Insights into temperature adaptation in the Thermotogae gained through transcriptomics and comparative genomics

by

Stephen Michael Jerome Pollo

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Microbiology and Biotechnology

Department of Biological Sciences
University of Alberta

ABSTRACT

Thermophilic microbes are extremophiles that live at high temperatures. In order to survive and maintain function of their biological molecules, they have a suite of characteristics not found in organisms that grow at moderate temperatures (mesophiles) that range from the cellular to the protein level. These fundamental differences presumably present a barrier to transitioning between the two lifestyles, yet many lineages are thought to have transitioned between thermophily and mesophily at least once. Studying groups of closely related thermophilic and mesophilic organisms can provide insight into these transitions.

The bacterial phylum Thermotogae comprises hyperthermophiles (growing up to 90°C), thermophiles (50-70°C) and mesophiles (<45°C), thus presenting an excellent opportunity to study bacterial temperature adaptation. One Thermotogae species, *Kosmotoga olearia,* grows optimally at 65°C but grows over an extraordinarily broad temperature range of ~25 - 79°C. To investigate how this bacterium can tolerate such an enormous temperature range, RNA-seq experiments were performed on cultures grown across its permissive temperature range. Multivariate analyses of the resulting transcriptomes showed that the temperature treatments separated into three groups: heat-stressed (77°C), intermediate (65°C and 40°C), and cold-stressed (30°C and 25°C). Among the genes differentially expressed, unsurprisingly, were genes with known temperature responses like chaperones, proteases, cold-shock proteins, and helicases. Intriguingly however, increased expression of genes involved in carbohydrate metabolism and transport at supra-optimal temperature, contrasted with increased expression of genes involved in amino acid metabolism and transport at sub-optimal temperature suggests

global metabolism is changed by growth temperature. This may allow *K. olearia* to play distinct roles across a range of thermal environments.

Among the differentially expressed genes in *K. olearia* are genes shared with mesophilic *Mesotoga* spp. but none of the other thermophilic Thermotogae. Many of these genes have inferred regulatory functions implying that large regulatory changes accompany low temperature growth. In agreement with this, more genes were found to be differentially expressed at low temperatures compared to optimal than at high temperatures compared to optimal in *K. olearia*. Further genomic comparisons between *K. olearia* and the related *K. arenicorallina*, which has a narrower growth temperature range of 35 - 70°C, identified 243 genes that could be important for the wide temperature range of *K. olearia*.

Clarifying mechanisms by which Bacteria adapt to temperature changes in isolation can inform studies of complex microbial communities in environments that experience fluctuations in temperature as well as provide a starting place to predict the responses of microbial communities to long term temperature change.

ACKNOWLEDGEMENTS

I am very grateful to all of the people who have helped in the realization of this research. In particular I would like to thank my supervisors Dr. Camilla Nesbø and Dr. Julia Foght for mentoring me throughout my program. I would also like to thank Dr. Yan Boucher for being part of the examining committee. I would like to thank The Applied Genomics Core at the University of Alberta, as well as Rhianna Charchuk for help with genome sequencing and assembly and Dr. Karen Budwill at Alberta Innovates – Technology Futures for culture headspace analysis. Many thanks go to the Foght lab members past and present, as well as the residents of the Micro wing on the fourth floor, who were always supportive and really made working in the lab a joy.

For financial support I would like the Natural Sciences and Engineering Research Council of Canada. Funding for this research was provided by the Norwegian Research Council.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 Background

Extremophiles are organisms that thrive under extreme conditions that most other organisms cannot survive. As such, they are of great interest for examining the limits of conditions that are permissible for life to exist. Such studies have a wide range of applications from bio-prospecting desirable organisms, to aiding the field of astrobiology in the search for life on other planets, and the discovery and better identification of environments on Earth where life may be found (Rothschild and Mancinelli 2001). Not only are extremophiles appealing from an academic perspective, but due to their intrinsically "extreme" nature these organisms are also desired for biotechnology to study function under conditions that render other organisms and their enzymes inactive.

When examining temperature as a parameter that can either permit or exclude life there are two types of extremophiles; psychrophiles that grow optimally at temperatures <15°C and thermophiles that grow optimally at temperatures >50°C. The non-extreme organisms that grow optimally at moderate temperatures are known as mesophiles (Pikuta et al. 2007). Thermophiles are of particular interest due to their ability to withstand the denaturing effect that higher temperatures can have on biological molecules such as proteins and DNA (Li et al. 2005), and are also the focus of this thesis.

The phylogenetic position of the two bacterial lineages harbouring hyperthermophilic organisms, the Thermotogae and the Aquificae, at the base of the 16S-rRNA tree of life (Fig. 1.1),

has been used as support for the hypothesis that the ancestor of the bacterial domain was a hyperthermophile (Achenbach-Richter et al. 1987). A similar pattern is found in Archaea, where thermophilic lineages are found at the base of the rRNA gene tree (Fig. 1.1). Together with the high temperature conditions of early Earth, this led to the hypothesis that the last universal common ancestor (LUCA) was thermophilic in nature (Pace 1991). There are, however, other lines of evidence that suggest that LUCA was a mesophile or a moderate thermophile (Boussau et al. 2008; Brochier-Armanet and Forterre 2006).

Regardless of the optimal growth temperature of LUCA however, the ancestors of present day bacterial and archaeal lineages may have had to modify perhaps their entire cellular structure and protein composition to transition between mesophilic and thermophilic lifestyles at least once during their evolutionary history (Boussau et al. 2008). Moreover, Hobbs et al. (2012) have presented evidence, using protein thermo-stability as a proxy for thermal adaptation, that thermophily has been lost and gained throughout the evolutionary history of the genus *Bacillus*, suggesting that these transitions may not be too difficult and that temperature adaptations may be constantly changing traits in many lineages. Using ancestral gene sequence reconstruction, presumptive ancestral versions of different ages of the LeuB enzyme (3-isopropylmalate dehydrogenase) were recreated and shown to transition from thermophilic to mesophilic and back to thermophilic temperature optima when going forward in evolutionary time from the ancestor of *Bacillus* (Hobbs et al. 2012). This chapter will discuss properties of thermophiles and how these may relate to transitions between thermophily and mesophily, focusing on the bacterial phylum Thermotogae.

1.2 The Thermotogae

Bacteria belonging to the Thermotogae were first isolated by Karl Stetter and colleagues in 1986 from geothermally heated sea floors (Huber et al. 1986). Their name derives from the unique outer membrane that balloons over each end of the cell, known as the "toga" (Fig. 1.2) (Conners et al. 2006). These Bacteria grow at temperatures up to 90°C, are all anaerobes and are strict organotrophs, capable of growing on a wide range of complex substrates. They are found in hot ecosystems all over the world including thermal springs, hydrothermal vents and petroleum reservoirs (Huber and Hannig 2006; Ollivier and Cayol 2005).

Much of the research on these organisms has focused on high temperature hydrogen and sulfur metabolism (Schut et al. 2012) and polysaccharide degradation (Chhabra et al. 2002), both of which have potential applications in biotechnology. In particular, carbohydrate utilization by the Thermotogae model strain *Thermotoga maritima* has been examined by studying its carbohydrate transporters and glycoside hydrolases (Chhabra et al. 2002; Frock et al. 2012), and by investigating its entire carbohydrate utilization network through a comparative genomic and transcriptomic approach (Rodionov et al. 2013).

The study of temperature adaptation in this group has mainly focused on the hyperthermophilic organism *T. maritima*, the first Thermotogae to be isolated, or on the phylum as a whole. Wang et al. (2012) examined the *T. maritima* proteome at four temperatures spanning its growth range. They found that at supra-optimal temperature both chaperones and carbohydrate metabolism proteins have higher abundance. Other temperature studies have focused on the accumulation of organic compounds and/or polyamines and the elucidation of their biosynthetic pathways, which was examined in *T. maritima* and *Petrotoga miotherma* (Jorge et al. 2007; Oshima et al. 2011; Rodionova et al. 2013; Rodrigues et al. 2009; Zellner and Kneifel 1993).

Dahle et al. (2011) on the other hand, examined 33 Thermotogae genomes to link evolutionary distance to temperature optimum variability. They examined the evolution of the 16S rRNA gene sequences and used these differences to predict the temperature optimum for uncultured members of the family. Another study by Green et al. (2013) used 30 Thermotogae genomes to perform ancestral sequence reconstruction for every node of their phylogenetic tree. Using these presumptive ancestral sequences they suggested that the thermotolerant Thermotogae lineages are under directional selection and that it is easier to transition from high to low optimal temperature than vice versa.

The adaptation of the *T. maritima* genome to life at high temperatures was highlighted by a multiple 'omics' approach (Latif et al. 2013). They found that its genome is streamlined with little intergenic space and a higher number of genes per transcription unit, which was inferred to result from life at higher temperatures. When the genome of *T. maritima* MSB8 was sequenced (Nelson et al. 1999), it was discovered that 24% of its open reading frames (ORFs) showed greatest similarity to sequences from Archaea, suggesting that these genes have been acquired by lateral gene transfer (LGT) from Archaea in the same environment. Comparative genomic analyses of additional Thermotogae genomes have confirmed a strong lateral connection to Archaea and an even larger fraction of Firmicutes genes in these genomes (Mongodin et al. 2005; Nesbø et al. 2009; Zhaxybayeva et al. 2009; Zhaxybayeva et al. 2012). As a major force that has shaped the genomes of the Theremotogae, LGT may have also been important for the acquisition and development of the temperature ranges of the various Thermotogae lineages. Indeed the inferred archaeal origin of the enzyme reverse gyrase has been used to suggest that hyperthermophily was acquired by Bacteria from Archaea by LGT (Forterre et al. 2000).

Although it was long thought that the Thermotogae only harbored thermophiles and hyperthermophiles (10 of 11 genera are entirely composed of thermophiles or hyperthermophiles) (Fig. 1.3), mesophilic Thermotogae members have recently been detected and isolated from cool hydrocarbon-impacted sites such as oil reservoirs and polluted sediments (Nesbø et al. 2006; Nesbø et al. 2010; Nesbø et al. 2012). The genus *Mesotoga* and its closest thermophilic relative, *Kosmotoga*, could provide an ideal model for studying the transition from a thermophilic to a mesophilic lifestyle.

There is a large knowledge base for some of the thermophilic members of the order, and protein structures, for instance, are available for a large portion of the *T. maritima* proteome (DiDonato et al. 2004; Lesley et al. 2002). Studying the *Kosmotoga* and *Mesotoga* lineages could therefore provide insight into the changes that took place during transition of their temperature tolerance and how this may have happened. In particular, *Kosmotoga olearia*, the focus of this thesis, has an unusually wide temperature growth range. Understanding this phenomenon could clarify mechanisms by which Bacteria adapt to new temperatures.

1.3 Thermal adaptation mechanisms

Regardless of whether cells are responding to transient temperature increases within their growth range or evolving to an alternate growth range, changes in temperature require major modifications across the cell to optimize cell function and growth. Table 1.1 summarizes some of the various strategies used by cells for temperature adaptation discussed below.

1.3.1 General cellular adaptations to thermophily

1.3.1.1 Cellular membranes; it's easier for Archaea

The cell membrane is critical to cell function for maintaining and separating the interior cell environment from the exterior environment. In order to serve its function, a lipid membrane must be impermeable to most solutes and maintain a liquid crystalline phase, even under stress (Koga 2012). As temperatures increase however, membrane integrity and impermeability become compromised until cell death occurs (Chang 1994). Thermophiles therefore must maintain their membranes under conditions that could inactivate those of mesophiles. Bacteria and Archaea handle this challenge very differently due to the dissimilar structures of their membrane lipids, reviewed by Koga and Morii (2005), Koga (2012), and Mansilla et al. (2004) and summarized below.

