally annotated in this region (four of which appear to have signal peptides), although we have identified at least 12 more possible ORFs **(Table 6)**. Certainly, some of these possible ORFs may not be translated, but others look promising and are preceded by possible ribosome binding sites. One of the originally annotated ORFs (TM1299) is nearly identical to one that was not annotated (TM1314-1) **(Table 6)**, so it seems likely that either both of these ORFs are translated or neither of them is. Limited tiling array data are publicly available for *T. maritima* (GEO series accession GSE28822), and they were gathered under growth conditions where transcription from the TM1300s region is low. Despite this limitation, the tiling array data do provide evidence for the transcription of two unannotated ORFs identified here (TM1323-1 and TM1333-1) **(Table 6)**. TM1323-1 (55 aa) is especially interesting because it contains 11 cysteines, and there are three other similar ORFs nearby that each contain at least 7 cysteines. A number of antimicrobial peptides have similarly cysteine-rich sections, including defensins (6) and lantibiotics (2). However, comparative genomic analysis of another uncharacterized group of cysteine-rich

peptides, the SCIFF peptides (six cysteines in forty-five residues), suggests that they might not be phermomones or bacteriocins (16), so further investigation is needed to determine the function of small proteins in the TM1300s region.

Conclusion. Efforts to determine the role of specific ORFs, and small ORFs in particular, in the TM1300s region are ongoing. However, as indicated **Tables 1 and 4**, microbial genomes encode many small ORFs, which, as is the case for *T. maritima*, are transcribed and potentially translated. Experimental and bioinformatics tools for studying the roles of these small ORFs are currently limited, as is knowledge about growth conditions that trigger their activation. However, the roles of these small ORFs in microbial physiology and ecology merit further attention as insights about how microorganisms interact with other organisms could be forthcoming.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy GTL Program (DG-FG02-08ER64687) and the National Science Foundation (CBET-0730091). ADF acknowledges support from NIH Pre-doctoral Biotechnology Traineeships (NIH T32 GM008776-06). Helpful discussions with T. Klaenhammer (North Carolina State University), K. Noll (University of Connecticut) and D. Rodionov (Sanford-Burnham Medical Research Institute, La Jolla, CA) are also acknowledged.

REFERENCES

1. **Alix, E., and A. B. Blanc-Potard.** 2009. Hydrophobic peptides: novel regulators within bacterial membrane. Mol Microbiol **72:**5-11.

2. **Banerjee, S., and J. N. Hansen.** 1988. Structure and expression of a gene encoding the precursor of subtilin, a small protein antibiotic. J Biol Chem **263:**9508-14.

3. **Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-12.

4. **Basrai, M. A., P. Hieter, and J. D. Boeke.** 1997. Small open reading frames: beautiful needles in the haystack. Genome Res **7:**768-71.

5. **Belkin, S., C. O. Wirsen, and H. W. Jannasch.** 1986. A new sulfur-reducing, extremely thermophilic eubacterium from a submarine thermal vent. Appl Environ Microbiol **51:**1180-1185.

6. **Bensch, K. W., M. Raida, H. J. Magert, P. Schulz-Knappe, and W. G. Forssmann.** 1995. hBD-1: a novel beta-defensin from human plasma. FEBS Lett **368:**331-5.

7. **Boucher, N., and K. M. Noll.** 2011. Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of *Thermotoga maritima* MSB8 screened by differential scanning fluorimetry. Appl. Environ. Microbiol. **77:**6395-6399.

8. **Chhabra, S. R., K. R. Shockley, S. B. Conners, K. L. Scott, R. D. Wolfinger, and R. M. Kelly.** 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. J. Biol. Chem. **278:**7540-7552.

9. **Chhabra, S. R., K. R. Shockley, D. E. Ward, and R. M. Kelly.** 2002. Regulation of endo-acting glycosyl hydrolases in the hyperthermophilic bacterium *Thermotoga maritima* grown on glucan- and mannan-based polysaccharides. Appl. Environ. Microbiol. **68:**545-554.

10. **Eguchi, Y., E. Ishii, M. Yamane, and R. Utsumi.** 2012. The connector SafA interacts with the multi-sensing domain of PhoQ in Escherichia coli. Mol Microbiol **85:**299-313.

11. **Fluhe, L., T. A. Knappe, M. J. Gattner, A. Schafer, O. Burghaus, U. Linne, and M. A. Marahiel.** 2012. The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A. Nat Chem Biol **8:**350-7.

12. **Frock, A. D., S. R. Gray, and R. M. Kelly.** 2012. Hyperthermophilic Thermotoga species differ with respect to specific carbohydrate transporters and glycoside hydrolases. Appl Environ Microbiol **78:**1978-86.

13. **Gassel, M., T. Mollenkamp, W. Puppe, and K. Altendorf.** 1999. The KdpF subunit is part of the K(+)-translocating Kdp complex of Escherichia coli and is responsible for stabilization of the complex in vitro. J Biol Chem **274:**37901-7.

14. **Gonzalez-Pastor, J. E., E. C. Hobbs, and R. Losick.** 2003. Cannibalism by sporulating bacteria. Science **301:**510-3.

15. **Haft, D. H.** 2011. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. BMC Genomics **12:**21.

16. **Haft, D. H., and M. K. Basu.** 2011. Biological systems discovery *in silico*: Radical *S*-adenosylmethionine protein families and their target peptides for posttranslational modification. J Bacteriol **193:**2745-2755.

17. **Hobbie, J. E., R. J. Daley, and S. Jasper.** 1977. Use of nuclepore filters for counting bacteria by fluorescence microscopy. Appl Environ Microbiol **33:**1225-8.

18. **Hobbs, E. C., F. Fontaine, X. Yin, and G. Storz.** 2011. An expanding universe of small proteins. Curr Opin Microbiol **14:**167-73.

19. **Huber, R., T. A. Langworthy, H. Konig, M. Thomm, C. R. Woese, U. B. Sleytr, and K. O. Stetter.** 1986. *Thermotoga maritima* sp-nov represents a new genus of unique extremely thermophilic eubacteria growing up to 90 degrees C. Arch. Microbiol. **144:**324-333.

20. **Ikeuchi, Y., N. Shigi, J. Kato, A. Nishimura, and T. Suzuki.** 2006. Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. Mol Cell **21:**97-108.

21. **Johnson, M. R., C. I. Montero, S. B. Conners, K. R. Shockley, S. L. Bridger, and R. M. Kelly.** 2005. Population density-dependent regulation of exopolysaccharide formation in the hyperthermophilic bacterium *Thermotoga maritima*. Mol. Microbiol. **55:**664-674.

22. **Kaletta, C., and K. D. Entian.** 1989. Nisin, a peptide antibiotic: cloning and sequencing of the nisA gene and posttranslational processing of its peptide product. J Bacteriol **171:**1597-601.

23. **Le Fourn, C., M. L. Fardeau, B. Ollivier, E. Lojou, and A. Dolla.** 2008. The hyperthermophilic anaerobe *Thermotoga maritima* is able to cope with limited amount of oxygen: insights into its defence strategies. Environ Microbiol **10:**1877-87.

24. **Lee, H., J. J. Churey, and R. W. Worobo.** 2009. Biosynthesis and transcriptional analysis of thurincin H, a tandem repeated bacteriocin genetic locus, produced by Bacillus thuringiensis SF361. FEMS Microbiol Lett **299:**205-13.

25. **Magnuson, R., J. Solomon, and A. D. Grossman.** 1994. Biochemical and genetic characterization of a competence pheromone from B. subtilis. Cell **77:**207-16.

26. **Montero, C.** 2005. Molecular microbial ecology of the Hyperthermophilic bacterium *Thermotoga maritima*: Transcriptional and physiological response to antibiotic challenge and inter-species interactions. . NCSU, Raleigh.

27. **Montero, C. I., D. L. Lewis, M. R. Johnson, S. B. Conners, E. A. Nance, J. D. Nichols, and R. M. Kelly.** 2006. Colocation of genes encoding a tRNA-mRNA hybrid and a putative signaling peptide on complementary strands in the genome of the hyperthermophilic bacterium *Thermotoga maritima*. J. Bacteriol. **188:**6802-6807.

28. **Mulliez, E., D. Padovani, M. Atta, C. Alcouffe, and M. Fontecave.** 2001. Activation of class III ribonucleotide reductase by flavodoxin: a protein radical-driven electron transfer to the iron-sulfur center. Biochemistry **40:**3730-6.

29. **Nanavati, D., K. M. Noll, and A. H. Romano.** 2002. Periplasmic maltose- and glucose-binding protein activities in cell-free extracts of *Thermotoga maritima*. Microbiology **148:**3531-3537.

30. **Nanavati, D. M., T. N. Nguyen, and K. M. Noll.** 2005. Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*. J. Bacteriol. **187:**2002-2009.

31. **Nanavati, D. M., K. Thirangoon, and K. M. Noll.** 2006. Several archaeal homologs of putative oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. Appl. Environ. Microbiol. **72:**1336-1345.

32. **Novichkov, P. S., O. N. Laikova, E. S. Novichkova, M. S. Gelfand, A. P. Arkin, I. Dubchak, and D. A. Rodionov.** 2010. RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. Nucleic Acids Res. **38:**D111-8.

33. **Okusu, H., D. Ma, and H. Nikaido.** 1996. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of Escherichia coli multiple-antibiotic-resistance (Mar) mutants. J Bacteriol **178:**306-8.

34. **Pysz, M. A., D. E. Ward, K. R. Shockley, C. I. Montero, S. B. Conners, M. R. Johnson, and R. M. Kelly.** 2004. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8:**209-217.

35. **Ravcheev, D. A., X. Li, H. Latif, K. Zengler, S. A. Leyn, Y. D. Korostelev, A. E. Kazakov, P. S. Novichkov, A. L. Osterman, and D. A. Rodionov.** 2012.

Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex. J Bacteriol **194:**1145-57.

36. **Rea, M. C., C. S. Sit, E. Clayton, P. M. O'Connor, R. M. Whittal, J. Zheng, J. C. Vederas, R. P. Ross, and C. Hill.** 2010. Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. Proc Natl Acad Sci U S A **107:**9352-7.

37. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1992. A new subfamily of bacterial ABC-type transport systems catalyzing export of drugs and carbohydrates. Protein Sci **1:**1326-32.

38. **Rodionova, I. A., D. A. Scott, N. V. Grishin, A. L. Osterman, and D. A. Rodionov.** 2012. Tagaturonate-fructuronate epimerase UxaE, a novel enzyme in the hexuronate catabolic network in Thermotoga maritima. Environ Microbiol **14:**2920-34.

39. **Rodionova, I. A., C. Yang, X. Li, O. V. Kurnasov, A. A. Best, A. L. Osterman, and D. A. Rodionov.** 2012. Diversity and versatility of the Thermotoga maritima sugar kinome. J Bacteriol **194:**5552-63.

40. **Sevin, E. W., and F. Barloy-Hubler.** 2007. RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. Genome Biol **8:**R155.

41. **Sofia, H. J., G. Chen, B. G. Hetzler, J. F. Reyes-Spindola, and N. E. Miller.** 2001. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Res **29:**1097-106.

42. **Stockdreher, Y., S. S. Venceslau, M. Josten, H. G. Sahl, I. A. Pereira, and C. Dahl.** 2012. Cytoplasmic sulfurtransferases in the purple sulfur bacterium Allochromatium vinosum: evidence for sulfur transfer from DsrEFH to DsrC. PLoS One **7:**e40785.

43. **Waters, L. S., M. Sandoval, and G. Storz.** 2011. The Escherichia coli MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. J Bacteriol **193:**5887-97.

44. **Wecksler, S. R., S. Stoll, H. Tran, O. T. Magnusson, S. P. Wu, D. King, R. D. Britt, and J. P. Klinman.** 2009. Pyrroloquinoline quinone biogenesis: demonstration that PqqE from Klebsiella pneumoniae is a radical S-adenosyl-L-methionine enzyme. Biochemistry **48:**10151-61.

45. **Yang, C., D. A. Rodionov, I. A. Rodionova, X. Li, and A. L. Osterman.** 2008. Glycerate 2-kinase of *Thermotoga maritima* and genomic reconstruction of related metabolic pathways. J Bacteriol **190:**1773-82.

46. **Zdobnov, E. M., and R. Apweiler.** 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17:**847-8.

47. **Zheng, G., L. Z. Yan, J. C. Vederas, and P. Zuber.** 1999. Genes of the sbo-alb locus of *Bacillus subtilis* are required for production of the antilisterial bacteriocin subtilosin. J Bacteriol **181:**7346-55.

Table 4.1. Distribution of ORFs annotated as "hypothetical proteins" as a function of encoded number of amino acids in selected microbial genomes (based on NCBI).