In archaeal membrane polar lipids the hydrocarbon chains consist of methyl-branched isoprenoids bound to glycerol by ether linkages (Koga and Morii 2005). In addition, many species contain bipolar lipids that span the membrane via a tetraether core (Koga and Morii 2005). Maintaining their membranes in the liquid crystalline state at different temperatures is not a problem for archaeal cells since archaeol- and caldarchaeol- (the core lipids that link to the polar head groups) based lipid membranes are considered to be in a liquid crystalline phase in the temperature range of 0-100°C (Blocher et al. 1984; Koga and Morii 2005; Koga 2012; Ulrih et al. 2009). This would greatly facilitate transition between thermophily and mesophily as membrane functions would be preserved and would require fewer modifications in response to temperature changes since the core lipid structure would not need to change.

Bacterial polar membrane lipids, on the other hand, consist mainly of straight-chain fatty acids, which are bound to the glycerol mainly by ester linkages (Koga and Morii 2005). Bacteria respond to various temperatures by altering the composition (length, degree of branching and degree of unsaturation) of their fatty acid chains (Table 1.1) to maintain membrane fluidity (Mansilla et al. 2004). This also alters the phase transition temperature of the lipid membranes, resulting in a narrower temperature range at which the membranes are in their liquid crystalline state and at their lowest permeability (Koga and Morii 2005). This presents a substantial barrier to transition between thermophily and mesophily because Bacteria are limited by the types of fatty acids they can produce. *E. coli*, for example, can produce three: one saturated fatty acid, palmitic (hexadecanoic) acid, and two unsaturated fatty acids, palmitoleic (cis-9-hexadecenoic) acid and cis-vaccenic (cis-11-octadecenoic) acid (Mansilla et al. 2004), which limits the range of temperatures to which it can respond. In contrast, hyperthermophilic Thermotogae have unusual membrane-spanning diabolic fatty acids in their membrane, which are thought to be an adaptation to high temperature growth (Carballeira et al. 1997; Damsté et al. 2007). In agreement with this, these diabolic fatty acids are not found in the membrane of the mesophilic *Mesotoga prima* (Nesbø et al. 2012).

In addition to the lipid structure of cell membranes, studies of integral membrane proteins of the DedA family have shown that integral membrane proteins in general can be important for the temperature tolerance of an organism (Thompkins et al. 2008). While the functions of proteins in this family are not fully understood, mutations to some members has been shown to cause temperature sensitivity and cell division defects (Thompkins et al. 2008). So while the fatty acid composition of the membrane is crucial for its function, integral membrane proteins can also play a significant role, particularly with respect to the upper temperature limit of an

organism's growth range. However, Doerrler et al. (2013) found that although there are DedA homologs in the Thermotogae phylum, there are also at least five Thermotogae genomes (*Thermotoga lettingae* TMO, *Thermotoga maritima* MSB8, *Fervidobacterium nodosum* Rt17 B1, *Thermosipho melanesiensis* BI429, *Petrotoga mobilis* SJ95) from both hyperthermophiles and moderate thermophiles that lack any significant DedA homolog.

1.3.1.2 Nucleic acids; keeping the strands together

High temperatures denature double stranded DNA and secondary structures in RNA. This presents a problem for thermophiles, particularly hyperthermophiles that can grow as high as 122°C (Takai et al. 2008). These organisms must maintain their chromosomes in an orderly state for both efficient packaging as well as coordinated gene expression.

To survive the damaging effects of high temperature thermophiles can either continuously repair their DNA, or protect it. The archaeon *Pyrococcus abyssi* has a highly efficient DNA repair system that continuously repairs temperature-induced DNA damage (Jolivet et al. 2003). One of the primary ways thermophiles protect their DNA is with thermostable proteins analogous to eukaryotic histones (Table 1.1): Higashibata et al. (1999) found that HpkA and HpkB from *Thermococcus kodakaraensis* (formerly *Pyrococcus kodkaraensis*) dramatically increase the melting temperature of a given DNA sequence upon binding, with HpkB being able to raise the melting temperature of poly(dA-dT) DNA by > 20°C. They suggested that these proteins play a major role in the stabilization of *T. kodakaraensis* chromosomes *in vivo*. Furthermore, Mukherjee et al. (2008) found that the histone-like protein HU in the bacterium *T. maritima* stabilizes and protects its DNA.

Thermophiles can also use polyamine compounds to stabilize their DNA and RNA, as well as many other cellular components. Higashibata et al. (2000) found that multivalent polyamine compounds such as putrescine, spermidine, and spermine, or their acetylated forms, could further compact histone-bound DNA in *T. kodakaraensis*, stabilizing it at temperatures as high as 90°C. Zellner and Kneifel (1993) found that the polyamines caldopentamine and caldohexamine increased in concentration with increased temperature in *Thermotoga* species, suggesting a role in thermal adaptation. Moreover, Mikulik and Anderova (1994) found that polyamine compounds were present in the bacterium *Calderobacterium hydrogenophilum*, where they stabilized the 70S initiation complex of ribosomes.

In addition to polyamine compounds thermophiles can use unique RNA modifications that confer thermo-stability (McCloskey et al. 2001). Such modifications, particularly to the tRNAs such as 2'-*O*-methyladenosine or N^2,2'-*O*-dimethylguanosine, are often growth temperature-specific, even among closely related lineages (McCloskey et al. 2001).

Another way thermophiles can stabilize their genomic DNA is with the enzyme reverse gyrase, which is found almost exclusively in hyperthermophiles (Forterre 2002). This enzyme positively supercoils DNA to increase its thermo-stability. However, while deletion of the gene encoding this enzyme from *T. kodakaraensis* results in slower growth at high temperatures (90°C), it does not abolish growth, showing that this enzyme is not essential for hyperthermophilic growth, as was once thought (Atomi et al. 2004).

While some of these adaptations for nucleic acid stabilization have only been found in thermophiles (such as reverse gyrase (Forterre 2002), certain RNA modifications (McCloskey et al. 2001) and thermostable histones (Higashibata et al. 1999)), others are found in mesophiles as well. For instance, the same polyamines found in *Thermotoga* are also found in mesophilic

microalgae (Nishibori et al. 2009). Transition between thermophily and mesophily then may only require a re-purposing of certain cellular constituents, rather than removing and/or acquiring them.

In addition to cellular components interacting with nucleic acids for stabilization, the nucleic acids themselves have adapted to thermo-stability in thermophiles. The structure of G:C bonds was long thought to play a part in optimal growth temperature due to the three hydrogen bonds they share compared to the two shared by A:T bonds. However, while genome-wide G+C content has been shown not to correlate with optimal growth temperature (Galtier and Lobry 1997; Hurst and Merchant 2001; Zeldovich et al. 2007), the G+C content of specific genetic features does. The G+C content of secondary structures of rRNA and tRNA molecules, specifically in the stem structures, has been found to increase with optimal growth temperature (Galtier and Lobry 1997; Kimura et al. 2013; Zhaxybayeva et al. 2009). This has enabled the 16S rRNA gene to be used in temperature studies in the Thermotogae as outlined above (Dahle et al. 2011; Green et al. 2013).

1.3.1.3 Compatible solutes; the power of redundancy

Compatible solutes are organic compounds that are accumulated by cells under stressful conditions such as osmotic stress and heat stress (Santos et al. 2011). Several compatible solutes are associated with thermophiles including di-*myo*-inositol phosphate, mannosyl-di-*myo*-inositol phosphate, mannosylglyceramide, and diglycerol phosphate (Borges et al. 2010; Gonçalves et al. 2012) and studies of thermophiles often identify novel solutes (Jorge et al. 2007; Rodrigues et al. 2009). However, while these compounds are thermophile-specific, they are not the only

compatible solutes used to deal with heat stress. When the ability to synthesize di-*myo*-inositol phosphate was removed from *T. kodakarensis* by deleting a key synthesis gene, Borges et al. (2010) found that the growth of this archaeon was unaffected and that aspartate accumulated as an alternative compatible solute. In addition, studies of the solute pools in Thermotogae bacteria found accumulation of multiple solutes under the stress conditions examined (Jorge et al. 2007; Rodrigues et al. 2009). This suggests compatible solutes play a yet-unknown role in thermal protection, but that there is functional redundancy among the solutes that is not specific to an optimal growth temperature.

1.3.1.4 Chaperones, chaperonins and heat shock response

Chaperones are large protein complexes that assist in the proper folding and re-folding of proteins, while the chaperonins represent a subclass of chaperones that have been extensively studied, particularly GroEL in *E. coli* (Large et al. 2009). Distribution of the various chaperone families varies across Bacteria and Archaea, but some are considered indispensable, such as the chaperonins (Large et al. 2009). These proteins help fold new polypeptides, as well as re-fold and rescue proteins that have been inactivated due to stress (Techtmann and Robb 2010). A major stressor that triggers chaperone-mediated protein repair is heat shock, which has resulted in many chaperones being named heat shock proteins (Hsp) (Large et al. 2009). By preventing inactivation and aggregation of proteins at high temperatures, this ubiquitous system is thought to be especially important in thermophiles like *T. maritima*, which employ chaperones in both their unstressed and heat-stressed states (Pysz et al. 2004). For example, the predicted chaperone TM1083 in *T. maritima* is thought to stabilize the DNA gyrase enzyme at temperatures near

optimal growth (Canaves 2004). Furthermore, Shockley et al. (2003) found that the hyperthermophilic archaeon *P. furiosus* relies on chaperone-mediated rescue, stabilization by compatible solutes, and proteolysis of damaged proteins to deal with heat shock-induced thermal stress.

A key distinction between well-studied bacterial mesophiles and the hyperthermophile *T. maritima* is the seemingly lacking regulation of most of its proteases in response to temperature stress (Conners et al. 2006). Its genome lacks genes encoding proteins that are major regulators of the mesophilic proteolytic response, including any major heat-shock sigma factor (Conners et al. 2006; Pysz et al. 2004), suggesting that *T. maritima* may gain a survival advantage from constitutive expression of most proteases. Similarly these regulators rpoH and ctsR are also not found in the genome of *K. olearia* (Swithers et al. 2011), another species of the Thermotogae.

1.3.2 Thermal adaptation at the protein level: Thermophiles vs. Mesophiles

Although chaperones aid in proper folding and maintenance of proteins under high temperature conditions, proteins from thermophilic organisms are themselves adapted to high temperature. This adaptation is required to maintain activity at temperatures that would denature mesophilic homologs and is found at all levels of protein structure, from primary through quaternary (See below). However, complicating matters, Gu and Hilser (2009) found that protein thermo-stability is modified to different extents across the proteome and depends on the functional role of the protein, with those involved in metabolism appearing to be the main targets of thermal adaptation.

While there are many examples of specific thermal stabilizing characteristics and interactions at each of the four levels of globular protein structure (reviewed by Imanaka (2011)

and Li et al. (2005), there is no universal property that confers thermo-stability. Rather, it is the unique combination of factors at all levels of structure that grants high temperature activity in globular proteins. Taylor and Vaisman (2010) concluded that increased thermo-stability is often due to slight differences in sequence and structure and that thermophilic and mesophilic orthologs are typically very similar proteins. Below is a brief overview of some of the pathways to temperature adaptation in proteins.

Protein primary structure is the amino acid sequence of the polypeptide chain. Ultimately, the properties and sequence of the amino acids determine the final higher level structures of the protein. Comparisons of thermophilic and mesophilic protein compositions have revealed several trends in amino acid composition at the global proteome level. One of the first such trends detected is the greater numbers of charged (D,E,K,R) versus polar (N,Q,S,T) amino acids (CvP bias) in soluble proteins from hyperthermophiles (Cambillau and Claverie 2000). Cambillau and Claverie (2000) suggested this bias reflects stronger ionic interactions between charged amino acids being favored over hydrogen-bonding interactions to increase stability of thermophilic proteins. This global proteome relationship has been confirmed by several subsequent studies (Gao and Wang 2012; Holder et al. 2013; Suhre and Claverie 2003), and may even hold true for individual soluble proteins (Taylor and Vaisman 2010).

Zeldovich et al. (2007) found that the total fraction of the amino acids IVYWREL in a proteome was most strongly correlated with optimal growth temperature. They also suggest that an increased A+G content in coding DNA among thermophiles is due to adaptation to favor these amino acids. Zhaxybayeva et al. (2009) found a linear correlation between optimal growth temperature and compositional features of proteins using both CvP bias and proportion of IVYWREL amino acids. Using this correlation they found that the composition of the

reconstructed ancestral Thermotogae genome suggested that this organism had a slightly higher optimal growth temperature than present day Thermotogae bacteria.

Finally, thermo-stable proteins also tend to be enriched in amino acids that contribute to a strong hydrophobic core. Larger, aliphatic amino acids with more branches are favored at positions that fill cavities, which ultimately strengthens the protein through increased hydrophobic interactions (Clark et al. 2004). Taylor and Vaisman (2010), however, found that this was only a moderately good indicator of protein thermo-stability.

Protein secondary structure describes the local folding of polypeptide sequences. This includes regular structures like α-helices and β-sheets, or irregular structures like β-turns, coils and loops. These are formed primarily by hydrogen bond interactions between the backbone and side chain elements of the amino acids. In addition to having secondary structures that facilitate tighter packing and rigidity at the tertiary level, thermophilic proteins also tend to have secondary structures that are more stabilized than mesophilic counterparts (Facchiano et al. 1998; Koga et al. 2008; Prakash and Jaiswal 2010). For example, thermo-stable proteins have been reported to have a larger fraction of their amino acid residues arranged in α-helices than mesophilic proteins (Prakash and Jaiswal 2010).

Protein tertiary structure is the arrangement of a folded polypeptide chain in three-dimensional space. This folding is influenced by disulfide bridges and electrostatic interactions within the molecule, in addition to hydrophobic interactions and hydrogen bonding within the molecule as well as between the molecule and solvent. Thermophilic proteins tend to have conformations that are more rigid and more tightly packed, with reduced entropy of unfolding and conformational strain compared to their mesophilic counterparts (Li et al. 2005). Taylor and Vaisman (2010) found that the strongest contributors to thermo-stability are increased ion pairs on

the protein surface combined with a more strongly hydrophobic interior. In agreement with this, Knapp et al. (1997) found that the glutamate dehydrogenase enzymes of the hyperthermophilic bacterium *T. maritima* and hyperthermophilic archaeon *P. furiosus* have smaller hydrophobic accessible surface area (ASA), and greater charged ASA than glutamate dehydrogenase from the mesophilic bacterium *Clostridium symbiosum*. Since few differences were found between the thermophilic and mesophilic enzymes, this tighter packing is thought to contribute to the thermal stability of the proteins.

Protein quaternary structure is the arrangement of multiple folded polypeptide chains into a multimeric complex. In globular proteins this level of structure is formed and maintained by many of the same forces that contribute to the tertiary structure of a protein, but between polypeptide chains rather than within them. These forces include disulfide bridges, electrostatic interactions, hydrophobic interactions and hydrogen bonding. In thermo-stable proteins, greater numbers of these interactions are favored, or stronger interactions are favored over weaker ones (Li et al. 2005).

One way of achieving greater stability is to increase the number of subunits in thermophilic proteins. One example of this is the malate dehydrogenase (MDH) enzyme of the thermophilic bacterium *Chloroflexus aurantiacus*. Bjørk et al. (2003) found that MDH, which is usually a dimer in mesophiles, is a tetramer in *C. aurantiacus*. They hypothesized that the additional dimer-dimer interface of the tetrameric MDH provided thermal stability due to the higher number of inter-polypeptide interactions compared to the mesophilic dimers. To test this, they introduced a disulfide bridge that would strengthen dimer-dimer interaction further, and found that the new enzyme had a melting temperature 15°C higher than the wild-type enzyme. In addition, Bjørk et al. (2004) found that removing excess negative charge at the dimer-dimer

interface by replacing a glutamate residue with either glutamine or lysine resulted in an increase of apparent melting temperature by ~ 24°C.

1.4 Cooling down – Not a simple transition for the individual gene

As outlined above, adaptation to a high optimal growth temperature is achieved differently by Bacteria and Archaea, differently by one species than another, and differently even by one protein than another within the same organism. Given that all of these factors combine in unique ways to permit growth within a specific temperature range, how could a shift in permissive temperature range be accomplished? While some of these properties are universal to thermophiles and mesophiles, such as chaperones and compatible solutes, others like membrane properties would have to be radically altered to accommodate large changes in temperature range (Table 1.1).

Changing a few key proteins may have global stabilizing effects on the whole cell. For instance, Gu and Hilser (2009) found that some of the proteins whose stability appears most affected by thermal adaptation are involved in production of compatible solutes that stabilize other proteins. Such changes would reduce the need to modify the stability of all components of the proteome. Similarly, Endo et al. (2006) found that it is possible to lower the maximal growth temperature of an organism through changes to a single protein. By replacing the chromosomal copy of *groEL* in *Bacillus subtilis* 168 (growth range 11 -52°C) with a psychrophilic *groEL* from *Pseudoalteromonas* sp. PS1M3 (growth range 4-30°C), they noted a 2°C reduction in the maximal growth temperature of the newly constructed *B. subtilis* strain. While this does not constitute a true *shift* in growth temperature range or even a change to optimal growth temperature, it suggests that changes to a key protein involved in temperature adaptation and

response could alter the temperature at which an organism is able to grow. Even a few changes that reduce the thermo-stability of such a protein could have this effect, provided the protein still has activity at lower temperatures. If this were to happen in a thermophile that has a wide temperature growth range, and so presumably already has mechanisms for growth at lower temperatures, the growth temperature range could effectively be lowered. Theoretically, if this process were repeated over time with adaptation after each shift to new lower growth temperatures, thermophily could eventually give way to mesophily.

1.5 Role of Lateral Gene Transfer in Temperature Adaptation – Acquisition of Already 'Adapted' Genes

Lateral gene transfer (LGT) is a major force in prokaryotic evolution, allowing rapid adaptation to changes in the environment by acquiring clusters of genes or single genes that confer a selective advantage (Boucher et al. 2003). Genes encoding proteins linked to adaptation to higher or lower growth temperatures have indeed been shown to be involved in LGT (reviewed by Boucher et al. 2003). Reverse gyrase is a classic example of transfer of a single gene that is thought to have been crucial for evolutionary adaptation to high temperatures by hyperthermophilic Bacteria (Brochier-Armanet and Forterre 2006). As previously discussed, this protein is not essential for hyperthermophilic growth by *T. kodakarensis*, but loss of this gene leads to a retardation in growth at higher temperatures (Atomi et al. 2004). Its role in adaptation to hyperthermophily is probably linked to protection of DNA against degradation and damage (Brochier-Armanet and Forterre 2006). Phylogenetic analyses suggest two ancient acquisitions of this gene by bacterial

lineages from Archaea, followed by secondary transfers among Bacteria (Brochier-Armanet and Forterre 2006).

Similarly, the compatible solute di-*myo*-inositol phosphate is thought to be important for heat tolerance in thermophiles and hyperthermophiles (Borges et al. 2010). Two key genes involved in the synthesis of this compound (inositol-1-phosphate cytidylyltransferase and di-*myo*-inositol phosphate phosphate synthase) are suggested to have been horizontally transferred from an archaeal lineage to hyperthermophilic marine *Thermotoga* species, while in other lineages they are predicted to have fused before being horizontally transferred among Bacteria and other archaeal lineages (Gonçalves et al. 2012).

Since the Thermotogae have been reported to have undergone extensive lateral gene transfer (Nelson et al. 1999), this force may have also shaped their temperature tolerance. Indeed a large fraction of the mesophilic *M. prima* genome was likely acquired by lateral gene transfer, implicating LGT in its mesophilic lifestyle (Zhaxybayeva et al. 2012).

1.6 Temperature adaptation in Kosmotoga and Mesotoga

The discovery of a mesophilic Thermotogae lineage (*Mesotoga*) raised the possibility that (hyper)thermophily was not ancestral to the phylum. However, analysis of protein signatures (CvP and IVYWREL) suggests that ancestral Thermotogae proteomes were thermophilic (Zhaxybayeva et al. 2009). Furthermore, the G+C content of ribosomal RNA, which correlates with optimal growth temperature, also suggests that the ancestral Thermotogae 16S rRNA belonged to a thermophile (Zhaxybayeva et al. 2009).

The genus *Mesotoga* is unique among the Thermotogae due to optimal growth between 37-45°C (Nesbø et al. 2012). Initially *Mesotoga* was only detected using molecular tools, which was further complicated by the fact that its 16S rRNA gene appears to amplify poorly using traditional PCR primers (Nesbø et al. 2006; Nesbø et al. 2010). *Mesotoga prima* is the first described member of the genus (Nesbø et al. 2012), which also includes two other members: *Mesotoga sp.* PhosAc3 (Ben Hania et al. 2011) and *Mesotoga infera* (Ben Hania et al. 2013). The genome of *M. prima* at 2.97 Mb is considerably larger than any previously sequenced Thermotogae genome, which range between 1.86 and 2.30 Mb (Zhaxybayeva et al. 2012). This larger size is due to both higher numbers of protein-coding genes and larger intergenic regions. In particular, the *M. prima* genome contains more genes encoding proteins involved in regulatory functions, for instance those involved in regulation of transcription. Phylogenetic analysis of all the ORFs in its genome suggests this lineage has undergone extensive LGT with diverse mesophilic lineages, and that LGT has aided its transition from a thermophilic to a mesophilic lifestyle (Zhaxybayeva et al. 2012). Analyses of the amino acid composition of its proteome gave a unimodal distribution of CvP values for its protein-coding genes in the mesophilic range, indicating that native *M. prima* proteins have also changed in response to its evolved mesophilic lifestyle.

The closest relative of the *Mesotoga* lineage is the thermophilic lineage *Kosmotoga* (Fig. 1.3). Members of this genus have been found in hydrothermal sediment (L'Haridon et al. 2014; Nunoura et al. 2010) and oil production fluid (DiPippo et al. 2009; Feng et al. 2010). They are anaerobic chemoorganotrophs able to ferment carbohydrates and peptides (Nunoura et al. 2010) and to produce molecular hydrogen (DiPippo et al. 2009; Feng et al. 2010). The first bacterium of this genus to be isolated was *Kosmotoga olearia* (DiPippo et al. 2009), which is the type strain (Nunoura et al. 2010). *K. olearia* grows optimally at 65°C and has a reported growth range of 20-

80°C (DiPippo et al. 2009). Not only does this include an uncommonly low temperature of growth for a thermophilic Thermotogae, but it also represents an unusually wide temperature growth range.

This property is intriguing for several reasons. First, in order for an organism to grow it must maintain protein activity and membrane integrity. Every living organism has adapted to do this at a certain temperature range, but how these requirements can be maintained over a 60°C range is cryptic. Second, as explained above, this may very likely have facilitated the transition of *Mesotoga* from thermophily to mesophily because the capacity to grow at lower temperatures already exists. If maximal growth temperature is lowered (e.g., by loss of thermal stability of a key protein), growth and further adaptation can still continue at lower temperatures.

Because a 60°C temperature growth range is exceptionally wide, many questions can be asked regarding the nature of growth of *K. olearia*. Given that it is a bacterium, and Bacteria have to modify their membrane lipid composition at different temperatures, how does it maintain membrane fluidity and integrity at both 80°C and 20°C? How do proteins that are designed to stabilize the DNA at high temperature still allow chromosome unwinding at low temperature for critical processes like transcription and replication? Are the genes encoding enzymes involved in the synthesis of polyamines and other compatible solutes only expressed under stress conditions or under the high temperature un-stressed conditions too? At what points within the growth temperature range is *K. olearia* "stressed"? Is its broad temperature growth range the result of highly effective stress responses? How does *K. olearia* maintain functionality of all its proteins across its growth temperatures? Essentially, how does a seemingly thermophilic microbe grow at mesophilic temperatures?