	Predicted ORFs within amino acid count range (Total) and those annotated as 'hypothetical' (Hyp)													
	1-100		**101-200**		**201-300**		**301-400**		**401-500**		**501-600**		**>600**	
Microbe	**Total**	**Hyp**	**Total**	**Hyp**	**Total**	**Hyp**	**Total**	**Hyp**	**Total**	**Hyp**	**Total**	**Hyp**	**Total**	**Hyp**
T. maritima MSB8	167	113	382	209	453	178	383	108	237	67	87	23	149	35
P. furiosus DSM 3638	286	210	545	330	502	228	395	140	225	67	64	18	108	24
C. saccharolyticus DSM 8903	233	143	631	289	588	142	512	125	341	59	159	17	215	34
B. anthracis A0248	860	692	1239	570	1159	341	817	163	510	58	217	19	239	26
H. pylori J99	123	77	331	165	349	151	272	79	189	50	77	14	147	24
E. coli K-12 MG1655	423	20	887	1	949	0	785	0	539	0	224	0	338	0
P. aeruginosa PAO1	331	237	1259	788	1357	534	1073	346	756	182	294	73	501	129
TOTAL	2423	1492	5274	2352	5357	1574	4237	961	2797	483	1122	164	1697	272
% HYPOTHETICAL	**61.6%**		**44.6%**		**29.4%**		**22.7%**		**17.3%**		**14.6%**		**16.0%**	

Table 4.2. Total ORFs differentially transcribed for *Thermotoga maritima* for growth phase transition and medium composition		
	↑ Exponential vs. Stationary	↑ Stationary vs. Exponential
1 g/l yeast extract (YE)	173	81
5 g/l yeast extract (YE)	187	172
Both 1 g/l and 5 g/l YE	80	58
	↑ 1 g/l vs. 5 g/l YE	↑ 5 g/l vs. 1 g/l YE
Exponential phase	18	4
Stationary phase	70	45
Both Exponential and Stationary	3	0

Table 4.3. Selected ORFs in *T. maritima* responding to 1 and 5 g/L yeast extract (YE) during exponential and stationary phase. (E, S = exponential, stationary phase; 1, 5 = 1, 5 g/L YE)

Locus tag	Function/Products	Average fold change*			
		1E vs. 5E	1S vs. 5S	1S vs. 1E	5S vs. 5E
TM1429-TM1431	glycerol utilization	5.9			6.5
TM1288-TM1292	MinD superfamily, radical SAM	2.1			
TM1792-TM1802	CRISPR-associated proteins	2.1		-2.2	
TM0733-TM0736	mannose-6-phosphate isomerase			-2.6	
TM0063-TM0069	uronate utilization			-3.5	-4.5
TM0868-TM0870	cell division and redox			-2.8	-2.5
TM1695-TM1699	protein synthesis/processing			-2.3	-2.3
TM1659-TM1666	SurE, peptide deformylase			-2.2	-2.1
TM0087-TM0092	unknown			-2.7	-2.3
TM0272-TM0275	central carbon metabolism				-3.0
TM1200-TM1204	maltose ABC transporter				-2.1
TM0101-TM0105	sugar ABC transporter				-2.4
TM0785-TM0786	ferritin and encapsulin				-3.8
TM1468-TM1505	ribosomal proteins				-3.2
TM0024-TM0032	beta-glucoside utilization		2.7		-3.1
xylE2F2K2	glucose/xylose ABC transporter		3.1		-4.3
treEFG	trehalose transporter		3.4		-3.5
TM0979-TM0983	intracellular sulfur oxidation		2.2		
TM0816-TM0817	multidrug efflux		5.1	3.8	
TM0430-TM0432	digalacturonate ABC transporter			3.7	5.2
TM0402-TM0403	ammonium assimilation			6.8	9.3
TM0441-TM0443	mannonate to gluconate-6-P pathway			2.1	2.7
TM0009-TM0012	NADP-reducing hydrogenase			2.1	2.9
TM0081-TM0086	flagella			2.0	3.4
TM0394-TM0398	amino acid metabolism			2.1	2.8
TM1120-TM1122	glycerol-3-P ABC transporter				3.7
TM1609-TM1617	ATP synthase				2.6
TM1243-TM1253	thiamine biosynthesis				2.4
TM1400-TM1401	serine metabolism				2.1
TM0548-TM0556	leucine biosynthesis				2.9
TM0729-TM0732	includes (p)ppGpp synthetase				2.1
TM0787-TM0788	thiamine biosynthesis				2.6
TM0132-TM0134	flagella				2.4
TM0849-TM0851	heat shock response				2.6
TM1538-TM1546	flagella		-2.1		2.1
TM0823-TM0825	TetR family, radical SAM		-2.3		3.4

*Log2(fold changes) were averaged for the indicated ORFs, then transformed to fold changes

Table 4.4. Transcription of hypothetical ORFs < 100 AA. Signal peptides (SP) predicted by SignalP 4.1 with "eukaryotes" selected as organism group.

			Least square mean (LSM)							Least square mean (LSM)			
ID	#AA	SP	1E	1S	5E	5S	ID	#AA	SP	1E	1S	5E	5S
TM0001	41	N					TM0999	32	N				
TM0002	41	N					TM1025	32	N				
TM0003	40	N					TM1107	80	Y				
TM0004	31	N					TM1108	78	N				
TM0026	68	N					TM1112	89	N				
TM0045	74	N					TM1186	68	N				
TM0046	39	N					TM1206	82	N				
TM0154	76	N					TM1208	70	N				
TM0170	73	N					TM1236	62	N				
TM0179	92	N					TM1242	30	N				
TM0210	69	N					TM1244	82	N				
TM0223	77	N					TM1266	82	N				
TM0242	40	N					TM1268	81	N				
TM0314	59	N					TM1275	87	N				
TM0315	83	N					TM1285	41	N				
TM0319	64	N					TM1298	89	N				
TM0338	35	N					TM1299	37	N				
TM0340	94	N					TM1300	79	Y				
TM0341	58	N					TM1305	43	N				
TM0357	68	N					TM1308	31	N				
TM0377	32	N					TM1309	68	N				
TM0391	31	N					TM1313	75	N				
TM0450	100	N					TM1315	57	Y				
TM0486	94	N					TM1316	31	N				
TM0504	42	N					TM1333	60	Y				
TM0518	64	N					TM1338	44	N				
TM0566	64	N					TM1386	69	N				
TM0611	56	N					TM1392	98	N				
TM0612	39	N					TM1412	47	N				
TM0617	30	N					TM1413	35	N				
TM0623	79	N					TM1420	77	N				
TM0636	69	N					TM1457	94	N				
TM0637	41	N					TM1462	81	N				
TM0693	90	N					TM1538	89	N				
TM0773	73	N					TM1567	75	N				
TM0781	71	N					TM1630	100	N				
TM0783	33	N					TM1649	93	N				
TM0826	56	N					TM1669	34	N				
TM0871	47	N					TM1671	31	N				
TM0898	76	N					TM1680	67	N				
TM0969	94	N					TM1681	71	N				
TM0970	41	N					TM1688	79	N				
TM0971	30	N					TM1690	92	N				
TM0974	96	N					TM1711	88	N				
TM0976	33	N					TM1712	88	N				
TM0977	66	N					TM1732	98	N				
TM0983	79	N					TM1786	59	N				
TM0994	37	Y					TM1829	30	N				
TM0995	38	N					TM1850	59	N				
TM0996	80	N					TM1871	91	N				

Table 4.5. Transcription of the TM1300s region. Gray shading indicates that the ORF changed less than 2-fold for the specific comparison.

Probe ID	Product	Strand	Fold changes			
			1E vs. 5E	1S vs. 5S	1S vs. 1E	5S vs. 5E
TM1298	hypothetical protein	-				
TM1299	hypothetical protein	+				
TM1300	hypothetical protein	+				
TM1301	radical SAM protein	+				
TM1302	ABC-2 transporter, ATP-binding protein	+		8.7	14.0	2.7
TM1303	ABC-2 transporter, permease protein	+		7.6	9.5	
TM1304	ABC-2 transporter, permease protein	+				
TM1305	hypothetical protein	+				
TM1306	ABC-2 transporter, permease protein	+				
TM1307	hypothetical protein	+		3.0	5.8	2.5
TM1308	hypothetical protein	+				
TM1309	hypothetical protein	+			2.2	
TM1310	ABC-2 transporter, ATP-binding protein	+		2.0	2.7	
TM1311	RASTA-predicted toxin	-				
TM1312	RASTA-predicted antitoxin	-				
TM1313	RASTA-predicted toxin, HicB-like	-				
TM1314	Contains HTH_XRE and TPR domains	-				
TM1315	hypothetical protein	+		3.6	4.5	
TM1316	hypothetical protein	+	6.1	27.3	34.7	7.7
TM1317	radical SAM protein	+		7.4	12.6	
TM1318	ABC-2 transporter, ATP-binding protein	+		7.1	5.4	
TM1319	ABC-2 transporter, ATP-binding protein	+		3.1		
TM1320	RASTA-predicted antitoxin	-				
TM1321	RASTA-predicted toxin, HicB-like	-				
TM1322	hypothetical protein	-				
TM1323-1333	hypothetical protein	+			3.3	3.4
TM1324	radical SAM protein	+	not on array			
TM1325	radical SAM protein	+	not on array			
TM1326	ABC-2 transporter, permease protein	+			-2.5	-2.3
TM1327	ABC-2 transporter, ATP-binding protein	+	not on array			
TM1328	ABC-2 transporter, ATP-binding protein	+		6.0	3.6	
TM1329	hypothetical protein	+	not on array			
TM1330	RASTA-predicted antitoxin, HTH_XRE domain	+				
TM1331	RASTA-predicted toxin	+	not on array			
TM1332-1322	hypothetical protein	-				
TM1333	hypothetical protein	+				
TM1334	radical SAM protein	+				3.8
TM1335	hypothetical protein	+				
TM1336	major facilitator superfamily, bacilysin exporter-like	+			2.1	2.4

Table 4.6. Putative small ORFs in the TM1300s region. ORFs highlighted in gray are supported by microarray data from this study or tiling array data (GEO series accession GSE28822). ORFs that were not originally annotated have been given a unique ID by appending "-x" onto the nearest annotated ID. Underlined AAs are predicted signal peptides.

ID	Coordinates		#AA	pI	AA sequence
	Start	End			
TM1299	1324834	1324947	37	7.8	MVLSKVDDFSILSLQNYYFYAILSHCCVTLGVKLHKT
TM1300	1325051	1325290	79	9.3	**MKIVRRIILTIVILLSLLYTLVFP**SPSDDYQTTVRKKTVVHTLLLEKSHSEDLTAGNQILRFNKSLARLLISNERGDKG
TM1305	1329161	1329292	43	6.0	MKIGLVQAGGLVAVLVVAIFLEKSARGCYFAKISALSESADID
TM1307-1	1329812	1329913	33	10.1	MPLKRSSIRFFGCSFTQHFLHMDIKRCTMRELS
TM1308	1330202	1330297	31	4.4	MAALGLLTRDINMLLGVAEMMVFILSVQVFL
TM1309	1330309	1330515	68	9.5	MFFRWLSQLLPLSRSVESIIYALQGFKMKAWISSAWEAAVGLAYLVLAYFTLRLIEFAARKYGNIDMY
TM1314-1	1334887	1335000	37	7.8	MVLSKVDDFSILSLQNYYFYAILSHCCVTLGVKLHKI
TM1315	1335104	1335277	57	6.0	**MKIVRRIILTIVILLSLLYTLVFP**SPSDDYQTTVRKKTVVHTFLIEKLHSEDLTVGN
TM1315-1	1335040	1335186	48	10.0	MAFQNRKSLKDHKRSVLIRKGGENREKDHSHNSYSAFSLIHACVSQSF
TM1316	1335354	1335449	31	8.6	MQKMPKPVVPNDYLNPGKGCGGAGLVCGWCG
TM1316-1	1335410	1335547	45	12.4	MWRSWSRLWLVWLRKNSSRFFPFPLAKGQGKGFIKRLYFILPRRF
TM1322-1	1342532	1342696	54	6.0	MKSGAENVLLERILEQPKLENHIRFDKVMTGDQNAKGFDDNLFDTRSPVYNHFR
TM1323	1342623	1342829	68	5.6	**MIRMRRVLMIIFLILVLLSTIIFA**DKVKTDNETHSWKSEITEQVQVAPKSAATCEVTFKGSTAGNQSF
TM1323-1	1342934	1343101	55	3.7	MEFLNNPTIGKIGLE---CCQSNGNCSECFNPDCNGICDPS-CNAVCD-CNAL--CTLCIVP
TM1323-2	1342992	1343132	46	8.8	MGTVLSVSIRIVMGYVTLHAMRYVIVMLYAHCVLCHKKEGENLWSS
TM1323-3	1343120	1343236	38	3.9	MEFLNDPLEVFKPVGDIMCCDKNGNCD-CINTVCND-CACC
TM1323-4	1343262	1343432	56	3.6	MEFLNFPDNSLSPL----CCSTNPNCD-CWDLNCSG-CFDDMCAG-CLVCSPNANCTSCLAKN
TM1325-1	1346088	1346273	61	4.6	MEFLNDPVVDEKVYLSFCDSLKDRCGDRVGNCDECTNSHCVGLCNTECDHVPWWKRIFLWW
TM1332-1	1352560	1352724	54	8.1	MKSGAENVLLKRILEQPKLENHIRFDKVMTGDQNAKGFDDNLFDTRSPVYNHFR
TM1333	1352651	1352857	68	5.6	**MIRMRRVLMIIFLILVLLSTIIFA**DKVKTDNETHSWKSEITEQVQVAPKSAATCEVTFKGSTAGNQSF
TM1333-1	1353021	1353185	54	10.6	MAPCRCGKVPELLRRILILMARQTQGTKCKLFERVYIISFNILFRRSEKDRWLK

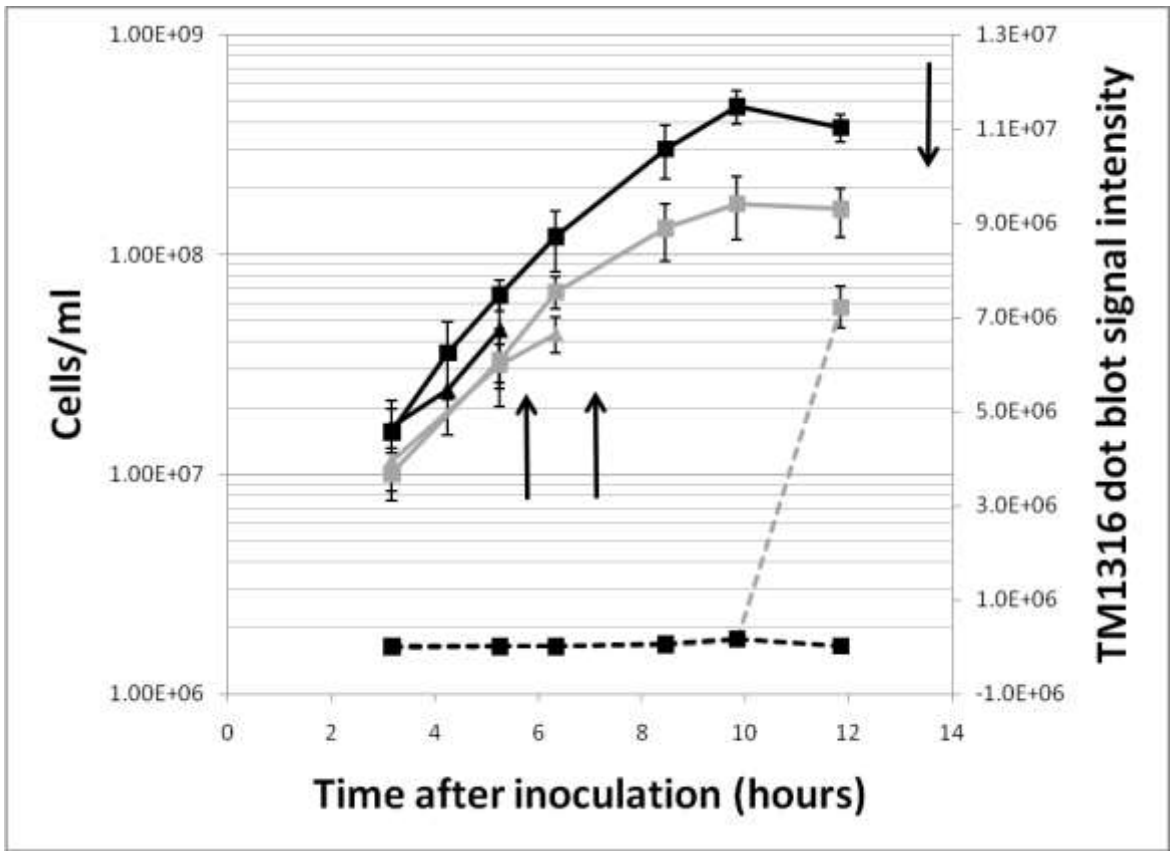

Figure 4.1. Growth of *T. maritima* on 1 g/L yeast extract (gray) and 5 g/L yeast extract (black). TM1316 concentrations were estimated by immunoblots (dotted lines). For each growth condition, cultures were harvested for transcriptomic analysis during exponential phase (triangles) and stationary phase (squares) at the times indicated by the black arrows. Note that TM1316 levels were significantly higher in stationary phase for 1 g/L yeast extract medium.