1.7 Thesis Outline

In this thesis I examine the responses of *K. olearia* to various temperatures within its wide temperature growth range and investigate thermal adaptation in the Thermotogae. RNA-Sequencing (RNA-Seq) was used to generate a transcriptome of *K. olearia* at different temperatures (Chapter 2) to determine which genes are differentially expressed. Candidate genes include ribosomal proteins, cold shock proteins and DEAD/DEAH box RNA helicases, as found by other temperature studies (Watanabe et al. 2012). Additionally, the expression of chaperones and heat shock proteins is likely to differ, particularly between the extreme high and low temperature transcriptomes. Furthermore, this approach can be used to infer putative functions for some of the many hypothetical genes in the *K. olearia* genome.

Once important temperature-responsive genes were identified, I investigated those found in both *K. olearia* and *Mesotoga* spp. to identify genes likely important for low temperature growth (Chapter 3). Furthermore, I used genomic comparisons with a *Kosmotoga* species having a narrower temperature range to help discriminate between genes that may be important for a wide growth temperature range and those that are part of a more general temperature response.

Table 1.1. Summary of some properties that differ between thermophiles and mesophiles.

Property	Thermophiles	Mesophiles
Cellular Membranes	Tend to contain more saturated fatty acids	Often have more unsaturated and branched fatty acids
Nucleic Acids	High GC content associated with stem loop structures in rRNA Can be stabilized by positive supercoiling and thermostable proteins	Lower GC content in stem loop structures in rRNA
Compatible Solutes	Stabilize nucleic acids and proteins (especially positively charged solutes)	Associated with stress responses
Protein Structure	High charged versus polar (CvP) bias Overrepresentation of IVYWREL amino acids	Low CvP bias
Protein Function	Optimally active at elevated temperatures that would denature many mesophilic proteins	Require moderate temperature for activity

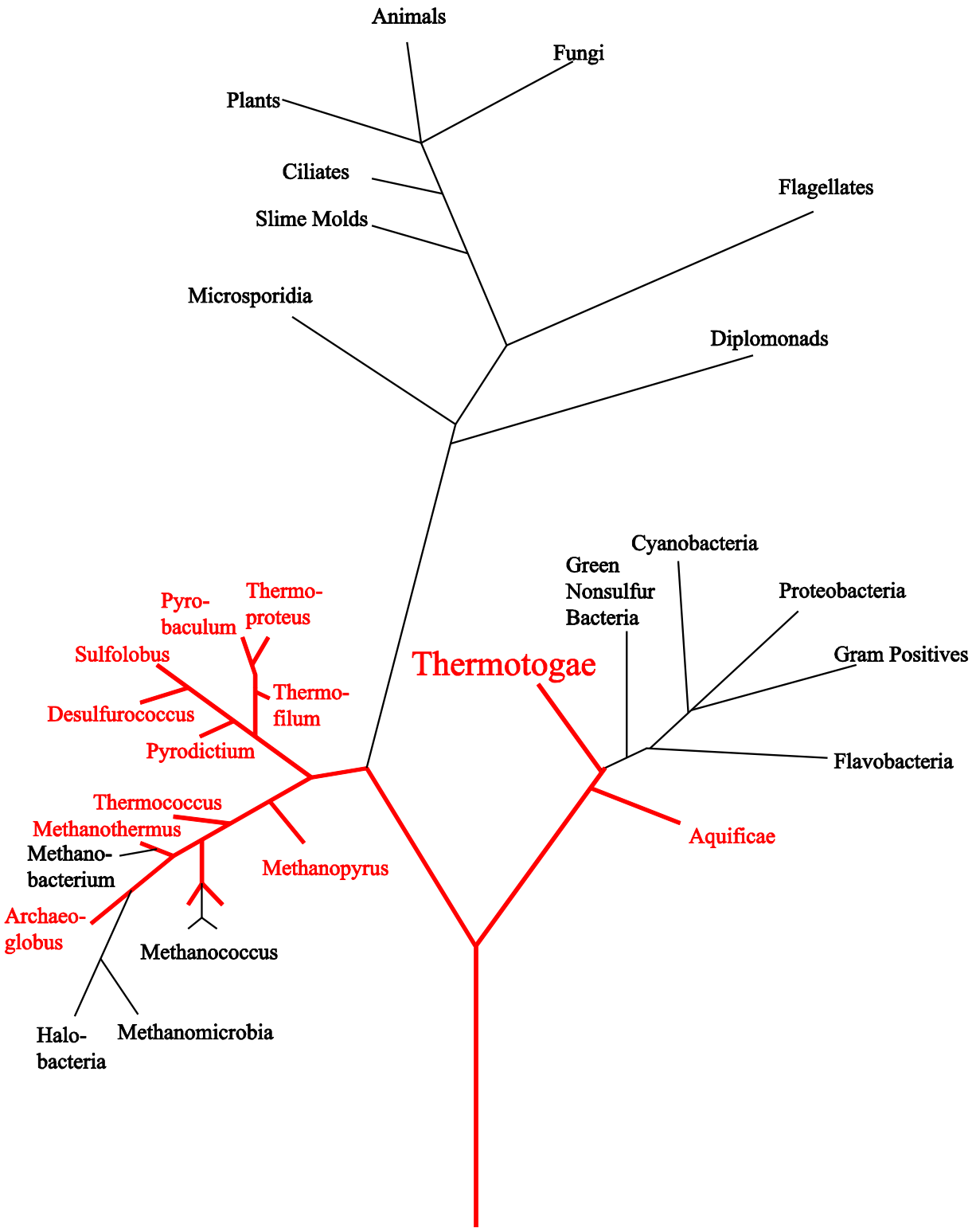


Fig. 1.1. Small subunit rRNA-based tree of life adapted from Imanaka (2011). Red lines and type indicate thermophiles and hyperthermophiles.

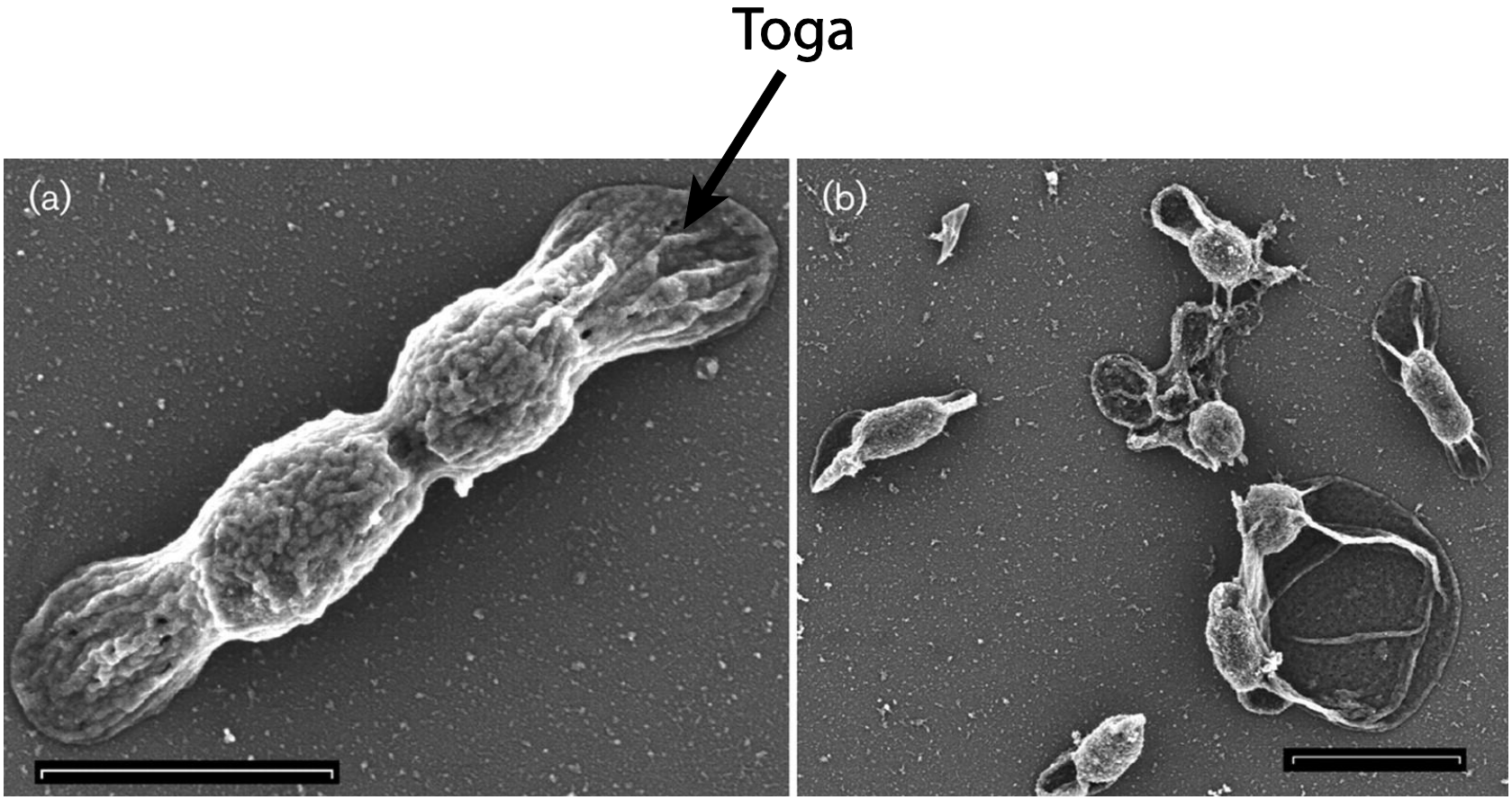


Fig. 1.2. Scanning electron micrographs of cells of *Kosmotoga olearia* TBF 19.5.1 adapted from Dipippo et al. (2009). The toga structure, characteristic of all Thermotogae, is indicated by the arrow. Bars denote 2 µm (a) and 1 µm (b).

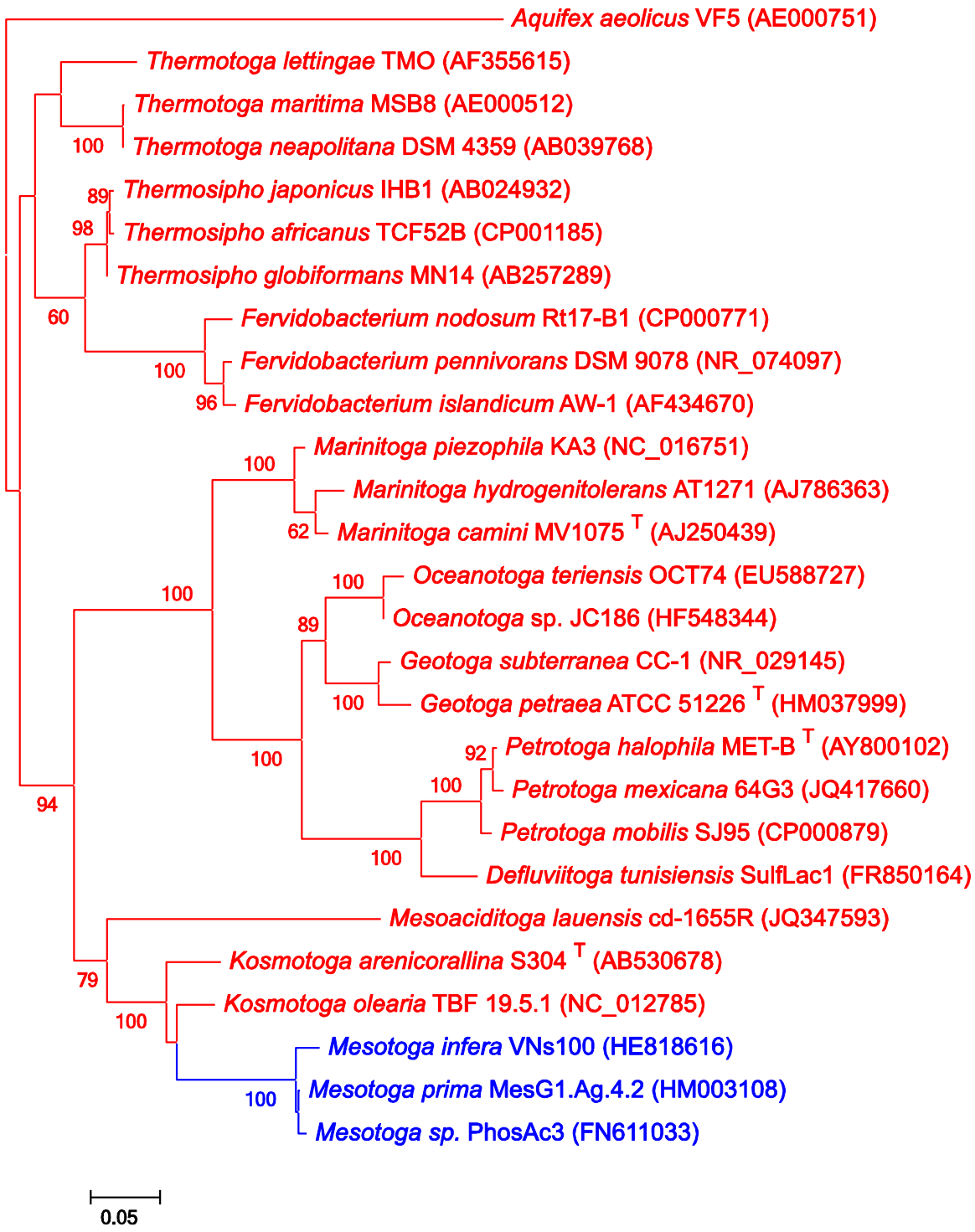


Fig. 1.3. Maximum likelihood tree constructed from 16S rRNA sequences of representatives from all known Thermotogae genera using RAxML (GTR gamma model). MUSCLE alignment of the sequences was edited in GBlocks (default settings) and manually. The Thermotogae tree was rooted on the branch leading to *Aquifex*. Red indicates thermophiles or hyperthermophiles, blue indicates mesophiles. Numbers on branches indicate percentage of occurrence in 100 bootstrap replicates and are shown if they are >50.

1.8 References

Achenbach-Richter, L., Gupta, R., Stetter, K.O., and Woese, C.R. 1987. Were the original eubacteria thermophiles? Syst. Appl. Microbiol. **9**(1-2): 34-39.

Atomi, H., Matsumi, R., and Imanaka, T. 2004. Reverse gyrase is not a prerequisite for hyperthermophilic life. J. Bacteriol. **186**(14): 4829-4833.

Ben Hania, W., Ghodbane, R., Postec, A., Brochier-Armanet, C., Hamdi, M., Fardeau, M., and Ollivier, B. 2011. Cultivation of the first mesophilic representative (mesotoga) within the order *Thermotogales*. Syst. Appl. Microbiol. **34**(8): 581-585.

Ben Hania, W., Postec, A., Aüllo, T., Ranchou-Peyruse, A., Erauso, G., Brochier-Armanet, C., Hamdi, M., Ollivier, B., Saint-Laurent, S., Magot, M., and Fardeau, M. 2013. *Mesotoga infera* sp. nov., a mesophilic member of the order *Thermotogales*, isolated from an underground gas storage aquifer. Int. J. Syst. Evol. Microbiol. **63**: 3003-3008.

Bjørk, A., Dalhus, B., Mantzilas, D., Sirevåg, R., and Eijsink, V.G.H. 2004. Large improvement in the thermal stability of a tetrameric malate dehydrogenase by single point mutations at the dimer-dimer interface. J. Mol. Biol. **341**(5): 1215-1226.

Bjørk, A., Dalhus, B., Mantzilas, D., Eijsink, V.G.H., and Sirevåg, R. 2003. Stabilization of a tetrameric malate dehydrogenase by introduction of a disulfide bridge at the dimer-dimer interface. J. Mol. Biol. **334**(4): 811-821.

Blocher, D., Gutermann, R., Henkel, B., and Ring, K. 1984. Physicochemical characterization of tetraether lipids from *Thermoplasma acidophilum*. Biochemica et Biophysica Acta **778**: 74-80.

Borges, N., Matsumi, R., Imanaka, T., Atomi, H., and Santos, H. 2010. *Thermococcus kodakarensis* mutants deficient in di-*myo*-inositol phosphate use aspartate to cope with heat stress. J. Bacteriol. **192**(1): 191-197.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbø, C.L., Case, R.J., and Doolittle, W.F. 2003. Lateral gene transfer and the origins of prokaryotaic groups. Annu. Rev. Genet. **37**: 283-328.

Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., and Gouy, M. 2008. Parallel adaptations to high temperatures in the Archaean eon. Nature **456**(7224): 942-945.

Brochier-Armanet, C., and Forterre, P. 2006. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. Archaea **2**(2): 83-93.

Cambillau, C., and Claverie, J. 2000. Structural and genomic correlates of hyperthermostability. The Journal of biological chemistry **275**(42): 32383-32386.

Canaves, J.M. 2004. Predicted role for the Archease protein family based on structural and sequence analysis of TM1083 and MTH1598, two proteins structurally characterized through structural genomics efforts. Proteins **56**(1): 19-27.

Carballeira, N.M., Reyes, M., Sostre, A., Huang, H., Verhagen, M.F., and Adams, M.W. 1997. Unusual fatty acid compositions of the hyperthermophilic archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*. J. Bacteriol. **179**(8): 2766-2768.

Chang, E.L. 1994. Unusual thermal stability of liposomes made from bipolar tetraether lipids. Biochem. Biophys. Res. Commun. **202**(2): 673-679.

Chhabra, S.R., Shockley, K.R., Ward, D.E., and Kelly, R.M. 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.

Clark, A.T., McCrary, B.S., Edmondson, S.P., and Shriver, J.W. 2004. Thermodynamics of core hydrophobicity and packing in the hyperthermophile proteins Sac7d and Sso7d. Biochemistry **43**: 2840-2853.

Conners, S.B., Mongodin, E.F., Johnson, M.R., Montero, C.I., Nelson, K.E., and Kelly, R.M. 2006. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. FEMS Microbiol. Rev. **30**(6): 872-905.

Dahle, H., Hannisdal, B., Steinsbu, B.O., Ommedal, H., Einen, J., Jensen, S., Larsen, Ø, Øvreås, L., and Norland, S. 2011. Evolution of temperature optimum in *Thermotogaceae* and the prediction of trait values of uncultured organisms. Extremophiles **15**(4): 509-516.

Damsté, J.S.S., Rijpstra, W.I.C., Hopmans, E.C., Schouten, S., Balk, M., and Stams, A.J.M. 2007. Structural characterization of diabolic acid-based tetraester, tetraether and mixed

ether/ester, membrane-spanning lipids of bacteria from the order *Thermotogales*. Arch. Microbiol. **188**(6): 629-641.

DiDonato, M., Deacon, A.M., Klock, H.E., McMullan, D., and Lesley, S.A. 2004. A scaleable and integrated crystallization pipeline applied to mining the *Thermotoga maritima* proteome. J. Struct. Funct. Genomics **5**: 133-146.

DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N., and Noll, K.M. 2009. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int. J. Syst. Evol. Microbiol. **59**: 2991-3000.

Doerrler, W.T., Sikdar, R., Kumar, S., and Boughner, L.A. 2013. New functions for the ancient DedA membrane protein family. J. Bacteriol. **195**(1): 3-11.

Endo, A., Sasaki, M., Maruyama, A., and Kurusu, Y. 2006. Temperature adaptation of *Bacillus subtilis* by chromosomal *groEL* replacement. Biosci. Biotechnol. Biochem. **70**(10): 2357-2362.

Facchiano, A., Colonna, G., and Ragone, R. 1998. Helix-stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. Protein Eng. **11**(9): 753-760.

Feng, Y., Cheng, L., Zhang, X., Li, X., Deng, Y., and Zhang, H. 2010. *Thermococcoides shengliensis* gen. nov., sp. nov., a new member of the order *Thermotogales* isolated from oil-production fluid. Int. J. Syst. Evol. Microbiol. **60**: 932-937.

Forterre, P., Bouthier De La Tour C, Philippe, H., and Duguet, M. 2000. Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from Archaea to Bacteria. Trends in genetics : TIG **16**(4): 152-154.

Forterre, P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. Trends in genetics : TIG **18**(5): 236-237.

Frock, A.D., Gray, S.R., and Kelly, R.M. 2012. Hyperthermophilic Thermotoga species differ with respect to specific carbohydrate transporters and glycoside hydrolases. Appl. Environ. Microbiol. **78**(6): 1978-1986.

Galtier, N., and Lobry, J.R. 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. **44**(6): 632-636.

Gao, J., and Wang, W. 2012. Analysis of structural requirements for thermo-adaptation from orthologs in microbial genomes. Annals of Microbiology **62**(4): 1635-1641.

Gonçalves, L.G., Borges, N., Serra, F., Fernandes, P.L., Dopazo, H., and Santos, H. 2012. Evolution of the biosynthesis of di-*myo*-inositol phosphate, a marker of adaptation to hot marine environments. Environ. Microbiol. **14**(3): 691-701.

Green, A.G., Swithers, K.S., Gogarten, J.F., and Gogarten, J.P. 2013. Reconstruction of ancestral 16S rRNA reveals mutation bias in the evolution of optimal growth temperature in the Thermotogae phylum. Mol. Biol. Evol. **30**(11): 2463-2474.

Gu, J., and Hilser, V.J. 2009. Sequence-based analysis of protein energy landscapes reveals nonuniform thermal adaptation within the proteome. Mol. Biol. Evol. **26**(10): 2217-2227.

Higashibata, H., Fujiwara, S., Takagi, M., and Imanaka, T. 1999. Analysis of DNA compaction profile and intracellular contents of archaeal histones from *Pyrococcus kodakaraensis* KOD1. Biochem. Biophys. Res. Commun. **258**(2): 416-424.

Higashibata, H., Fujiwara, S., Ezaki, S., Takagi, M., Fukui, K., and Imanaka, T. 2000. Effect of polyamines on histone-induced DNA compaction of hyperthermophilic Archaea. J. Biosci. Bioeng. **89**(1): 103-106.

Hobbs, J.K., Shepherd, C., Saul, D.J., Demetras, N.J., Haaning, S., Monk, C.R., Daniel, R.M., and Arcus, V.L. 2012. On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. Mol. Biol. Evol. **29**(2): 825-835.

Holder, T., Basquin, C., Ebert, J., Randel, N., Jollivet, D., Conti, E., Jékely, G., and Bono, F. 2013. Deep transcriptome-sequencing and proteome analysis of the hydrothermal vent annelid *Alvinella pompejana* identifies the CvP-bias as a robust measure of eukaryotic thermostability. Biology direct **8**(2): 1-16.

Huber, R., and Hannig, M. 2006. Thermotogales. *In* The Prokaryotes. *Edited by* M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, Stackebr and E. t. Springer New York pp. 899-922.

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

Hurst, L.D., and Merchant, A.R. 2001. High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. Proc. R. Soc. Lond. B **268**(1466): 493-497.

Imanaka, T. 2011. Molecular bases of thermophily in hyperthermophiles. Proc. Japan Acad. , Ser. B **87**(9): 587-602.

Jolivet, E., Matsunaga, F., Ishino, Y., Forterre, P., Prieur, D., and Myllykallio, H. 2003. Physiological responses of the hyperthermophilic archaeon "*Pyrococcus abyssi*" to DNA damage caused by ionizing radiation. J. Bacteriol. **185**(13): 3958-3961.