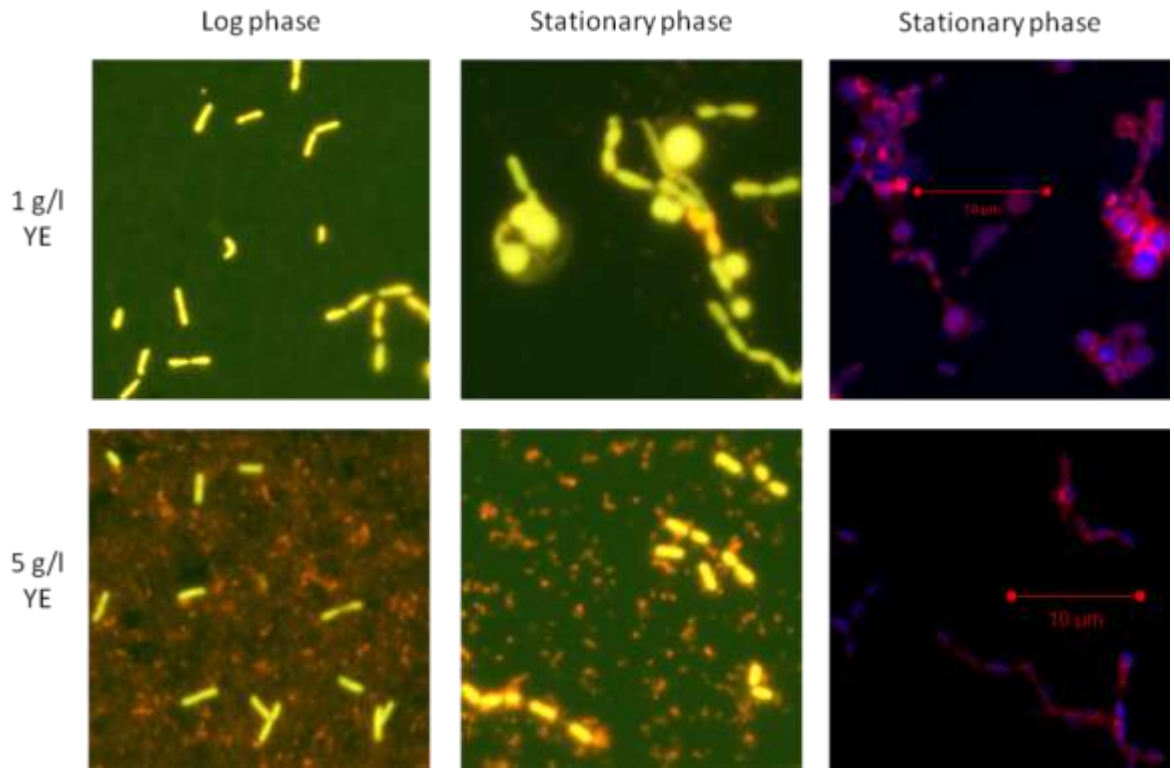

Figure 4.2. Fluorescence microscopy of *T. maritima* grown on media containing 1 g/L and 5 g/L yeast extract, during log phase and stationary phase. In the left and middle images, the cells were stained with acridine orange. In the images on the right, the cells were stained with DAPI and CellMask Deep Red plasma membrane stain (Invitrogen). The orange background apparent in the bottom left image was attributed to yeast extract.

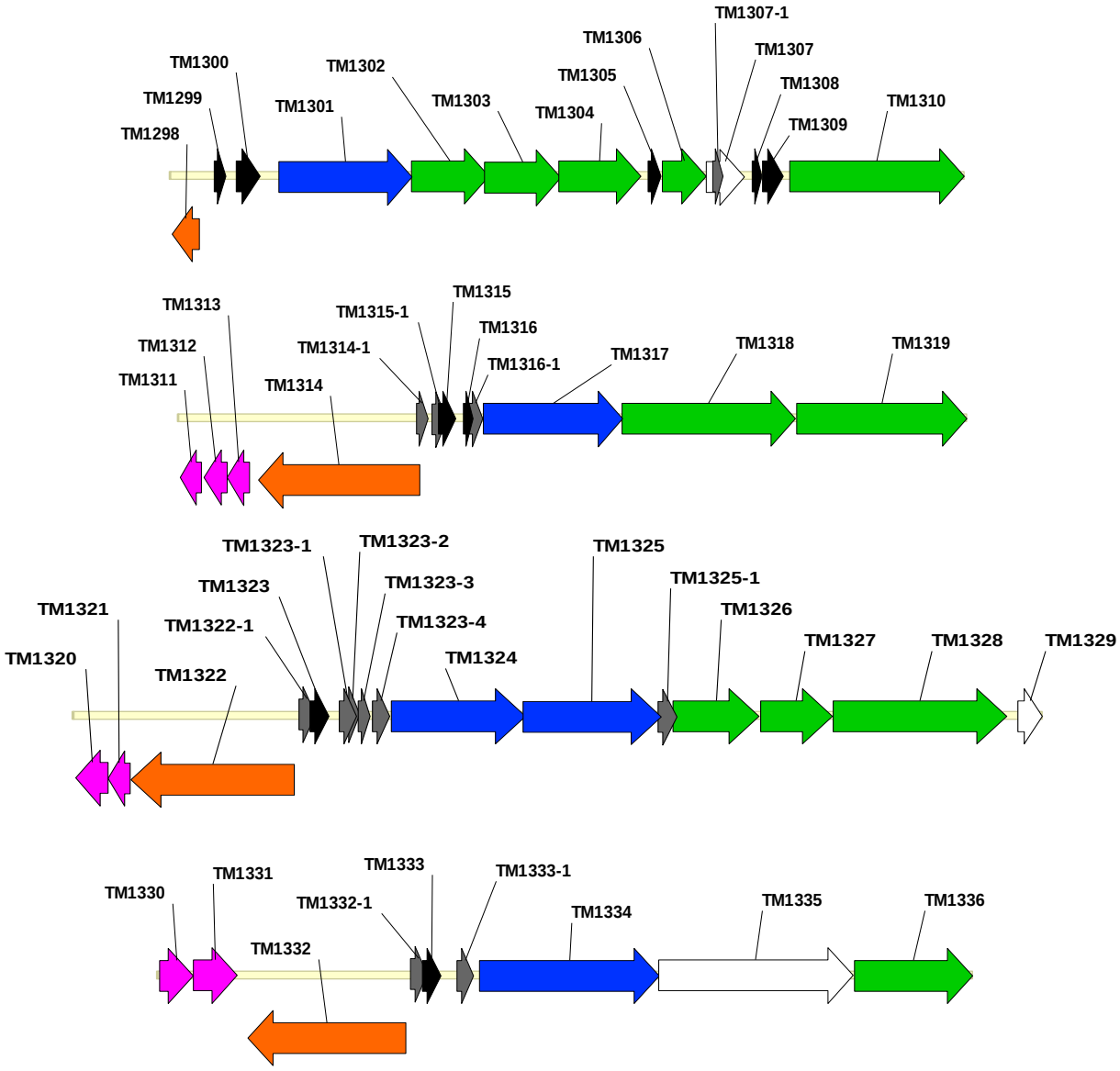

Figure 4.3. Genomic organization of the TM1300s region. Schematic generated using Vector NTI (Invitrogen). This region includes small hypothetical ORFs (black), radical SAM proteins (blue), transporters (green), toxins/antitoxins (pink), other hypothetical proteins (white), previously unannotated small ORFs (gray), and proteins with a highly similar section of ~80 amino acids long (orange), one of which (TM1314) appears to be a transcriptional regulator based on identifiable TPR and XRE family HTH domains.

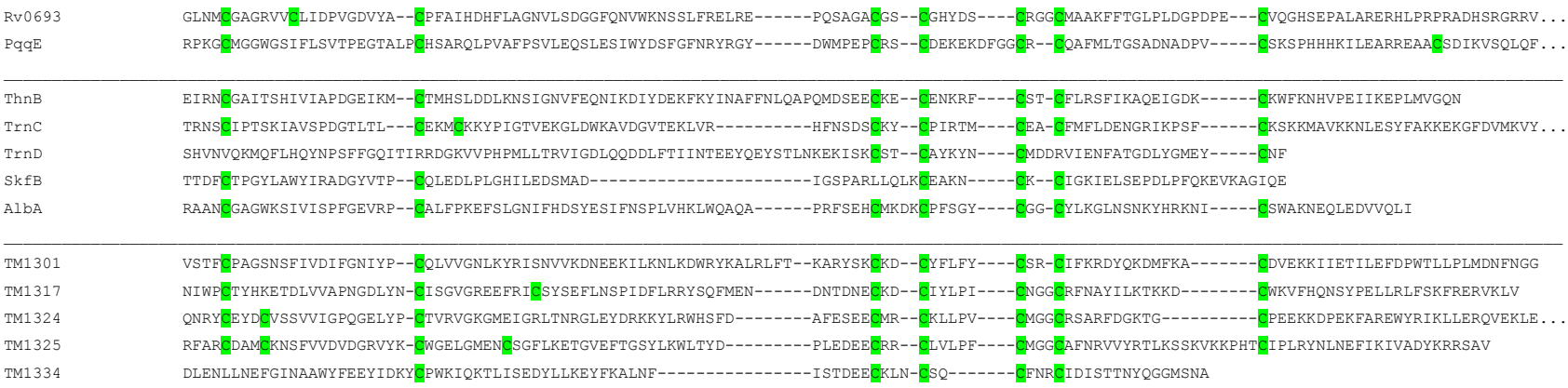

Figure 4.4. Alignment of conserved cysteines in the C-termini of radical SAM enzymes. The first two (Rv0693 from *Mycobacterium tuberculosis* and PqqE from *Klebsiella pneumonia*) are presumably not involved in bacteriocin synthesis. The next five (ThnB, TrnC, and TrnD are from *Bacillus thuringiensis*; SkfB and AlbA from *B. subtilis*) are involved in bacteriocin synthesis. The final five are putative radical SAM enzymes identified in the TM1300s region of the *T. maritima* genome.

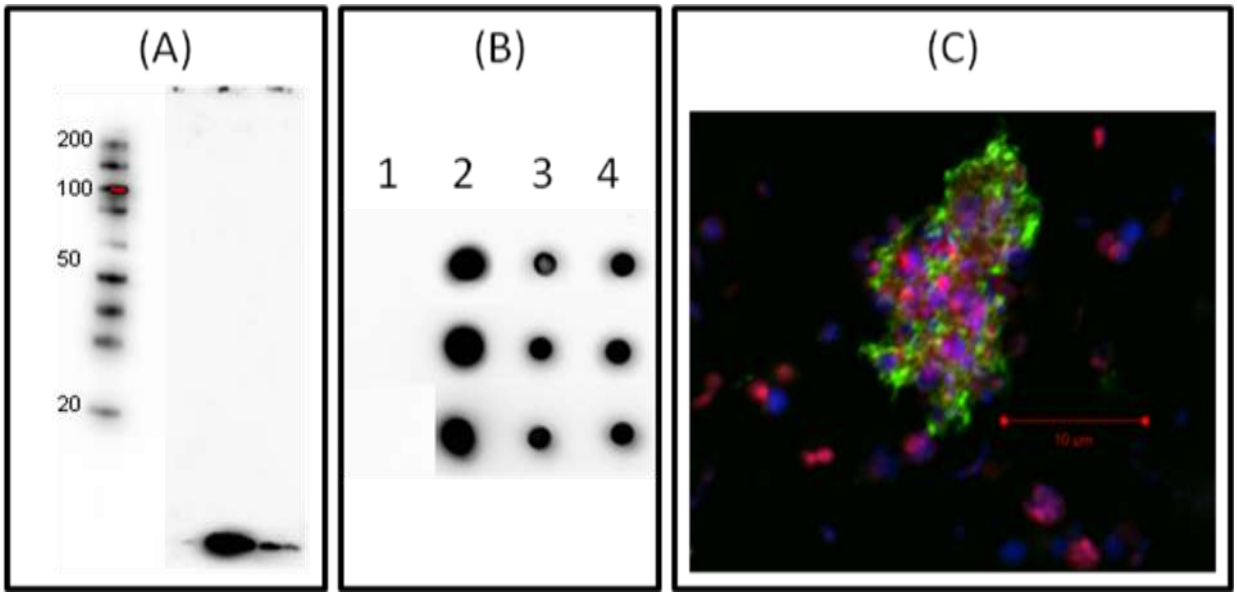

Figure 4.5. Detection of TM1316 in fractionated *T. maritima* cells. (A) A single band is detected in cell lysate by western blot. (B) Dot blot of cell fractions including: (1) cell-free culture supernatant, (2) whole-cell lysate, (3) lysate clarified by centrifugation, (4) lysate clarified by centrifugation and sterile filtration. (C) Detection of TM1316 associated with aggregated cells by immunofluorescence. Cells were stained with DAPI (blue) and CellMask Deep Red membrane stain (red) and incubated with rabbit anti-TM1316 serum and goat anti-rabbit IgG-DyLight 488 conjugate (green).

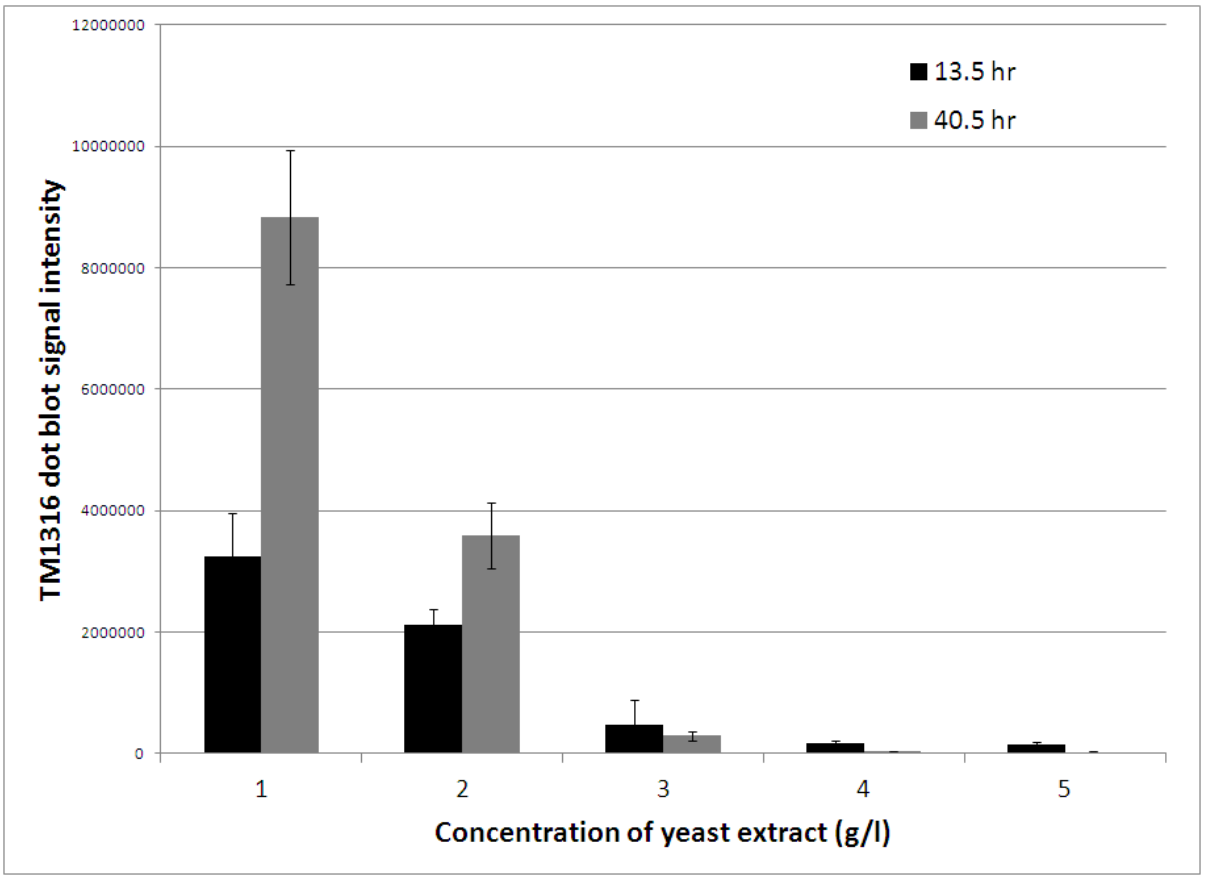

Figure 4.6. Production of TM1316 as a function of medium yeast extract concentration. TM1316 detected by immunoblot in unprocessed stationary phase culture broth after 13.5 and 40.5 hours of growth.

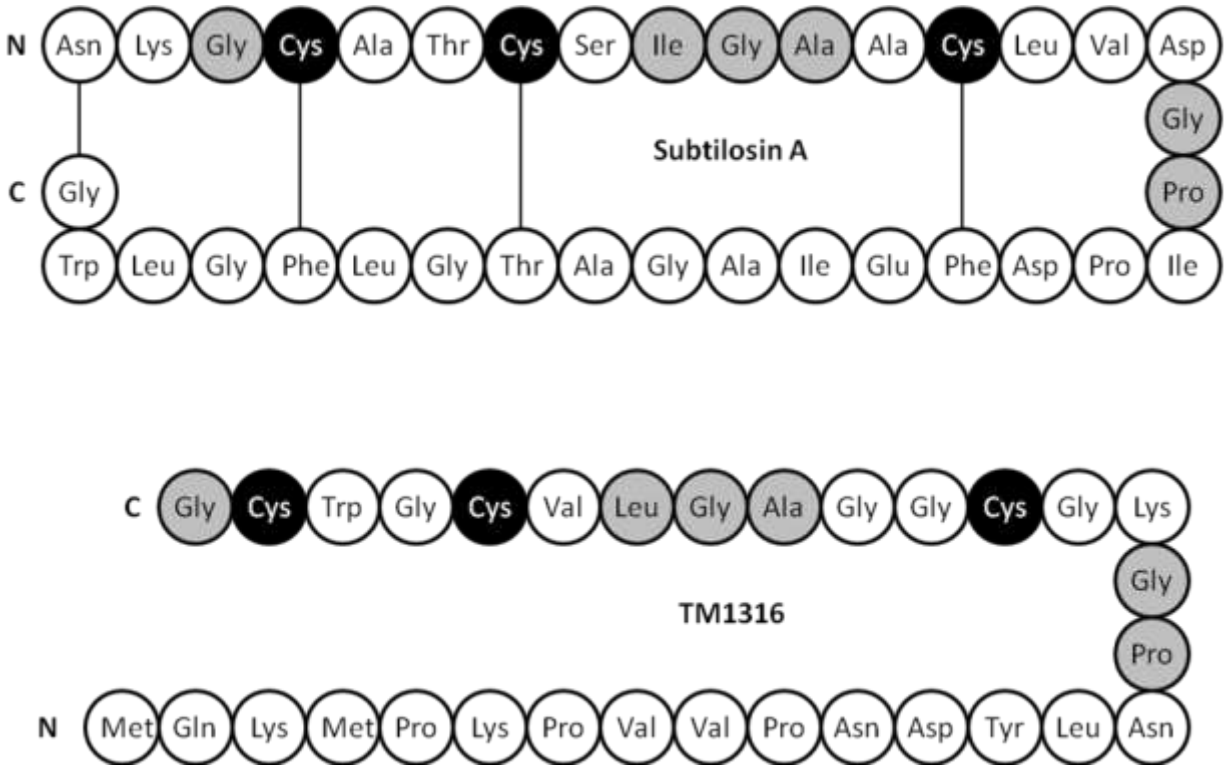

Figure 4.7. Comparison of TM1316 and the mature form of subtilosin A. Black lines indicate known intra-chain linkages in subtilosin A. Cysteines are highlighted in black, other similar or identical amino acids common to the two peptides are highlighted in gray.

Genes differentially transcribed during exponential and stationary growth of *T. maritima* on 1 and 5 g/L yeast extract

Table 4.A1. Differentially transcribed ORFs responding to medium YE concentration or growth phase transition

Probe ID	Product	Fold change			
		1E vs. 5E	1S vs. 5S	1S vs. 1E	5S vs. 5E
TM0006	muconate cycloisomerase				2.3
TM0007	transglutaminase/cysteine proteinase		-2.8		
TM0009	rubrerythrin/ferritin			2.1	2.7
TM0011	NADP-reducing hydrogenase, subunit B			2.5	3.4
TM0012	NADP-reducing hydrogenase, subunit A			4.8	5.8
TM0023	methyl-accepting chemotaxis protein			-3.4	-2.9
TM0024	laminarinase LamA		2.4		-2.8
TM0025	beta-glucosidase BglB			-2.7	-3.5
TM0028	beta-glucoside ABC transporter, ATP-binding protein BglK				-4.5
TM0029	beta-glucoside ABC transporter, permease protein BglG		3.8		-4.3
TM0031	beta-glucoside ABC transporter, periplasmic sugar-binding protein BglE	2.1	3.2	-2.6	-3.8
TM0032	cellobiose-responsive regulator of beta-glucoside utilization BglR		5.8		-4.4
TM0042	aminopeptidase P, putative			-2.4	-4.2
TRQ2_0901 (TM0044)	outer membrane protein, beta-barrel domain containing (opp strand of TM0044)		2.0		
TM0044	outer membrane protein, beta-barrel domain containing (opp strand of TM0044)		2.2		
TM0051	iron(II) transport protein B	-2.4	2.1		-5.6
TM0057	xylan oligosaccharide ABC transporter, ATP-binding protein XtpL			-2.4	
TM0058	xylan oligosaccharide ABC transporter, ATP-binding protein XtpK			-2.8	
TM0063	hypothetical protein			-6.1	-6.9
TM0064	hexuronate isomerase UxaC			-3.5	-2.7
TM0065	transcriptional regulator of glucuronate utilization KdgR			-5.5	-8.1
TM0067	2-keto-3-deoxygluconate kinase KdgK			-3.6	-5.0
TM0068	D-mannonate dehydrogenase, NADH-dependent UxuB				-2.6
TM0069	D-mannonate dehydratase UxuA			-3.8	-3.9
TM0071	xylobiose ABC transporter, periplasmic sugar-binding protein XloE		5.3	2.3	
TM0074	xylobiose ABC transporter, ATP-binding protein XloK			-2.1	
TM0081	flagellar assembly factor FliW				2.1
TM0082	flagellar hook-associated protein 3 FlgL				2.7
TM0084	FlgN protein			3.1	4.2
TM0085	FlgM anti-sigma-28 factor		-2.1	3.0	7.6
TM0086	proposed peptidoglycan lipid II flippase MurJ/virulence factor MviN-related protein		-3.4		7.2
TM0087	Memo-like protein			-4.8	-2.2
TM0088	Type II/III secretion system secretin-like protein			-4.1	-3.7
TM0090a	hypothetical protein			-2.2	-2.5
TM0090b	hypothetical protein			-2.5	-2.5
TM0091	hypothetical protein			-4.5	-3.3
TM0101	hypothetical protein				-2.5
TM0103	sugar ABC transporter, ATP-binding protein			-2.2	-3.6
TM0104	sugar ABC transporter, permease protein				-2.9
TM0107	diguanylate cyclase/phosphodiesterase-domain containing protein			2.2	2.1

TM0113	acetyl xylan esterase XylU		2.1		
TM0115	xylose ABC transporter, ATP-binding protein			-2.9	
TM0132	flagellin protein		-2.1		2.7
TM0133	isochorismatase-related protein				2.0
TM0146	ATP-dependent Clp protease, ATPase subunit clpX		-2.2		2.1
TM0150	LSU ribosomal protein L32			2.2	
TRQ2_0791 (upstream of TM0158)	Domain of unknown function DUF370				-2.6
TM0162	alpha/beta hydrolase			-2.1	
TM0166	folylpolyglutamate synthase/dihydrofolate synthase			-2.6	
TM0173	DNA reverse gyrase			-2.9	
TM0174	pyrophosphatase, proton-translocating			-2.3	-3.4
TRQ2_R0030 (TMrrnaA16S)	16S-rRNA			2.7	2.1
TM0179	hypothetical protein				2.1
TM0186	response regulator			-2.7	
TM0190	iron(III) ABC transporter, permease protein, putative		-2.3		
TM0199	DNA repair protein RadA		-2.0		2.0
TM0204	ABC transporter, ATP-binding protein			-4.3	-2.3
TM0216	glycyl-tRNA synthetase, alpha subunit				-2.1
TM0229	membrane transport protein				2.9
TM0231	UDP-N-acetylmuramate--alanine ligase			-2.3	
TM0232	UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase			-2.4	
TM0239	glucose-1-phosphate adenylyltransferase			-2.4	
TM0243a	biotin synthase-like MiaB/NifB radical SAM superfamily			-3.1	
TM0262	DNA polymerase III, beta subunit			-2.2	
TM0266	DNA-binding protein, HU		2.2		-2.2
TM0270	methylenetetrahydrofolate reductase				2.1
TM0272	pyruvate,orthophosphate dikinase				-3.7
TM0273	fructose-bisphosphate aldolase				-2.8
TM0274	acetate kinase				-2.7
TM0275	transcriptional regulator of arabinose utilization AraR			-2.7	-2.9
TM0282	L-arabinose epimerase AraM			-2.1	-2.8
TM0283	L-ribulose-5-phosphate epimerase AraD				-3.0
TM0284	ribulokinase AraB				-2.4
TM0295	transaldolase			2.3	
TM0297	oxidoreductase, short chain dehydrogenase/reductase family			-2.1	-2.5
TM0298	alcohol dehydrogenase, zinc-containing			-3.6	
TM0304	xyloglucan ABC transporter, ATP-binding protein			-2.4	
TM0308	alpha-xylosidase			-2.2	-2.3
TM0309	predicted xylose oligosaccharide ABC transporter, periplasmic substrate-binding protein XtpN				-2.3
TM0310	beta-D-galactosidase BgaL			-2.5	-2.4
TM0321	hypothetical protein			2.2	2.0
TM0330	metal-dependent hydrolase			-2.1	
TM0336	alpha/beta hydrolase			-2.5	
TM0338	hypothetical protein			2.1	
TM0344	prephenate dehydrogenase				2.7
TM0349	3-dehydroquinase dehydratase				-2.1