Jorge, C.D., Lamosa, P., and Santos, H. 2007. Alpha-D-mannopyranosyl-(1->2)-alpha-D-glucopyranosyl-(1->2)-glycerate in the thermophilic bacterium *Petrotoga miotherma* - structure, cellular content and function. FEBS J. **274**(12): 3120-3127.

Kimura, H., Mori, K., Yamanaka, T., and Ishibashi, J. 2013. Growth temperatures of archaeal communities can be estimated from the guanine-plus-cytosine contents of 16S rRNA gene fragments. Environ. Microbiol. Rep. **5**(3): 468-474.

Knapp, S., Vos, W.M.D., Rice, D., and Ladenstein, R. 1997. Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. J. Mol. Biol. **267**: 916-932.

Koga, Y. 2012. Thermal adaptation of the archaeal and bacterial lipid membranes. Archaea : doi:10.1155/2012/789652.

Koga, Y., and Morii, H. 2005. Recent advances in structural research on ether lipids from Archaea including comparative and physiological aspects. Biosci. Biotechnol. Biochem. **69**(11): 2019-2034.

Koga, Y., Katsumi, R., You, D., Matsumura, H., Takano, K., and Kanaya, S. 2008. Crystal structure of highly thermostable glycerol kinase from a hyperthermophilic archaeon in a dimeric form. FEBS J. **275**(10): 2632-2643.

Large, A.T., Goldberg, M.D., and Lund, P.A. 2009. Chaperones and protein folding in the Archaea. Biochem. Soc. Trans. **37**: 46-51.

Latif, H., Lerman, J.A., Portnoy, V.A., Tarasova, Y., Nagarajan, H., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Lee, D., Qiu, Y., and Zengler, K. 2013. The genome organization of *Thermotoga maritima* reflects its lifestyle. PLoS genetics **9**(4): e1003485.

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., Wooley, J., Taylor, S.S., Hodgson, K.O., Wilson, I.A., Schultz, P.G., and Stevens, R.C. 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. U. S. A. **99**(18): 11664-11669.

L'Haridon, S., Jiang, L., Alain, K., Chalopin, M., Rouxel, O., Beauverger, M., Xu, H., Shao, Z., and Jebbar, M. 2014. *Kosmotoga pacifica* sp. nov., a thermophilic chemoorganoheterotrophic bacterium isolated from an East Pacific hydrothermal sediment. Extremophiles **18**(1): 81-88.

Li, W.F., Zhou, X.X., and Lu, P. 2005. Structural features of thermozymes. Biotechnol. Adv. **23**(4): 271-281.

Mansilla, M.C., Cybulski, L.E., Albanesi, D., and Mendoza, D.D. 2004. Control of membrane lipid fluidity by molecular thermosensors. J. Bacteriol. **186**(20): 6681-6688.

McCloskey, J.A., Graham, D.E., Zhou, S., Crain, P.F., Ibba, M., Konisky, J., Söll, D., and Olsen, G.J. 2001. Post-transcriptional modification in archaeal tRNAs: identities and phylogenetic relations of nucleotides from mesophilic and hyperthermophilic *Methanococcales*. Nucleic Acids Res. **29**(22): 4699-4706.

Mikulik, K., and Anderova, M. 1994. Role of polyamines in the binding of initiator tRNA to the 70S ribosomes of extreme thermophilic bacterium *Calderobacterium hydrogenophilum*. Arch. Microbiol. **161**: 508-513.

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. 2005. Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. J. Bacteriol. **187**(14): 4935-4944.

Mukherjee, A., Sokunbi, A.O., and Grove, A. 2008. DNA protection by histone-like protein HU from the hyperthermophilic eubacterium *Thermotoga maritima*. Nucleic Acids Res. **36**(12): 3956-3968.

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. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature **399**(6734): 323-329.

Nesbø, C.L., Dlutek, M., Zhaxybayeva, O., and Doolittle, W.F. 2006. Evidence for existence of Mesotogas, members of the order *Thermotogales* adapted to low-temperature environments. Appl. Environ. Microbiol. **72**(7): 5061-5068.

Nesbø, C.L., Kumaraswamy, R., Dlutek, M., Doolittle, W.F., and Foght, J. 2010. Searching for mesophilic *Thermotogales* bacteria: "Mesotogas" in the wild. Appl. Environ. Microbiol. **76**(14): 4896-4900.

Nesbø, C.L., Bradnan, D.M., Adebusuyi, A., Dlutek, M., Petrus, A.K., Foght, J., Doolittle, W.F., and Noll, K.M. 2012. *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the Thermotogales. Extremophiles **16**(3): 387-393.

Nesbø, C.L., Bapteste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., Dlutek, M., Bowman, S., Zhaxybayeva, O., Birkeland, N., and Doolittle, W.F. 2009. The genome of *Thermosipho africanus* TCF52B: lateral genetic connections to the *Firmicutes* and *Archaea*. J. Bacteriol. **191**(6): 1974-1978.

Nishibori, N., Niitsu, M., Fujihara, S., Sagara, T., Nishio, S., and Imai, I. 2009. Occurrence of the polyamines caldopentamine and homocaldopentamine in axenic cultures of the red tide flagellates *Chattonella antiqua* and *Heterosigma akashiwo (Raphidophyceae)*. FEMS Microbiol. Lett. **298**(1): 74-78.

Nunoura, T., Hirai, M., Imachi, H., Miyazaki, M., Makita, H., Hirayama, H., Furushima, Y., Yamamoto, H., and Takai, K. 2010. *Kosmotoga arenicorallina* sp. nov. a thermophilic and obligately anaerobic heterotroph isolated from a shallow hydrothermal system occurring within a coral reef, southern part of the Yaeyama Archipelago, Japan, reclassification of

Thermococcoides shengliensis as *Kosmotoga shengliensis* comb. nov., and emended description of the genus *Kosmotoga*. Arch. Microbiol. **192**(10): 811-819.

Ollivier, B., and Cayol, J. 2005. Fermentative, iron-reducing, and nitrate-reducing microorganisms. *In* Petroleum Microbiology. *Edited by* B. Ollivier and M. Magot. ASM Press, Washington, DC. pp. 71-88.

Oshima, T., Moriya, T., and Terui, Y. 2011. Identification, chemical synthesis, and biological functions of unusual polyamines produced by extreme thermophiles. Methods Mol. Biol. **720**: 81-111.

Pace, N.R. 1991. Origin of life- Facing up to the physical setting. Cell **65**: 531-533.

Pikuta, E.V., Hoover, R.B., and Tang, J. 2007. Microbial extremophiles at the limits of life. Crit. Rev. Microbiol. **33**(3): 183-209.

Prakash, O., and Jaiswal, N. 2010. Alpha-amylase: An ideal representative of thermostable enzymes. Appl. Biochem. Biotechnol. **160**(8): 2401-2414.

Pysz, M.A., Ward, D.E., Shockley, K.R., Montero, C.I., Conners, S.B., Johnson, M.R., and Kelly, R.M. 2004. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8**(3): 209-217.

Rodionov, D.A., Rodionova, I.A., Li, X., Ravcheev, D.A., Tarasova, Y., Portnoy, V.A., Zengler, K., and Osterman, A.L. 2013. Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. Front. Microbiol. **4**: 244.

Rodionova, I.A., Leyn, S.A., Burkart, M.D., Boucher, N., Noll, K.M., Osterman, A.L., and Rodionov, D.A. 2013. Novel inositol catabolic pathway in *Thermotoga maritima*. Environ. Microbiol. **15**(8): 2254-2266.

Rodrigues, M.V., Borges, N., Almeida, C.P., Lamosa, P., and Santos, H. 2009. A unique beta-1,2-mannosyltransferase of *Thermotoga maritima* that uses di-*myo*-inositol phosphate as the mannosyl acceptor. J. Bacteriol. **191**(19): 6105-6115.

Rothschild, L.J., and Mancinelli, R.L. 2001. Life in extreme environments. Nature **409**(6823): 1092-1101.

Santos, H., Lamosa, P., Borges, N., Gonçalves, L.G., Pais, T., and Rodrigues, M.V. 2011. Organic compatible solutes of prokaryotes that thrive in hot environments: The importance of ionic compounds for thermostabilization. *In* Extremophiles Handbook. *Edited by* K. Horikoshi. Springer Japan, Tokyo. pp. 498-516.

Schut, G.J., Boyd, E.S., Peters, J.W., and Adams, M.W.W. 2012. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. FEMS Microbiol. Rev. **37**(2): 182-203.

Shockley, K.R., Ward, D.E., Chhabra, S.R., Conners, S.B., Montero, C.I., Robert, M., and Kelly, R.M. 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microbiol. **69**(4): 2365-2371.

Suhre, K., and Claverie, J. 2003. Genomic correlates of hyperthermostability, an update. J. Biol. Chem. **278**(19): 17198-17202.

Swithers, K.S., DiPippo, J.L., Bruce, D.C., Detter, C., Tapia, R., Han, S., Goodwin, L.A., Han, J., Woyke, T., Pitluck, S., Pennacchio, L., Nolan, M., Mikhailova, N., Land, M.L., Nesbø, C.L., Gogarten, J.P., and Noll, K.M. 2011. 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. **193**(19): 5566-5567.

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., and Horikoshi, K. 2008. Cell proliferation at 122°C and isotopically heavy CH_4 production by a hyperthermophilic methanogen under high-pressure cultivation. Proc. Natl. Acad. Sci. U. S. A. **105**(31): 10949-10954.

Taylor, T.J., and Vaisman, I.I. 2010. Discrimination of thermophilic and mesophilic proteins. BMC Struct. Biol. **10**: S5.

Techtmann, S.M., and Robb, F.T. 2010. Archaeal-like chaperonins in bacteria. Proc. Natl. Acad. Sci. U. S. A. **107**(47): 20269-20274.

Thompkins, K., Chattopadhyay, B., Xiao, Y., Henk, M.C., and Doerrler, W.T. 2008. Temperature sensitivity and cell division defects in an *Escherichia coli* strain with mutations in *yghB* and *yqjA*, encoding related and conserved inner membrane proteins. J. Bacteriol. **190**(13): 4489-4500.

Ulrih, N.P., Gmajner, D., and Raspor, P. 2009. Structural and physicochemical properties of polar lipids from thermophilic archaea. Appl. Microbiol. Biotechnol. **84**(2): 249-260.

Wang, Z., Tong, W., Wang, Q., Bai, X., Chen, Z., Zhao, J., Xu, N., and Liu, S. 2012. The temperature dependent proteomic analysis of *Thermotoga maritima*. PloS one **7**(10): e46463.

Watanabe, T., Kojima, H., and Fukui, M. 2012. Draft genome sequence of a psychrotolerant sulfur-oxidizing bacterium, *Sulfuricella denitrificans* skB26, and proteomic insights into cold adaptation. Appl. Environ. Microbiol. **78**(18): 6545-6549.

Zeldovich, K.B., Berezovsky, I.N., and Shakhnovich, E.I. 2007. Protein and DNA sequence determinants of thermophilic adaptation. PLoS computational biology **3**(1): e5.

Zellner, G., and Kneifel, H. 1993. Caldopentamine and caldohexamine in cells of *Thermotoga* species, a possible adaptation to the growth at high temperatures. Arch. Microbiol. **159**(3753): 472-476.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. U. S. A. **106**(14): 5865-5870.

Zhaxybayeva, O., Swithers, K.S., Foght, J., Green, A.G., Bruce, D., Detter, C., Han, S., Teshima, H., Han, J., Woyke, T., Pitluck, S., Nolan, M., Ivanova, N., Pati, A., Land, M.L., Dlutek, M., Doolittle, W.F., Noll, K.M., and Nesbø, C.L. 2012. Genome sequence of the mesophilic Thermotogales bacterium *Mesotoga prima* MesG1.Ag.4.2 reveals the largest Thermotogales genome to date. Genome Biol. Evol. **4**(8): 700-708.