TM0364	4-alpha-glucanotransferase				-2.0
TM0365	M18-family aminopeptidase, putative				-2.7
TM0366	endonuclease III				-2.5
TM0369	putative regulatory domain (TIGR03879)	2.3			2.9
TM0373	dnaK protein			3.3	
TM0380	alkylhydroperoxidase-like protein AhpD			2.3	2.2
TRQ2_0545 (TM0380)	alkylhydroperoxidase-like protein AhpD			2.0	
TM0384	anaerobic ribonucleoside-triphosphate reductase class III				2.5
TM0390	Hypothetical protein			2.6	3.6
TM0392	trehalose synthase TreT			-6.1	-4.3
TM0394	glutamate synthase, alpha subunit			4.0	
TM0395	NADH oxidase, putative				2.4
TM0397	glutamate synthase, alpha subunit				3.4
TM0402	ammonium transporter			4.6	4.9
TM0403	nitrogen regulatory protein P-II			10.1	17.9
TM0408	protein-glutamate methylesterase, chemotaxis response regulator CheB			-3.4	-3.2
TRQ2_0523 (TM0422)	xylose/hydroxypyruvate isomerase-like			3.2	3.9
TM0422	xylose/hydroxypyruvate isomerase-like		-2.4		4.7
TM0423	glycerol dehydrogenase				3.3
TM0427	succinate dehydrogenase flavoprotein subunit			-2.8	
TM0430	digalacturonate ABC transporter, permease protein AguG			2.8	2.8
TM0431	digalacturonate ABC transporter, permease protein AguF			4.0	5.0
TM0432	digalacturonate ABC transporter, periplasmic substrate-binding protein AguE			4.6	10.0
TM0441	D-mannonate dehydrogenase, NADPH-dependent UxaD			2.6	2.7
TM0443	gluconate kinase GntK				2.2
TM0446	phosphoribosylaminoimidazole carboxylase, catalytic subunit				-2.5
TM0448	hypothetical protein			-2.9	-2.1
TM0456	LSU ribosomal protein L10			-2.6	-5.8
TM0457	LSU ribosomal protein L7/L12				-2.9
TM0459	DNA-directed RNA polymerase, beta' subunit				2.3
TM0462	Ribosomal large subunit pseudouridine synthase D				-2.4
TM0463	lipoprotein signal peptidase			-3.7	-3.9
TM0472	glutamine amidotransferase				2.3
TM0475	pyrazinamidase/nicotinamidase-related protein				2.1
TM0485	hydroxymethylpyrimidine ABC transporter, permease protein, putative				2.1
TM0486	Domain of unknown function DUF77				2.1
TM0501b	oligopeptide ABC transporter, ATP-binding protein				-2.0
TM0506	groEL protein				-2.1
TM0510	iron-dependent transcriptional repressor, putative		-2.0		2.2
TM0511	Uracil DNA glycosylase superfamily		-2.6		3.4
TM0517	hypothetical protein				2.7
TM0531	oligopeptide ABC transporter, periplasmic oligopeptide-binding protein		2.1		-2.1
TM0533	oligopeptide ABC transporter, permease protein			-2.6	-2.6
TM0537	P-loop containing nucleoside triphosphate hydrolase			-2.2	
TM0540	fumarate hydratase				2.3
TM0545	homoserine kinase			2.9	2.2
TM0546	threonine synthase				4.1

TM0548	acetolactate synthase, large subunit				3.3
TM0549	acetolactate synthase, small subunit			3.0	4.8
TM0551	dihydroxy-acid dehydratase				3.7
TM0552	2-isopropylmalate synthase, putative				2.9
TM0553	2-isopropylmalate synthase				3.8
TM0554	3-isopropylmalate dehydratase, large subunit		-2.1		3.1
TM0555	3-isopropylmalate dehydratase, small subunit			2.8	3.7
TM0560	ferritin		4.0		-5.6
TRQ2_0374 (upstream of TM0563)	Transmembrane protein with signal peptide				2.2
TM0565	sugar fermentation stimulation protein, putative		-2.1	-2.1	
TM0571	heat shock serine protease, periplasmic HtrA				3.7
TM0573	acyltransferase, putative			-2.4	
TM0582	radical SAM enzyme, MiaB/NifB family			-2.0	
TM0583	lipopolysaccharide biosynthesis protein			-2.0	
TM0584	conserved hypothetical protein			-2.2	
TM0585	lipopolysaccharide biosynthesis protein BplA			-2.4	-2.4
TRQ2_0346 (TM0589)	hypothetical protein			-2.3	-2.9
TM0589	hypothetical protein			-2.4	-3.6
TM0590	peptidoglycan synthetase MrdA				-2.5
TM0604	single stranded DNA-binding protein, putative				-2.1
TM0606	hypothetical protein			-2.7	-3.7
TM0628	hypothetical protein			-2.1	
TM0630	UDP-glucose epimerase, putative			-3.4	-2.6
TM0631	lipopolysaccharide biosynthesis protein			-2.4	
TM0651	HAD superfamily hydrolase				2.4
TM0656	HTH_19 and Cupin_2 containing protein			-2.2	-2.8
TM0661	hypothetical protein			-2.2	
TM0671	M-related protein		-3.3	2.5	9.2
TM0672	hypothetical protein			2.1	4.3
TM0681	dehydrase-related protein		-2.1		4.9
TM0686	DNA polymerase III, gamma and tau subunit				3.1
TM0687	conserved hypothetical protein		-2.0		4.7
TM0688	glyceraldehyde-3-phosphate dehydrogenase				2.1
TM0690	arpartyl/glutamyl-tRNA amidotransferase subunit B-related				2.3
TM0698	flagellar biosynthesis protein FliP			-2.7	
TM0704	L-isoaspartate(D-aspartate) O-methyltransferase			-2.0	
TM0710	transcriptional regulator, MarR family				-3.4
TM0713	hypothetical protein				3.2
TM0723	conserved hypothetical protein			-2.2	-2.5
TM_rnpB	ribonuclease P, RNA component			2.8	2.9
TM0729	(p)ppGpp synthetase				2.6
TM0730	D-tyrosyl-tRNA(Tyr) deacylase				2.2
TM0731	Domain of unknown function DUF208				2.1
TM0734	tRNA uracil-5-methyltransferase Gid			-2.5	
TM0735	conserved hypothetical protein			-3.1	
TM0736	mannose-6-phosphate isomerase			-2.4	
TM0738	hypothetical protein				2.1
TM0742	serine/threonine protein phosphatase			-2.8	

TM0750	hypothetical protein				2.4
TM0756	galactosyltransferase-related protein		-2.3	-2.5	
TM0767	maltodextrin glycosyltransferase				-2.7
TM0771	DNA polymerase III, gamma subunit-related protein			-2.0	
TM0777	transposase			-2.3	-2.8
TM0784	hypothetical protein			2.6	3.3
TM0785	Linocin_M18-like encapsulin/protease		2.1		-4.2
TM0787	thiamine biosynthetic enzyme			2.2	2.9
TM0788	thiamine biosynthesis protein ThiC				2.3
TM0798	malonyl CoA-acyl carrier protein transacylase			-5.5	-3.4
TM0801	(3R)-hydroxymyristoyl-(acyl carrier protein) dehydratase				-2.4
TM0802	3-oxoacyl-(acyl carrier protein) synthase II			-2.7	
TM0807	peroxiredoxin, alkyl hydroperoxide reductase, putative		4.1	-2.0	-9.4
TM0816	transcriptional regulator, putative, Mar family		8.2	5.8	
TM0817	AcrB multidrug efflux transporter transmembrane domain superfamily		3.2	2.5	
TM0818	lipopolysaccharide biosynthesis protein, WecB glycosyltransferase				-2.2
TM0820	NADH-dependent butanol dehydrogenase, putative			-3.0	-2.7
TM0823	transcriptional regulator, TetR family		-4.6		5.6
TM0824	radical SAM protein, astB/chuR-related		-2.1		3.9
TM0826	hypothetical protein			2.3	2.9
TM0830	possible tRNA modifying radical SAM MiaB family protein			-2.0	
TM0838	hypothetical protein			-2.4	
TM0840	hypothetical protein				2.0
TM0842	response regulator		-2.4		2.3
TM0851	heat shock operon repressor HrcA		-2.2	2.3	5.9
TM0858	hypothetical protein				-2.1
TM0866	putative stomatin/prohibitin-family membrane protease subunit YbbK				-2.3
TM0868	glutaredoxin-related protein			-2.1	
TM0869	thioredoxin reductase			-3.7	
TM0870	cell division protein FtsI			-3.0	-2.8
TRQ2_0054 (TM0873)	ATP-dependent Clp protease, ATP-binding subunit ClpC			2.8	
TM0878	2-oxoglutarate ferredoxin oxidoreductase subunit alpha				-2.2
TM0882	O-acetylhomoserine sulfhydrylase			2.1	2.4
TM0888	thiamin pyrophosphokinase				2.1
TM0895	glycogen synthase				-2.1
TM0897	spoVS-related protein				-2.4
TM0902	RNA polymerase sigma-28 factor, putative			-2.2	
TM0909	flagellar biosynthesis protein FlhB				2.0
TM0934	Arsenical pump membrane protein, ArsB				2.3
TM0951	predicted D-arabinose isomerase			-2.0	
TM0952	predicted D-ribulose kinase			-2.4	
TM0958	ribose ABC transporter, periplasmic ribose-binding protein		2.0		
TM0961	LemA protein				-2.4
TM0962	conserved hypothetical protein			-2.8	
TM0968	ribosomal-protein-S5p-alanine acetyltransferase				2.2
TM0969	Domain of unknown function DUF234				2.0
TM0974	hypothetical protein			2.4	3.0
TM0975	hypothetical protein				2.1
TM0979	ACR, intracellular sulfur oxidation		2.0		
TM0980	hypothetical protein		3.0		

TM0982	conserved hypothetical protein		2.5		
TM0987	conserved hypothetical protein			-2.4	
TM0990	SNF2 family helicase			-3.1	
TM0992	hypothetical protein			-6.8	-4.5
TM1003	transposase-related protein			2.3	3.9
CTN_1574 (downstream of TM1003)	ISTma3, transposase			2.1	2.5
TM1014	conserved hypothetical protein			-2.0	
TM1016	hypothetical protein			2.2	3.0
TM1017	Multidrug resistance efflux transporter EmrE superfamily		-2.1		3.3
TM1021	permease, putative			-2.3	
TM1036	Imidazole glycerol phosphate synthase cyclase subunit			-2.3	
TM1047	septum site-determining protein MinC, putative			-2.5	-2.7
TM1048	M42 glutamyl aminopeptidase				-2.2
TM1050	endoglucanase			-2.6	
TM1051	hypothetical protein			-2.8	
TM1059	spoVS-related protein			-3.5	
TM1061	hypothetical protein				-2.7
TM1074	GH2 family sugar-binding protein			-2.2	
TM1077	pantoate--beta-alanine ligase				2.1
TM1094	RNA methyltransferase, putative			-2.4	
TM1096	adenylosuccinate synthetase			-2.1	
TM1097	ornithine carbamoyltransferase, anabolic			-2.8	-2.5
TM1101	Calcineurin-like phosphoesterase superfamily protein			-2.8	
TM1103	conserved hypothetical protein			-2.7	-2.4
TM1106	hypothetical protein			-2.0	
TM1111	hypothetical protein			4.1	3.4
Tpet_1632 (TM1111)	hypothetical protein			2.0	
TM1119	hypothetical protein			-4.6	-4.1
TM1120	glycerol-3-phosphate ABC transporter, periplasmic substrate binding protein			3.1	7.4
TM1121	glycerol-3-phosphate ABC transporter, permease protein				2.3
TM1122	glycerol-3-phosphate ABC transporter, permease protein				2.8
TM1123	flagellar hook-associated protein 2, putative			2.2	2.9
TM1124	flagellin protein FlaG			2.3	3.1
TM1134	hypothetical protein				-2.8
TM1135	branched chain amino acid ABC transporter, substrate-binding protein			2.3	4.5
TM1138	branched chain amino acid ABC transporter, ATP-binding protein		-2.2		
TM1141	cytochrome C-type biogenesis protein, putative, transmembrane	2.1			
TM1155	glucose-6-phosphate 1-dehydrogenase			2.4	4.3
TM1156	conserved hypothetical protein			-2.3	
TRQ2_1657 (downstream of TM1158)	Ureohydrolase domain				2.4
TM1160	esterase		2.2		
TM1164	2-oxoacid ferredoxin oxidoreductase, alpha subunit			-2.4	-2.3
TM1167	hypothetical protein		2.4	-2.0	-4.7
TM1172	hydroxylamine reductase	-3.1			-2.6

TM1185	methylglyoxal synthase				-2.1
TM1191	galactose-1-phosphate uridylyltransferase			-2.1	
TM1192	alpha-galactosidase			-2.4	
TM1196	oligopeptide ABC transporter, ATP-binding protein			-2.8	
TM1201	arabinogalactan endo-1,4-beta-galactosidase, putative			-4.3	-3.5
TM1204	maltose ABC transporter, periplasmic maltose-binding protein		2.2		
TM1216	NADH dehydrogenase, 49 kDa subunit, putative			-2.0	
TM1218	transcriptional regulator of mannoside/glucoside utilization CelR			-2.2	-2.1
TM1219	mannoside ABC transporter, ATP-binding protein			-2.0	
TM1223	mannoside ABC transporter, periplasmic substrate-binding protein	2.3			
TM1224	transcriptional regulator of mannan utilization ManR				2.7
TM1238	ATP-dependent DNA helicase				-2.0
TM1244	phosphoribosylformylglycinamidine synthase			2.4	5.1
TM1246	phosphoribosylformylglycinamidine synthase II				3.1
TM1248	phosphoribosylglycinamide formyltransferase			2.8	4.8
TM1250	phosphoribosylamine--glycine ligase				2.1
TM1254	beta-phosphoglucomutase, putative				3.3
TM1259	phosphate regulon transcriptional regulatory protein PhoB			-3.4	-2.9
TM1260	phosphate transport system regulator PhoU				-2.1
TM1266	hypothetical protein	-2.6			
TM1268	hypothetical protein	-2.1			
TM1270	cystathionine gamma-synthase			-2.2	
TM1275	Type II antitoxin Phd/YefM			-2.0	
TM1276	sugar ABC transporter, ATP-binding protein				-2.0
TM1281	6-phospho-beta-glucosidase				-2.5
TM1286	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase				3.8
TM1288	radical SAM domain protein	2.3		-2.1	
TM1289	ferredoxin	2.1			
TM1297	oxidoreductase, putative			-2.2	
TM1302	ABC-2 transporter, ATP-binding protein		8.7	14.0	2.7
TM1303	conserved hypothetical protein		7.6	9.5	
TM1307	hypothetical protein		3.0	5.8	2.5
TM1309	ABC-2 transporter			2.2	
TM1310	ABC transporter, ATP-binding protein		2.0	2.7	
TM1315	hypothetical protein		3.6	4.5	
TM1316	hypothetical protein	6.1	27.3	34.7	7.7
TM1317	radical SAM + TIGR04085 protein		7.4	12.6	
TM1318	ABC-2 transporter, ATP-binding protein, authentic frameshift		7.1	5.4	
TM1319	ABC transporter, ATP-binding protein		3.1		
TM1323-1333	hypothetical protein			3.3	3.4
TM1326	ABC-2 transporter, permease protein			-2.5	-2.3
TM1328	ABC transporter, ATP-binding protein		6.0	3.6	
CTN_1102-1078758-1078872 (near TM1333)	hypothetical protein			8.2	
TM1334	radical SAM protein, MiaB/NifB family				3.8
TM1336	permease, putative			2.1	2.4
TM1345	polynucleotide phosphorylase				-2.1
TM1364	flagellar basal-body rod protein FlgB		-2.3	-2.1	

TM1366	flagellar hook-basal body complex protein FliE			-2.2	
TM1373	phospholipase				-2.0
TM1384	adenine phosphoribosyltransferase				-2.2
TM1385	glucose-6-phosphate isomerase			-3.7	-2.1
TM1387	dephospho-CoA kinase			-2.3	-2.4
TM1388	RNA methyltransferase RsmD			-3.0	-2.7
TM1396	alanyl-tRNA synthetase			-3.0	-2.1
TM1399	ribosome recycling factor				2.1
TM1400	serine-pyruvate aminotransferase				2.2
TM1401b	hydroxypyruvate reductase			2.8	2.1
TM1404	antibiotic ABC transporter, transmembrane protein, putative			-2.1	
TM1411	helicase-related protein				3.3
TRQ2_1312 (TM1418)	nucleotidyl transferase				-2.0
TRQ2_1311 (TM1418)	CDP-alcohol phosphatidyltransferase				-3.8
TM1422	rnfB-related protein			-2.7	
TM1429	glycerol uptake facilitator protein	2.3			2.9
TM1430	glycerol kinase	6.9			6.6
TM1431	glycerol uptake operon antiterminator	12.7			14.7
TM1434	hypothetical protein				4.9
TM1435	succinyl-CoA synthetase alpha subunit-related protein				2.5
TRQ2_1330 (TM1438)	GH57 glycogen branching enzyme		2.0		-2.5
TM1438	GH57 glycogen branching enzyme			-4.0	-3.1
TM1439	hypothetical protein				-2.7
TM1442	anti-sigma factor antagonist, putative				2.3
TM1443	cytidylate kinase				2.6
TM1446	GTP-binding protein EngA		-2.1	-2.5	
TM1451	RNA polymerase sigma-A factor		-2.2		3.8
TM1455	Substrate-specific component RibU of riboflavin ECF transporter			2.4	
TM1457	hypothetical protein			2.1	
TM1458	LSU ribosomal protein L21			2.5	
TM1461	inner membrane protein, putative			-4.3	-2.4
TM1467	conserved hypothetical protein			-2.3	
TRQ2_1361 (downstream of TM1467)	hypothetical protein				-2.1
TM1470	transcription termination factor Rho				-2.8
TM1473	SSU ribosomal protein S4				-2.1
CTN_1016 (TM1477)	translation initiation factor IF-1				-2.3
TM1478	methionine aminopeptidase			-3.8	-3.8
TM1480	preprotein translocase SecY subunit		2.7	-2.7	-6.5
TM1481	LSU ribosomal protein L15				-2.8
TM1482	LSU ribosomal protein L30			-2.1	-3.4
TM1483	SSU ribosomal protein S5		2.7		-5.4
TM1484	LSU ribosomal protein L18		2.2		-4.1
TM1485	LSU ribosomal protein L6		3.0		-5.9
TM1486	SSU ribosomal protein S8		2.0		-3.1
TM1488	LSU ribosomal protein L5		2.0		-3.5

TM1489	LSU ribosomal protein L24		2.6		-4.9
TM1491	SSU ribosomal protein S17		2.4		-4.5
TM1492	LSU ribosomal protein L29		2.7		-5.2
TM1493	LSU ribosomal protein L16				-6.4
TM1494	SSU ribosomal protein S3			-2.8	-4.0
TM1495	LSU ribosomal protein L22		2.2		-3.8
TM1496	SSU ribosomal protein S19		2.3		-3.6
TM1497	LSU ribosomal protein L2			-2.2	-4.0
TM1498	LSU ribosomal protein L23		2.5	-2.2	-6.1
TM1499	LSU ribosomal protein L4				-7.5
TM1500	LSU ribosomal protein L3		2.1		-3.8
TM1501	SSU ribosomal protein S10		2.2		-3.6
TM1502	translation elongation factor Tu		4.0		-5.6
CTN_0990 (TM1503)	translation elongation factor G				-3.9
TM1503	translation elongation factor G		2.1		-4.1
TM1504	ribosomal protein S7				-2.1
TM1507	phoH-related protein				2.3
TM1508	conserved hypothetical protein		-2.6		2.0
TM1510	hypothetical protein				2.9
TM1513	conserved hypothetical protein			-2.2	
TM1519	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase-related			2.4	3.0
TM1520	dihydrodipicolinate reductase			2.1	2.3
TM1525	endoglucanase				3.9
TM1533	ferredoxin				2.4
TM1539	flagellar P-ring protein				2.2
TM1540	flagellar L-ring protein		-2.9		3.3
TM1541	flagellar protein FlgA, putative		-3.1		3.1
TM1542	flagellar basal-body rod protein FlgG		-2.1		2.3
TM1543	flagellar basal-body rod protein FlgF		-2.4		3.1
TM1544	rod shape-determining protein MreB		-2.3		3.5
TM1546	single stranded DNA-specific exonuclease, putative		-2.2		
TM1568	16S rRNA processing protein RimM, putative				2.1
TM1569	tRNA guanine-N1 methyltransferase				2.4
TM1574	pseudouridylate synthase I				2.1
TM1578	preprotein translocase SecA subunit		-2.8		
TM1588	conserved hypothetical protein, GGDEF domain				16.8
TM1594	diguanylate cyclase with PAS/PAC sensor			-2.3	
TM1598	RNA polymerase sigma-E factor		-2.6	2.3	7.9
TM1600	hypothetical protein				2.9
TM1611	ATP synthase F1, subunit gamma				2.8
TM1612	ATP synthase F1, subunit alpha				2.2
TM1615	ATP synthase F0, subunit c			2.7	3.6
TM1617	hypothetical protein			2.4	3.6
TM1623	LepA protein				-2.0
TM1624	beta-mannosidase, putative				-3.3
TM1640	glutamate synthase, beta subunit			-2.1	
TM1650	alpha-amylase, putative			-2.8	-2.2
TM1652	predicted periplasmic protein DUF2238			-2.6	-2.2
TM1660	conserved hypothetical protein			-2.3	
TM1662	stationary phase survival protein, nucleotidase SurE			-3.6	-3.2

TM1664	ribosomal RNA small subunit methyltransferase C			-3.6	-2.6
TM1665	signal transduction histidine kinase			-3.1	-2.3
TM1670	hypothetical protein				2.3
TM1677	transposase, putative			-2.1	
TM1680	hypothetical protein	2.1			
TM1684	ribosomal protein L31				-2.6
TM1685	S1 domain containing RNA-binding protein				-3.4
TM1695	conserved hypothetical protein				-2.1
TM1696	type IV prepilin peptidase			-2.4	-3.4
TM1697	RNA M5U methyltransferase family			-2.1	
TM1698	aspartate aminotransferase			-2.1	
TM1699	conserved hypothetical protein			-3.4	-2.4
TM1707	ribonucleotide reductase transcriptional regulator NrdR		-2.2		2.4
TM1708	conserved hypothetical protein		-2.1		3.2
TM1717	ribosome small subunit-dependent GTPase EngC			-2.1	
TM1725	vacuolar ATP synthase sub D-related protein, authentic point mutation			3.8	6.3
TM1731	conserved hypothetical protein				2.0
TM1734	phosphate transport system regulator PhoU, putative			-2.2	
TM1739	ribonuclease H-like domain protein			-2.5	
TM1745	hypothetical protein			2.2	3.9
TM1755	phosphate butyryltransferase		-2.8	-4.0	
TRQ2_1065 (downstream of TM1759)	thiamine pyrophosphate binding domain protein				2.0
TM1761	excinuclease ABC, subunit B			2.2	5.7
TM1763	translation elongation factor P		2.1		-3.4
TM1768	exodeoxyribonuclease VII, large subunit			-2.0	-2.2
TM1778	conserved hypothetical protein			2.1	
TM1782	N-acetyl-gamma-glutamyl-phosphate reductase				2.4
TRQ2_1029 (downstream of TM1792)	CRISPR-associated protein, Cmr5			-2.0	
TM1792	CRISPR-associated RAMP protein, Cmr4	2.4		-2.1	
TM1793	CRISPR-associated protein, Cmr3	2.4		-2.8	
TM1794	CRISPR-associated protein, Cmr2			-2.3	
TM1795	CRISPR-associated RAMP protein, Cmr1			-3.8	
TM1798	CRISPR-associated protein, Cas4a	2.6		-2.1	
TM1799	CRISPR-associated helicase, Cas3	2.9	-2.5	-3.1	2.3
TM1800	CRISPR-associated protein, Cas5	2.0			2.4
TM1801	CRISPR-associated protein, Csh2	2.1	-2.3		2.7
TM1802	CRISPR-associated protein, Csh1			-2.3	
TM1809	CRISPR-associated RAMP protein, Csm3		2.1		
TM1814	CRISPR repeat RNA endoribonuclease Cas6		2.4		
TM1820	GMP synthase		-2.8		2.4
TM1822	ftsH protease activity modulator HflK		-2.0		2.1
TM1823	ftsH protease activity modulator HflC			-2.6	-2.0
TM1826	GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase				2.6
TM1827	riboflavin synthase, alpha subunit			2.2	2.5
TM1831	transposase, putative			-2.0	

TM1832	transposase			-2.6	-2.6
TM1833	methyl-accepting chemotaxis-related protein			-2.5	-2.2
TM1834	alpha-glucosidase				-2.6
TM1847	transcriptional regulator of glucose/trehalose utilization GluR		2.2	-2.2	-5.0
TRQ2_0975-0976	transcriptional regulator of glucose/trehalose utilization GluR		2.1		-3.6
TRQ2_0975a	glucose/xylose ABC transporter, ATP-binding component XylK2		2.1		-3.1
TRQ2_0975b	glucose/xylose ABC transporter, ATP-binding component XylK2		3.6		-7.9
TRQ2_0975c	glucose/xylose ABC transporter, ATP-binding component XylK2				-2.6
CTN_0775	glucose/xylose ABC transporter, ATP-binding component XylK2		2.7		-5.6
TRQ2_0974-0975	glucose/xylose ABC transporter, permease component XylF2		3.0	-2.0	-7.6
CTN_0776	glucose/xylose ABC transporter, permease component XylF2		2.2		-4.8
TRQ2_0973b	glucose/xylose ABC transporter, substrate-binding component XylE2		14.4		-10.6
TRQ2_0973c	glucose/xylose ABC transporter, substrate-binding component XylE2	2.0	13.5		-8.5
CTN_0777	glucose/xylose ABC transporter, substrate-binding component XylE2		6.8		-4.8
TRQ2_0972-0973	trehalose ABC transporter, permease component TreG		4.2		-3.5
CTN_0778a	trehalose ABC transporter, permease component TreG				-2.6
TRQ2_0971	trehalose ABC transporter, permease component TreF		6.0		-5.1
CTN_0779a	trehalose ABC transporter, permease component TreF		2.6		-3.0
CTN_0779b	trehalose ABC transporter, permease component TreF		3.8		-3.3
TRQ2_0970a	trehalose ABC transporter, substrate-binding component TreE		10.1		-6.7
TRQ2_0970b	trehalose ABC transporter, substrate-binding component TreE		8.0		-7.8
CTN_0780	trehalose ABC transporter, substrate-binding component TreE		9.0		-5.9
TRQ2_0969b	beta-glucosidase A			-2.3	-4.0
CTN_0782	beta-glucosidase A				-2.5
TM1848	cellobiose-phosphorylase		-7.1	6.9	63.3
TM1849	hypothetical protein		-4.2	6.7	34.4
TM1850	Hypothetical protein		-2.0	17.5	41.6
TM1851	alpha-mannosidase, putative				3.3
TM1852	glycoside hydrolase family 43 protein				2.6
TM1858	RecX regulatory protein, putative			-2.5	-2.4
TM1865	endonuclease V			-2.6	-2.7
TM1870	septum site-determining protein MinD				-2.2
TM1874	cold shock protein CspA			-4.1	-8.0
TM1876	metallophosphoesterase				-2.0
TM1878	UDP-sugar hydrolase		-2.2		2.9

Chapter 5

Population dynamics of synthetic mixed cultures of hyperthermophilic *Thermotoga* species

Andrew D. Frock, William J. Kornahrens, and Robert M. Kelly

Department of Chemical and Biomolecular Engineering
North Carolina State University
Raleigh, NC 27695-7905

In preparation to be submitted to: *Applied and Environmental Microbiology*

ABSTRACT

The metabolic flexibility of mixed cultures would be desirable for any bioconversion in which the substrate input is heterogeneous and environmental conditions vary. The population dynamics of various two and four species co-cultures of hyperthermophilic *Thermotoga* species were analyzed by qPCR, and the results indicated co-cultures were not stable under the growth conditions employed in this study. Considerations and challenges of monitoring population composition of closely related species are discussed.