2 TRANSCRIPTIONAL RESPONSES OF *KOSMOTOGA OLEARIA* TO CULTIVATION TEMPERATURE

2.1 Introduction

Organisms on earth are able to grow at temperatures ranging at least from -20°C to 122°C (Takai et al. 2008; Tuorto et al. 2014), however individual cells have a much narrower growth range, usually spanning 20 - 40°C. *Escherichia coli* O157:H7 for example, grows from 19°C - 41°C (Raghubeer and Matches 1990), a 22°C range, while *Geobacillus thermoleovorans* grows from 35°C - 78°C (Nazina et al. 2001), a 43°C range. *Kosmotoga olearia* on the other hand has an unusually wide growth range spanning at least 54°C (determined here to be ~ 25°C – 79°C). What enables this large growth range? Since temperature affects many different cellular functions and structures (See Chapter 1), the broadest possible overview of *K. olearia*'s responses to temperature was desired in order to identify candidate genes, pathways and processes involved in its thermal adaptation. Several 'omics technologies can be used to study these responses, such as transcriptomics, proteomics and metabolomics. Indeed some of these techniques have been applied to other Thermotogae members, particularly *Thermotoga* spp. For example, a cDNA microarray-based transcriptomic approach was used to study the different preferences for carbohydrates among four *Thermotoga* spp. (Frock et al. 2012), while many genomic features of *Thermotoga maritima* were examined in a multi-omic approach consisting of genome resequencing, RNA-Seq-based transcriptomics and proteomics (Latif et al. 2013). Additionally, a proteomic study of *T. maritima*'s response to growth temperature (Wang et al.

2012) enables comparison of temperature responses between the well-studied *T. maritima* and *K. olearia*.

In order to study the changes in gene expression by *K. olearia* at different temperatures, transcriptomics was used to examine these changes at the transcriptional level. Additionally, transcript abundance has been shown to correlate with protein abundance in other Thermotogae bacteria (Latif et al. 2013), suggesting that the transcriptome likely also provides information about protein abundance in *K. olearia*. Transcriptomes are usually profiled using either hybridization-based technologies such as microarrays, or more recently, sequencing-based technologies like RNA-Sequencing (RNA-Seq) (van Vliet 2010). An RNA-Seq approach was used here due to the lack of *K. olearia*-specific microarrays and the advantages of RNA-Seq over microarray technology, such as single nucleotide resolution, better signal-to-noise ratio, and higher dynamic range (reviewed by Güell et al. (2011) and van Vliet (2010).

2.1.1 RNA-Seq

RNA-Seq procedures usually consist of the following steps: RNA extraction, selection of target RNA molecules, RNA fragmentation, cDNA synthesis, adapter ligation, library amplification, and sequencing (Febrer et al. 2011). Aside from the sequencing-related and library amplification bias issues reviewed by Febrer et al. (2011) and Oshlack et al. (2010), there are still several actively debated aspects of this new technique. Firstly, what is the required amount of sequencing for meaningful differential expression analysis? Secondly, what is the effect of rRNA depletion (the method to select target RNA molecules mentioned above) on the resultant transcriptome? Finally, how important is the complete elimination of DNA from the total RNA

preparation for the downstream analysis? The first two issues have been addressed by Haas et al. (2012) who sequenced transcriptomes of *E. coli*, *Mycobacterium tuberculosis*, and *Vibrio cholerae* to evaluate the issue of how many reads are required for transcript quantification. They found that, unsurprisingly, rRNA depletion allows for better detection of transcripts, but is not necessary if sufficient reads are present to saturate the transcriptome. Interestingly, however, based on random data reduction of their sequencing reads, they concluded that 2-3 million fragments per sample, even for biological replicates with low correlation, is sufficient for identification, with high statistical significance, of genes differentially expressed by more than 2-fold. The potential issue of co-isolating DNA with the desired RNA has been suggested to cause an increase in background signal when using strand-specific RNA-Seq in samples experiencing incomplete DNase digestion (Perkins et al. 2009), but the relative effect on differential expression analysis in non-strand-specific RNA-Seq experiments is unknown.

To examine global transcriptional responses to temperature in *K. olearia*, RNA-Seq was conducted here on RNA isolated from cultures of *K. olearia* grown at different temperatures. Based on the findings of Haas et al. (2012), rRNA depletion was performed on all samples with the goal of sequencing ~3 million reads per sample.

2.2 Materials and Methods

2.2.1 Cultivation of *Kosmotoga olearia*

Cultures of *K. olearia* were grown in KTM (*Kosmotoga* medium) containing (L^{-1} glass-distilled water): 30 g NaCl, 0.335 g KCl, 1.395 g $MgCl_2 \cdot 2H_2O$, 1.4 g $MgSO_4 \cdot 7H_2O$, 0.25 g NH_4Cl, 0.14 g $CaCl_2 \cdot 2H_2O$, 0.45 g KH_2PO_4, 10 mL trace mineral solution (Balch et al. 1979; Appendix A) adjusted to pH 7.0, 5.0 g Bacto yeast extract (BD), 0.5 g cysteine, 6.7 g PIPES, 1 mg resazurin. The pH was adjusted to 6.8 with 10 M NaOH. The medium was sparged with N_2 for 30 min with gentle heating and dispensed anaerobically under a stream of N_2. Serum bottles (Wheaton, 125 mL; Fisher Scientific) containing 50 mL KTM, or Hungate tubes (17 mL; Fisher Scientific) containing 9 mL KTM were sealed with butyl rubber stoppers and aluminum crimps and autoclaved at 121°C for 20 min. All bottles were then supplemented from sterile, anaerobic stocks with 2.5 mL 10% sodium pyruvate, 0.5 mL 1 M sodium thiosulfate, and 0.5 mL vitamin solution (Balch et al. 1979; Appendix A), and all tubes were supplemented with 0.5 mL 10% sodium pyruvate, 0.1 mL 1 M sodium thiosulfate, and 0.1 mL vitamin solution. All cultures were inoculated using a sterile syringe and needle and were incubated stationary in the dark.

Cultures used as inocula were stored at 4°C unless they were originally grown at 77°C. Actively growing cultures at 77°C had to be used directly as inoculum because if they were stored at 4°C transfer cultures would not grow and even at room temperature (~22°C) the culture would not re-grow upon transfer after two months of storage. All experiments at each temperature had the same volume of inoculum added, however, inoculum volume differed at different temperatures as the non-optimal temperatures needed larger volumes to grow. Table 2.1 shows the volumes of inoculum used for each temperature and vessel type. Due to the very slow growth rate of *K. olearia* at 25°C, cultures were first grown at 30°C then transferred into new medium as a 50% inoculum and incubated at 25°C for 24 h.

2.2.1.1 Growth curves of *K. olearia*

Growth curves were performed by using a sterile syringe and needle to place 1 mL of culture into a cuvette for measurement of optical density at OD_{600} using a spectrophotometer (Ultrospec 3100 pro). Two sets of triplicate bottles, inoculated from the same inoculum 12 h apart, were monitored to generate the growth curves at all temperatures except 30°C for which one set of triplicate bottles was monitored once daily. The ln OD_{600} was plotted against growth time for each culture and the curves were individually fitted with a linear trend line to determine the slope of the log phase (i.e. the growth rate). Full composite growth curves were constructed from all replicate culture data for each temperature to determine the length of time corresponding to each growth phase. Isothermic growth curves consisted of six replicate cultures except for the curves conducted at 30°C and 77°C for which there were three and 12 replicate cultures respectively. All shifted growth curves consisted of six replicate cultures except for the curves conducted at 40°C and 30°C for which there were four and three replicate cultures respectively.

2.2.2 RNA Isolation and Processing

All cultures used for RNA extraction were inoculated from cultures that had also grown at that temperature for at least three transfers. RNA was extracted in either mid-log phase or late-log phase as determined from the composite growth curve at that temperature. Multivariate analysis suggested that any differences between mid-log and late-log phase samples were not significant compared to the differences due to the other variables examined.

RNA for the first two transcriptomes generated, designated K65-D-IT and K65-E-IT, was isolated using the Zymo Research Fungal/Bacterial RNA MiniPrep Kit (Cedarlane Laboratories, Ltd.; Burlington, Ontario), using the manufacturer's protocols. Two culture bottles were opened and the entire contents were transferred to two 50-mL conical polypropylene centrifuge tubes (Corning 430290) for centrifugation at 6100 g for 30 min at 20°C. The RNA underwent in-column DNase digestion during RNA isolation (using Zymo Research DNase I set 250U; E1009) and was eluted in 30 µL of RNase-free water provided with the MiniPrep kit. The RNA from the two culture bottles was pooled and an aliquot was taken for library preparation resulting in the K65-D-IT transcriptome. The remaining RNA was repeatedly digested with DNase using the Zymo Research DNA-Free RNA Kit, following the manufacturer's instructions, until a 45-cycle PCR with the RNA as template did not show any amplification. This RNA was then used for library preparation resulting in the K65-E-IT transcriptome. The PCR used primers Kole_1838_rpob.944.F 5' – TCC CCA GAA TCA ACG CCG CA – 3' and Kole_1838_rpob.1174R 5' – GGG CCA AAT GTC TGG CCG CT – 3' targeting the RpoB gene in *K. olearia* (Kole_1838) and were subjected to an initial denaturation at 96°C for 5 min, followed by 45 cycles with denaturation at 96°C for 30 s, primer annealing at 52°C for 30 s, and primer extension at 72°C for 1 min, followed by a final extension at 72°C for 5 min. Two microlitres of the PCR products were run on a 1% agarose gel containing GelRed at 70 V for 40 min, with 6 µL of Fermentas GeneRuler 1kb DNA Ladder. The gel was then photographed under UV light. Based on the analysis of the K65-E-IT and K65-D-IT transcriptomes RNA used for all other transcriptomes was digested with DNase only once, during the isolation.

For all other transcriptomes, the cultures were treated by adding an equal volume of "stop solution" (10% phenol in ethanol) immediately upon removal from the incubator. The resulting

mixture was then transferred to a 250-mL polypropylene centrifuge bottle (Nalgene PPCO) and centrifuged at 4068 g for 20 min at 4°C in a Sorvall RC-5B Refrigerated Superspeed Centrifuge. The supernatant was decanted and the cell pellet was re-suspended in 1.5 mL cold 50/100 TE buffer [50 mM Tris (ICN) and 100 mM EDTA (BDH), pH 8] and transferred into two 2-mL microcentrifuge tubes (Sorenson Dolphin tubes, Sigma-Aldrich). The cells were re-pelleted at 21000 g for 5 min in an Eppendorf 5424 centrifuge. The supernatant was removed and the two tubes containing cells from the same culture were combined by re-suspending the cells from one tube in 750 µL cold 50/100 TE buffer, transferring the suspension to the other tube and re-suspending all cells. Cells were then re-pelleted in the same manner as before. Once the supernatant was removed the cell pellets were used in the Zymo Research Fungal/Bacterial RNA MiniPrep Kit to extract the total RNA following the manufacturer's instructions, performing the in-column DNase digestion once, and eluting in 30 µL of RNase-free water.

Quality and quantity of the total RNA were assessed on an Agilent 2100 bioanalyzer RNA Nano chip following the manufacturer's instructions for Prokaryote Total RNA. rRNA was then depleted using an Epicentre Ribo-Zero rRNA Removal Kit (Gram-Positive Bacteria) Magnetic Kit (MRZGP126), and the quantity and efficiency of rRNA depletion was assessed on the Agilent 2100 bioanalyzer RNA Nano chip (or RNA Pico chip), following the manufacurer's instructions. On average, two rRNA depletions were needed to generate sufficient input RNA (200-500 ng for Ion Torrent PGM and 10-400 ng for Illumina MiSeq), although some samples required as many as five rRNA depletions. Successfully depleted RNA was used to construct libraries for RNA-Seq following the manufacturer's instructions and sequenced on either an Ion Torrent PGM (RNA-Seq kit V2) or an Illumina MiSeq (TruSeq RNASeq v2 2x100 bp). The platform and RNA extraction technique used for each transcriptome is presented in Table 2.2.