INTRODUCTION

Employment of mixed cultures for the conversion of lignocellulosic and waste-derived feedstocks has been proposed as a means to increase the adaptability of the culture and bypass the need to engineer a single strain capable of all desired catalytic activities (1, 2, 6). Improved design of such industrial bioprocesses requires a more comprehensive understanding of the way both pure and co-cultures respond to changing environmental conditions and stresses. Previously, during growth on relatively lean medium, the hyperthermophilic bacterium *T. maritima* was shown to adopt unusual morphology and up-regulate the expression of small proteins likely involved in signaling, sensing, or competing upon entry into stationary phase (see Chapter 4). Because such molecules can be involved in interactions with closely related species or strains (14), co-cultures of *Thermotoga* species were investigated in search of physiological responses of one species to another.

In some cases, syntrophic consortia can simply be created by combining strains that have capabilities the other strains lack (1), but stable mixed cultures can be difficult to

establish because the metabolic rates are not matched, changing conditions will generally favor one species over another, and even spatial separation of the constituents can be critical (3, 11, 13, 15). Nevertheless, stable co-cultures of closely-related species have been reported (12, 17), although mechanistic explanations were lacking. Zeidan et al. (17) reported stable coexistence of *Caldicellulosiruptor saccharolyticus* and *C. kristjanssonii* in continuous culture. They accurately modeled the mixed culture by assuming that an unknown growth-associated product from *C. saccharolyticus* enhanced the growth of *C. kristjansonnii*.

Here, various co-cultures of *T. maritima*, *T. neapolitana, T. petrophila*, and *T.* sp. str. RQ2 were examined using quantitative real-time PCR (qPCR). While it is simple to determine which species dominates the co-culture, it is much more difficult to pinpoint the exact reason that species dominates. This could be the combination of multiple factors including length of lag phase, growth rate, tolerance to waste products, ability to use alternate carbon sources, and ability to survive stationary phase (16).

MATERIALS AND METHODS

Growth of *Thermotoga* species. *Thermotoga* species were grown anaerobically on either sea salts medium (SSM) (8) or artificial sea water (ASW) medium (4) in an oil bath at 80°C with shaking at 100 rpm. Either glucose (final concentration 2.5 g/L) or a mixture of polysaccharides (7) was provided as carbon source. Culture bottles were made anaerobic by sealing, sparging with nitrogen, and adding $Na_2S•9H_2O$ at a final concentration of 0.05% (w/v). For preparation of freezer stocks, 970 µl of cooled log-phase culture was added to 30 µl of dimethylsulfoxide and stored at -80°C. Growth was monitored by measurement of

optical density at 600 nm or enumeration of acridine orange-stained cells viewed with epifluorescence microscopy.

Establishment of mixed cultures. Prior to mixed culture experiments, each species was passed at least three times on the medium to be used. Then a “test growth curve” was performed for each species. The purpose of the “test growth curve” was to determine the amount of time required for each species to reach mid- to late-exponential phase (4-7 x 10^7 cells/mL). This information was used to inoculate a new set of pure cultures at staggered times so that each species reached 4-7 x 10^7 cells/mL at approximately the same time. In this way, the first pass of the mixed culture started with hot, actively-growing, exponential phase inoculate of each species. In principle, this strategy reduces lag phase and puts each species on ‘more equal footing’ at the beginning of the mixed culture. Cultures were harvested for genomic DNA extraction at 7,500 × g for 10 minutes. Genomic DNA was isolated either by a modified phenol-chloroform extraction method (9) or using the ChargeSwitch gDNA Mini Bacteria kit (Invitrogen) according to the manufacturer’s instructions.

Quantitative real-time PCR. Unique genes for each species were identified using Mauve multiple genome alignment software (5) and confirmed by BLAST. The genes CTN_1550, CTN_0355, TRQ2_0300, TRQ2_0637, Tpet_1773, and TM0625 were chosen as targets. Primers were designed manually with 40-55% GC content and 55-60°C melting temperature so that the length of the PCR products were between 75 and 150 bp. The primers for CTN_0355 and TRQ2_0637 were used only once to confirm that different primers give similar results. PCR was performed with each genomic DNA/primer set

combination to confirm primer specificity. Real-time PCR was performed with either SYBR Green Supermix or SsoFast EvaGreen Supermix (Bio-Rad) according to the manufacturer's instructions. SYBR Green was used with a Bio-Rad iCycler iQ system, while SsoFast EvaGreen was used with a Bio-Rad CFX96. Generally, 5-10 ng of mixed culture genomic DNA was used in a 20 µL reaction with primers at a final concentration of 500 nM. Series dilutions of genomic DNA from pure cultures were used as standards for absolute quantification. Unknown samples were assayed in triplicate.

RESULTS

Model for mixed culture dynamics. The study of mixed cultures first requires consideration of different ways that species might affect each other. For simplicity's sake, initially consider binary mixed cultures. If two species have complementary metabolism, they might grow faster or reach higher cell densities in mixed culture than in pure culture. This has been observed for a co-culture of *Thermotoga maritima* and *Methanocaldococcus jannaschii,* where an inhibitory waste product of *T. maritima* (hydrogen) is an energy source for *M. jannaschii* (10). Alternatively, if two species are not inhibited by the same waste products, but form a "division of resources" type of co-culture where nutrients are consumed in a noncompetitive manner, then one might expect they could grow at the same rate to the same final density in mixed culture as they do in pure culture. The mixed cultures being examined here would not be expected to fall into either of these first two categories. Because these four *Thermotoga* species have the same type of metabolism and are closely related, we would expect that they would utilize the same nutrients and produce the same waste products. Therefore, we would expect that these species in mixed culture might

become limited by the same nutrients, inhibited by the same waste products, and even respond to the same pheromones, such that the final cell density of a pure culture might be similar to the final cell density of all species combined in a mixed culture. However, if these species do not recognize each other as “self” and compete with one another, then they might also produce antibiotics against one another.

To determine whether or not any of the species affect each other’s growth, the mixed culture composition determined by quantitative PCR (qPCR) must be compared to what we would expect the mixed culture composition to be, based on the growth rates of the four species in pure culture. The exponential growth of each species individually can be modeled by the equation:

$$N_A(t) = N_{A0}e^{\mu t} \qquad [1]$$

where t is the time in hours, N_{A0} is the cell density or amount of DNA from species “A” present at t=0 and μ is the specific growth rate. For mixed cultures, if we assume that (1) the length of lag phase for each species is the same and (2) each species stops growing at the same *total* cell density, then the time of exponential growth for each species in a mixed culture is equal, so that we can relate the growth of two species (“A” and “B”) in a mixed culture by the following equation:

$$\frac{ln\frac{N_{AF}}{N_{A0}}}{\mu_A} = \frac{ln\frac{N_{BF}}{N_{B0}}}{\mu_B} \qquad [2]$$

where N_{AF} is the amount of cells or DNA from species “A” at the end of growth. Because the inoculum for each culture in these experiments was 1% (v/v), the total amount of DNA should effectively increase by 10,000-fold from the end of the first pass to the end of the third pass. With the real-time PCR data, we can calculate the percentage of the total mixed culture DNA that comes from each species at different time points. For example, we can

predict the percent of mixed culture DNA from each species at the end of pass 3 (X_{M3}, X_{N3}, X_{P3}, X_{R3}) from the experimentally determined pure culture growth rates (μ_M, μ_N, μ_P, and μ_R) and the composition of pass 1 determined by qPCR (X_{M1}, X_{N1}, X_{P1}, X_{R1}) by solving the following system of four equations with four unknowns:

$$\frac{ln\frac{10000X_{M3}}{X_{M1}}}{\mu_M} = \frac{ln\frac{10000X_{N3}}{X_{N1}}}{\mu_N} \quad [3]$$

$$\frac{ln\frac{10000X_{M3}}{X_{M1}}}{\mu_M} = \frac{ln\frac{10000X_{P3}}{X_{P1}}}{\mu_P} \quad [4]$$

$$\frac{ln\frac{10000X_{M3}}{X_{M1}}}{\mu_M} = \frac{ln\frac{10000X_{R3}}{X_{R1}}}{\mu_R} \quad [5]$$

$$X_{M3} + X_{N3} + X_{P3} + X_{R3} = 100 \quad [6]$$

Then we can compare the predicted composition of pass 3 to that which was determined by qPCR to look for evidence of interactions between the species. We must also assume that the number of cells in the mixed culture is proportional to the amount of DNA, because the pure culture growth rates are measured by cell counts, while the composition of the mixed culture is measured by the amount of DNA.

Four species mixed culture on SSM + 2.5 g/L glucose. Results for a four species mixture grown on SSM + 2.5 g/L glucose are shown in Table 1. 99.6% of the DNA present at the end of the third pass was estimated by qPCR to be *T.* sp. RQ2 (RQ2). Because this species had the highest growth rate (1.19 hr^{-1}) under these conditions, the third pass was predicted to be 97.6% RQ2, in agreement with the qPCR results. The predicted amount of *T. maritima* (Mar) in the third pass (0.9%) was also reasonably close what was observed

(0.4%). However, both *T. neapolitana* (Nea) and *T. petrophila* (Pet) were present in lower concentrations than expected in the third pass.

Four species mixed culture on ASW + polysaccharides. To test mixed culture dynamics under different conditions, these four species were co-cultured on richer media (ASW) with a mix of polysaccharides. In this instance, the pure culture growth rates were more similar, but RQ2 once again dominated, making up 88.4% of the DNA present after the third pass (Table 2). The mixed culture composition reflected the more similar growth rates in that RQ2 did not take over the culture as quickly as it did on SSM + glucose. However, on ASW + polysaccharides, Nea's growth rate is actually slightly higher than *T.* sp. RQ2's, so the population dynamics cannot be fully explained by the pure culture growth rates alone. Mar and Nea are both present at lower concentrations than expected in the third pass.

***T. neapolitana* in binary co-culture with other species.** In both instances described above, Nea did not persist in the four species mixed culture as expected. One possible explanation for this would be that one (or more) of the other three species inhibited Nea's growth. To test this, Nea was grown in co-culture with each of the three other species individually on SSM + 2.5 g/L glucose. The composition of the co-cultures was determined by qPCR at t = 0, at t = 6 hours, and at t = 12 hours. These time points were chosen to investigate whether Nea "loses ground" to the other species during exponential phase or stationary phase. Pure cultures were inoculated simultaneously with the mixed cultures, and the growth was monitored. Because of the number of cultures involved, growth was monitored by OD_{600} instead of cell counts (Figure 1).

In this experiment, Nea clearly outgrows Mar and Pet as expected, and is out-grown by RQ2 as expected. The Nea-Mar and Nea-Pet co-culture growth curves are nearly identical to the Nea pure culture growth curve. This raises the question of whether Mar and Pet grew at all. One could envision that if Nea reached final cell density before Mar or Pet exited lag phase, then the environment present at that time might not allow those species to begin growing. However, if these two species did not grow at all in the co-cultures, they should comprise less than 1% of the culture at the final time point. Therefore, the qPCR data suggests that Pet and Mar did in fact grow, but were out-grown by Nea. Unfortunately, the experimental design used for these co-cultures (multiple time points in a single pass of growth) does not allow for application of the model used for the multi-pass experiments described above. Nevertheless, the binary co-cultures do not provide evidence that Nea's growth is affected in the presence of any of the other three species.

Binary co-culture compared to pure culture by qPCR. The possibility remains that RQ2 somehow inhibits Nea. This was further tested with a simpler experiment. RQ2 and Nea pure cultures, along with a Nea-RQ2 co-culture were simultaneously inoculated. In late exponential phase, the co-culture was harvested for DNA. The pure cultures were combined and harvested for DNA. Both DNA samples were analyzed by qPCR. This way, the Nea:RQ2 ratio in the co-culture can be compared to the Nea:RQ2 ratio in the combined pure cultures. This also allows determination of the number of cells present in each pure culture, that go into the combined pure culture sample, whereas the number of cells in a co-culture cannot be determined because these species morphology is too similar.

The results suggest that from inoculation to late exponential phase, Nea and RQ2 do not grow any differently when they are together than they do when they are apart. According

to qPCR, 66.8% of the DNA in the co-culture was from RQ2, and 64.5% of the DNA in the combined pure cultures was from RQ2. Interestingly, the cell densities of the pure cultures at the time of harvest were 9.7 x 10^7 cells/mL (RQ2) and 8.8 x 10^7 cells/mL (Nea). Although the RQ2:Nea cell density ratio was 1.1:1, the RQ2:Nea DNA ratio was 1.8:1. This suggests that RQ2 contains about 65% more DNA per cell than Nea. This is consistent with the observation that RQ2 cells seem to appear slightly larger under the microscope.

DISCUSSION

Quantitative real-time PCR is a powerful technique that can be used to track the population dynamics of a mixed culture, even one consisting of very closely related species. However, interpretation of the results in the context of cooperative or competitive inter-species interactions can be more difficult. Here, we attempted to compare the mixed culture dynamics to expectations based on pure culture growth rates. We essentially assumed that all growth characteristics other than growth rate were identical for all four species. However, qPCR analysis of a sample of two combined pure cultures with known cell density suggests that each species does not have the same amount of DNA per cell. This complicates the assumptions made here, because the DNA content per cell in a mixed culture could change as the species composition of the mixed culture shifts. We also assumed that the length of lag phase and final cell density is the same for each species, although we have seen final cell densities vary from 1-2 x 10^8 cells/mL for Pet to 4-6 x 10^8 cells/ml for Nea. This is especially difficult to account for in any model since it is not known exactly what causes the cells to stop growing. We could assume that Nea would stop growing in a mixed culture when the total cell density reached 4-6 x 10^8 cells/ml, but that is presumably only true if the

concentrations of nutrients, waste products, and signaling molecules are the same in the mixed culture as they are in the pure culture. Our simplistic method of combining pure cultures for qPCR analysis suggested that Nea and RQ2 do not affect each other in co-culture during exponential phase, but this method would not be accurate for investigating activity during stationary phase, during which antagonistic interactions are more likely to occur.

SUMMARY

This study highlights the difficulties in quantitatively comparing pure and mixed culture growth in the absence of obvious interactions between the constituents. Nevertheless, this study supports previous assertions that stable artificial mixed cultures can only be established when the existence of each species is required for the survival of the others. This is a cautionary tale for bioprocess researchers who consider combining multiple strains with different catabolic or fermentative abilities in search of “synergistic” production of the targeted biomolecule. When a synthetic mixed culture is created, the null hypothesis should always be that all species except one will disappear after a relatively short time, even if that one species seems to have a growth rate that is only marginally faster.

Further efforts could improve upon the mixed culture model used here to account for the length of lag phase of each individual species. Alternatively, modeling continuous culture population dynamics would circumvent the need to account for phase-dependent growth. However, while medium composition is the primary consideration for batch growth, both medium composition and dilution rate are important factors for continuous culture, because inter-species signaling or defense mechanisms may only be active under certain

physicochemical conditions. The method used to monitor growth of the mixed culture could also be important for interpreting qPCR data. Some hyperthermophiles like *Thermotoga* species grow to relatively low final cell densities, such that cell counts can be more effective than OD_{600} measurements of growth. Both cell counts and OD_{600} may not always be proportional to the amount of biomass or DNA due to varitiations in cell size and morphology. Population dynamic data derived from qPCR would also be easier to interpret for mixed cultures with previously demonstrated synergistic or competitive interactions. For example, plaque assays are commonly used for identification of bactericidal activity, but qPCR analysis of a bacteriocin producer/bacteriocin target co-culture could provide further insight into factors affecting such biological “warfare”.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy GTL Program (DG-FG02-08ER64687) and the National Science Foundation (CBET-0730091). ADF acknowledges support from NIH Pre-doctoral Biotechnology Traineeships (NIH T32 GM008776-06). WJK is acknowledged for development of the qPCR assay and execution of some of the experiments.

REFERENCES

1. **Bernstein, H. C., S. D. Paulson, and R. P. Carlson.** 2012. Synthetic Escherichia coli consortia engineered for syntrophy demonstrate enhanced biomass productivity. J Biotechnol **157:**159-66.

2. **Brenner, K., L. You, and F. H. Arnold.** 2008. Engineering microbial consortia: a new frontier in synthetic biology. Trends Biotechnol **26:**483-9.

3. **Briones, A., and L. Raskin.** 2003. Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. Curr Opin Biotechnol **14:**270-6.

4. **Chhabra, S. R., and R. M. Kelly.** 2002. Biochemical characterization of *Thermotoga maritima* endoglucanase Cel74 with and without a carbohydrate binding module (CBM). FEBS Lett. **531:**375-380.

5. **Darling, A. E., B. Mau, and N. T. Perna.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147.

6. **Eiteman, M. A., S. A. Lee, and E. Altman.** 2008. A co-fermentation strategy to consume sugar mixtures effectively. J Biol Eng **2:**3.

7. **Frock, A. D., S. R. Gray, and R. M. Kelly.** 2012. Hyperthermophilic Thermotoga species differ with respect to specific carbohydrate transporters and glycoside hydrolases. Appl Environ Microbiol **78:**1978-86.

8. **Gao, J., M. W. Bauer, K. R. Shockley, M. A. Pysz, and R. M. Kelly.** 2003. Growth of hyperthermophilic archaeon *Pyrococcus furiosus* on chitin involves two family 18 chitinases. Appl. Environ. Microbiol. **69:**3119-3128.

9. **Geslin, C., M. Le Romancer, G. Erauso, M. Gaillard, G. Perrot, and D. Prieur.** 2003. PAV1, the first virus-like particle isolated from a hyperthermophilic euryarchaeote, "Pyrococcus abyssi". J Bacteriol **185:**3888-94.

10. **Johnson, M. R., S. B. Conners, C. I. Montero, C. J. Chou, K. R. Shockley, and R. M. Kelly.** 2006. The *Thermotoga maritima* phenotype is impacted by syntrophic interaction with *Methanococcus jannaschii* in hyperthermophilic coculture. Appl. Environ. Microbiol. **72:**811-818.

11. **Kim, H. J., J. Q. Boedicker, J. W. Choi, and R. F. Ismagilov.** 2008. Defined spatial structure stabilizes a synthetic multispecies bacterial community. Proc Natl Acad Sci U S A **105:**18188-93.

12. **Masset, J., M. Calusinska, C. Hamilton, S. Hiligsmann, B. Joris, A. Wilmotte, and P. Thonart.** 2012. Fermentative hydrogen production from glucose and starch using pure strains and artificial co-cultures ofClostridium spp. Biotechnol Biofuels **5:**35.

13. **Muralidharan, V., K. D. Rinker, I. S. Hirsh, E. J. Bouwer, and R. M. Kelly.** 1997. Hydrogen transfer between methanogens and fermentative heterotrophs in hyperthermophilic cocultures. Biotechnol Bioeng **56:**268-278.

14. **Waite, R. D., and M. A. Curtis.** 2009. Pseudomonas aeruginosa PAO1 pyocin production affects population dynamics within mixed-culture biofilms. J Bacteriol **191:**1349-54.

15. **Weibel, D. B.** 2008. Building communities one bacterium at a time. Proc Natl Acad Sci U S A **105:**18075-6.

16. **Zambrano, M. M., D. A. Siegele, M. Almiron, A. Tormo, and R. Kolter.** 1993. Microbial competition: Escherichia coli mutants that take over stationary phase cultures. Science **259:**1757-60.

17. **Zeidan, A. A., P. Radstrom, and E. W. van Niel.** 2010. Stable coexistence of two Caldicellulosiruptor species in a de novo constructed hydrogen-producing co-culture. Microb Cell Fact **9:**102.

Table 5.1. Four species mixed culture on SSM + 2.5 g/L glucose				
Species	**Pure culture growth rate, μ (hr^{-1})**	**% DNA (qPCR)**		**% DNA (expected)**
		Pass 1	**Pass 3**	**Pass 3**
T. maritima	0.66	33.76	0.4	0.9
T. neapolitana	0.92	6.48	0.00067	1.4
T. petrophila	0.55	14.08	0.016	0.1
T. sp. RQ2	1.19	45.68	99.6	97.6

Table 5.2. Four species mixed culture on ASW + polysaccharide mix				
Species	**Pure culture growth rate, μ (hr^{-1})**	**% DNA (qPCR)**		**% DNA (expected)**
		Pass 1	**Pass 3**	**Pass 3**
T. maritima	0.85	14.08	1.8	10.0
T. neapolitana	0.94	14.91	4.0	27.0
T. petrophila	0.70	14.40	5.9	2.1
T. sp. RQ2	0.89	56.61	88.4	60.9

Table 5.3. Dynamics of two and four species co-cultures involving *T. neapolitana*			
	% DNA (qPCR)		
Time (hours)	0	6	12
***T. neapolitana* + *T. maritima* co-culture**			
T. neapolitana	67.1	80.6	89.4
T. maritima	32.9	19.4	10.6
***T. neapolitana* + *T. petrophila* co-culture**			
T. neapolitana	82.6	94.0	97.2
T. petrophila	17.4	6.0	2.8
***T. neapolitana* + *T.* sp. RQ2 co-culture**			
T. neapolitana	52.6	41.0	33.0
T. sp. RQ2	47.4	59.0	67.0
Four species co-culture			
T. neapolitana	34.8	34.1	28.2
T. maritima	19.1	8.9	5.6
T. petrophila	7.6	1.6	1.0
T. sp. RQ2	38.6	55.4	65.2

Table 5.4. Comparison of pure and co-culture by qPCR			
	% DNA (qPCR)		
	Co-culture	**Combined pure cultures**	**Pure culture cells/mL**
T. neapolitana	33.2	35.5	8.8×10^7
***T.* sp. RQ2**	66.8	64.5	9.7×10^7

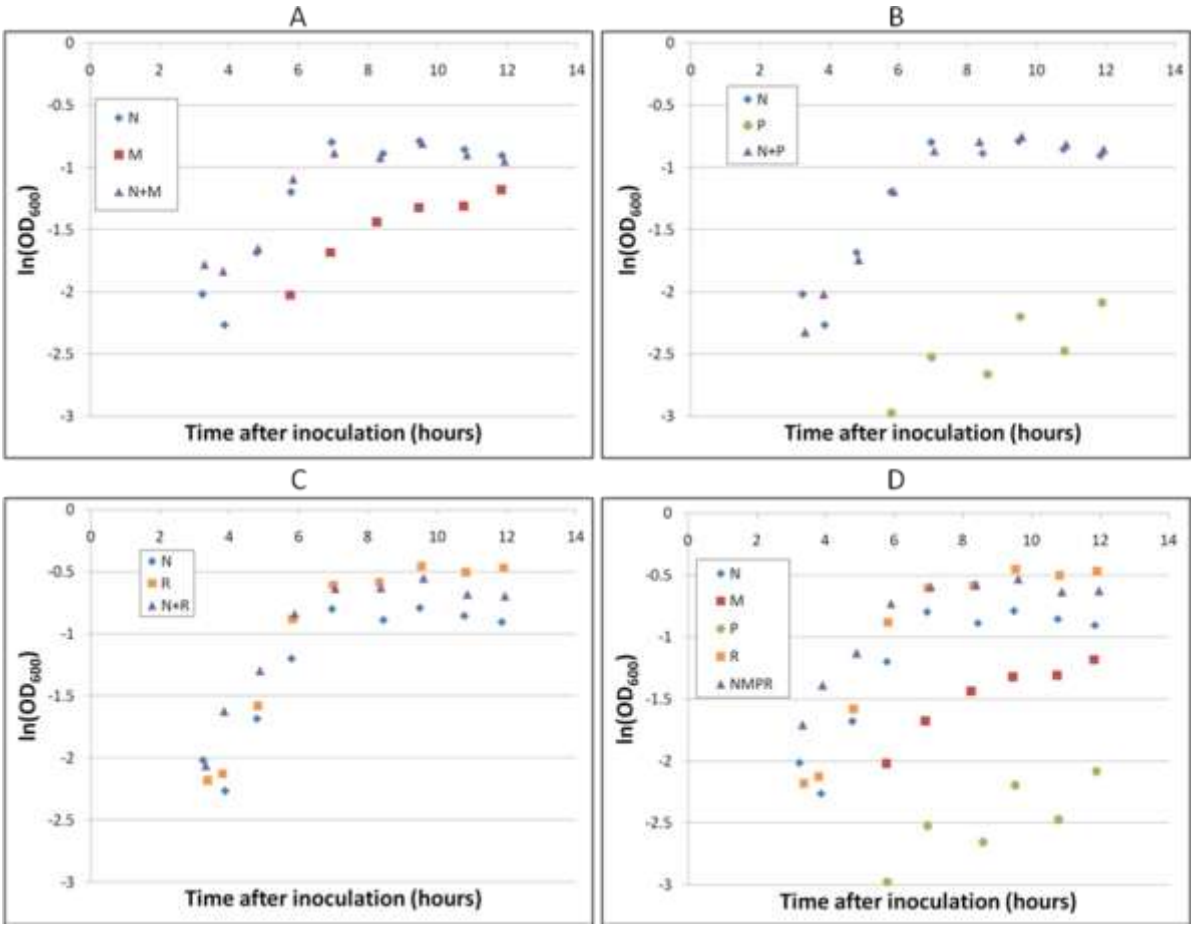

Figure 5.1. Pure culture, two species co-culture, and four species co-culture growth curves. (A) *T. neapolitana* + *T. maritima* co-culture, (B) *T. neapolitana* + *T. petrophila* co-culture, (C) *T. neapolitana* + *T.* sp. RQ2 co-culture, (D) All four species co-culture.

Chapter 5 Appendix

Primers used for qPCR

Table 5.A1. Primers used for qPCR						
Primer name	**Genome coordinates**		**Length of PCR product**	**Primer sequence**	**%GC**	**T_m (°C)**
CTN_1550 fwd	1512815	1512839	149	AACACCAGAGCTGTTTATGAGATGG	44.0	57.0
CTN_1550 rev	1512963	1512934		CGTGCATATTAGCCATCTTAGAAAAGTGTC	40.0	57.6
Tpet_1773 fwd	1779111	1779087	94	GTGGAGTTCTTGGAATTTCTGGTTG	44.0	56.3
Tpet_1773 rev	1779042	1779018		CGTCGTGAAATACCAAAACATCACT	40.0	55.8
TRQ2_0300 fwd	284071	284101	77	TTAATAGTAGACGGTACCCCCTATAGTGGTT	41.9	59.2
TRQ2_0300 rev	284147	284122		TGACCCACTTTCTTCGTCAGATACTT	42.3	57.6
TM0625 fwd	660057	660032	88	GTTGTCGTTTTGCTTCTCACAGGTAG	46.2	58.0
TM0625 rev	659970	659995		GATCACCAAGAATTTCATACCCTCCG	46.2	57.4
CTN_0355 fwd	344657	344635	76	AAGAGGGGAAACTTGTTGTGAGG	47.8	57.2
CTN_0355 rev	344601	344582		AACCGCAAACAATCCCTCCC	55.0	58.3
TRQ2_0637 fwd	646878	646898	81	TGCTCTTCTTCAGATGGTGGC	52.4	57.3
TRQ2_0637 rev	646958	646935		GAATAACGCAGAGGGAATAGGACT	45.8	56.2