2.2.3 Data Analysis

2.2.3.1 RNA-Seq analysis and calculation of expression values

Sequencing reads were analyzed using the RNA-Seq function in CLC genomics workbench version 7.0.4. Briefly, reads were first trimmed for quality and to reduce ambiguity using an ambiguous limit of 2 and a quality limit of 0.05 (maximum number of N's allowed in the read and the maximum probability of a base-call error based on the quality scores from the sequencer). They were then subjected to an initial RNA-Seq protocol with lax mapping parameters [mapping only to gene regions in the reference sequence, similarity fraction of 0.5, length fraction of 0.5, all other default settings (maximum number of hits for a read of 10, map to both strands, mismatch cost of 2, insertion cost of 3, deletion cost of 3, auto-detect paired distances, color space alignment, color error cost of 3)] and only rRNA genes as a "reference genome" to remove all rRNA reads from all samples (to remove biases from different rRNA depletion efficiency). All un-mapped reads were collected and subjected to another RNA-Seq protocol with strict mapping parameters (also mapping to inter-genic regions, similarity fraction of 0.95, and length fraction of 0.95, all other default settings) using the *K. olearia* annotated genome as a reference. Expression values were calculated for every gene as reads per kilobase per million mapped reads (RPKM) values. Reads that did not map during the analysis were discarded. In total, 16 transcriptomes were analyzed and comparisons were set up to determine fold changes between two temperatures at a time. The statistical analysis "Empirical Analysis of DGE" function in CLC (which employs the "Exact Test" for two-group comparisons developed

by Robinson and Smyth (2008)) was used to evaluate the differences in gene expression observed.

2.2.3.2 Multivariate Analysis

Principal Component Analysis (PCA) and Biplots overlaying variable vectors on top of PCAs were performed using R (R Core Team 2013) packages ade4 and bpca respectively (references available through http://cran.r-project.org/web/packages/ website). The input files used the RPKM values calculated by CLC genomics workbench. The data generated by the R packages were plotted in Microsoft Excel.

2.2.3.3 Differential expression analysis

All the transcriptomes were clustered according to similarity using hierarchical clustering in MeV (Saeed et al. 2003) using both Manhattan distance and Pearson correlation as distance metrics and average linkage. Based on this clustering comparisons between temperature groups were set up in CLC for calculation of fold changes and stats. Where applicable, combinations of transcriptomes between two temperatures included Illumina transcriptomes only, all transcriptomes, and all 65 transcriptomes except for K65-1-ILL due to its clustering with the 77°C transcriptomes. Only genes that were considered differentially expressed in all three of these comparisons for any given temperature comparison were considered to be differentially

expressed between the two temperatures. Additionally, all genes were clustered in MeV based on their expression values using the Pearson correlation.

For genes to be considered differentially expressed in any comparison they had to have (i) a statistically significant difference in expression (p-value <0.05), (ii) have greater than 20 reads (criterion used by Simon et al., 2013) in at least one Illumina transcriptome (due to reduced statistical power at low read counts; Oshlack and Wakefield 2009; Oshlack et al. 2010), and (iii) exhibit either < -2 or >2 fold change. Only genes that fulfilled all three criteria were considered to be differentially expressed between the two temperatures examined. Temperatures compared include all temperatures compared to optimal (25°C vs 65°C, 30°C vs 65°C, 40°C vs 65°C, 77°C vs 65°C), all temperatures compared to those temperatures closest to them (25°C vs 30°C, 30°C vs 40°C), and the two temperature extremes (77°C vs 25°C). Genes from all seven temperature comparisons were combined into a list of all genes showing a temperature response. These genes were sorted into the clusters of orthologous groups (COG categories) assigned to them in the integrated microbial genomes database (IMG) (Markowitz et al. 2014). Genes that were assigned to more than one category were counted in all categories assigned to them.

2.2.4 Analysis of Genes Involved in Growth on Pyruvate

The KEGG and BioCyc (Caspi et al. 2014) predicted pathways for pyruvate conversion to acetate in *K. olearia* were retrieved. Expression values for all those genes were examined together with those for the hydrogenases identified by Schut et al. (2012) and the F-type ATPase

subunits. Further, these genes were checked to see if the corresponding protein was detected in the proteomes of *K. olearia* generated by Green (2013).

2.3 Results

2.3.1 Growth of *Kosmotoga olearia*

Kosmotoga olearia is reported to grow at temperatures from 20°C - 80°C with optimal growth at 65°C (DiPippo et al. 2009). Initially, in the current study, the temperatures 20°C, 40°C, 65°C, and 80°C were desired for a transcriptomic investigation of *K. olearia*'s behaviour across this growth range. However, once growth of the isolate was achieved at 65°C, attempts to then cultivate it directly from this culture at other temperatures, particularly lower temperatures, failed. Rather, the growth temperature had to be changed in smaller increments to achieve an inoculum from which a culture could grow at the desired temperature. During this process, it was noted that the time it took for the culture to grow increased as the temperature diverged from 65°C. Additionally, cultures grown at a given temperature that had been inoculated from a culture grown at a different temperature took longer to grow than cultures that were inoculated from cultures grown at that same temperature. To quantify these observations growth curves were conducted at each of the temperature increments used to acclimatize *K. olearia* to non-optimal growth temperatures. Figures 2.1 and 2.2 show that the temperature shifts away from optimal temperature caused both a longer lag phase and a slower growth rate compared to cultures being maintained at the same temperature. Due to the growth rates observed, the

densities achieved, and the consistency of growth, the temperatures 30°C, 40°C, 65°C, 77°C and a 25°C shift from 30°C were chosen for transcriptomic study. Furthermore, other than for the 25°C shift, all cultures that were used for RNA extraction came from cultures that had also been grown at the same temperature for at least three transfers.

2.3.2 The RNA isolation method affects gene expression

The effect of DNA depletion versus DNA elimination on the RNA-Seq analysis was investigated. This was done by pooling RNA samples from two cultures grown at 65°C as described in the Materials and Methods (2.2.2) which were used to generate two transcriptomes designated K65-D-IT (DNA-depleted) and K65-E-IT (DNA-eliminated). K65-E-IT generated very few reads that mapped to the *K. olearia* genome (Table 2.2), making the comparison of the two transcriptomes difficult. Using the same criteria used for differential expression analysis (in this case differential detection), 17 genes were considered significantly differently detected (Appendix B.1), although many more were completely undetected in the K65-E-IT transcriptome. In addition to the RNA-Seq analysis, a *de novo* assembly was done from the K65-E-IT reads that did not map after the RNA-Seq analysis. BLAST searches of the resulting contigs revealed a high number of matches to *Bacillus sp.* and to *K. olearia* (data not shown). Since the reads used for this *de novo* assembly passed the quality thresholds set during trimming, it is unclear why the reads making up the contigs with BLAST matches to *K. olearia* did not map during the RNA-Seq analysis. Relaxing the mapping parameters (even to the settings used for *in silico* rRNA-read removal) did not significantly improve the proportion of reads mapping (increased to 12.47%, data not shown).

Importantly, there were regions of the genome (both intergenic and annotated) that had no reads mapping from K65-D-IT, suggesting that any DNA in the RNA sample did not persist in the final library and get sequenced. Additionally, the K65-E-IT transcriptome was of much lower quality with both low number of reads that mapped to the genome and the presence of many contaminating sequences, particularly *Bacillus* sequences. Taken together these data suggest that additional DNase digestions are not necessary and favour an approach that minimizes handling of the RNA.

Despite their differences, K65-D-IT and K65-E-IT both showed high expression of a cold shock-related gene (Kole_2064) (Fig. 2.3). Since cold-shock genes are among the candidates for differential expression between temperatures, and the high expression could have been a response to the cell harvesting conditions (rather than a property of optimal growth), an alternate RNA isolation method was sought. Therefore, the K65-SK-IT transcriptome was constructed using RNA extracted after adding a phenol-ethanol stop solution to the culture prior to harvesting the cells (as described in Materials and Methods, 2.2.2). Compared to K65-D-IT and K65-E-IT, K65-SK-IT had much lower expression of the cold shock-related gene, high expression of metabolism genes that are expected to be active (see section 2.3.3) as well as higher expression of an alcohol dehydrogenase (Kole_0742) (Fig. 2.3). This cell harvesting method was used for all subsequent transcriptomes because, even though the cells were still reacting to the RNA isolation procedure, a phenol-ethanol-related response was favoured over a temperature-related one for this temperature study.

2.3.3 Energy Generation Pathway Model

Since the 65°C transcriptomes were generated from cultures growing under the reported optimal conditions for *K. olearia* TBF 19.5.1, they can be used to examine central metabolism in this bacterium. This can provide insight into pyruvate metabolism (which is uncommon among the Thermotogae) and the unknown mechanism for growth enhancement provided by (thio)sulfate (DiPippo et al. 2009). Table 2.3 shows the average expression at each temperature of all of the genes predicted by KEGG and BioCyc to be involved in the conversion of pyruvate to acetate. Based on these values, the model presented in Figure 2.4 was constructed by selecting genes with similarly high expression. Gaseous end products of growth on either maltose or pyruvate with and without thiosulfate were measured and are shown in Table 2.4.

2.3.4 Temperature-related transcriptional responses in *K. olearia*

Once all of the transcriptomes were sequenced and their expression values calculated, multivariate analysis was conducted to cluster the transcriptomes to evaluate their similarity. Based on the results of hierarchical clustering and principal component analysis (PCA) when including K65-D-IT and K65-E-IT (data not shown) and the differences noted between them and K65-SK-IT (section 2.3.2), K65-D-IT and K65-E-IT were excluded from all further analysis.

Figure 2.5 shows the MeV hierarchical clustering of the transcriptomes using Manhattan distance. With the exception of K65-1-ILL, which clusters with the 77°C transcriptomes, all biological replicates of each temperature cluster together suggesting temperature-specific differences between test conditions. Pearson correlation gave similar results with some mixing of the 30°C and 25°C transcriptomes and K40-1-ILL clustering with these low temperature samples rather than with the other 40°C transcriptomes. The clustering pattern of the temperature

treatments in Fig. 2.5 illustrates how transcriptomes from one temperature were more similar to transcriptomes from the temperatures closest to them.

PCA plots show a similar pattern of replicates clustering together. Figure 2.6 shows a biplot where one such PCA (constructed from only the Illumina transcriptomes) has been overlaid with the underlying variable vectors. Biplots including the Ion Torrent transcriptomes (not shown) show a similar pattern, but also display some sequencing platform-specific differences. In addition to clustering of the biological replicates, a segregation between intermediate temperatures (65°C and 40°C) and extreme temperatures (77°C, 30°C and 25°C) along the vertical axis, and high temperatures (77°C and 65°C) and low temperatures (40°C, 30°C and 25°C) along the horizontal axis is evident (Fig. 2.6). This further illustrates the similarity between adjacent temperature treatments since the plotted transcriptomes increase in temperature moving from quadrant 3 to 4 in a clockwise manner in Fig. 2.6.

With these clustering patterns in mind, temperature comparisons were set up as described in Materials and Methods (section 2.2.3.3). The genes considered to be differentially expressed were collected from all seven temperature comparisons and combined together into a list of all genes responding to temperature, totalling 890 genes (Appendix C). Figure 2.7 shows the COG category breakdown of these genes compared to all of the genes in the *K. olearia* genome. These were further refined by examining only the temperature comparisons involving optimal temperature (25°C vs 65°C, 30°C vs 65°C, 40°C vs 65°C, 77°C vs 65°C) and including the trend in change of expression (Fig. 2.8, Appendix B.2).

Finally, the MeV hierarchical clustering of all the genes in *K. olearia* using Pearson correlation (from the transcriptome clustering) was used to create clusters of genes that show similar expression changes between temperature treatments. In this way genes of interest were

located that show similar patterns of expression across temperatures. This gave a total of 56 clusters of genes. Examination of the clustering of genes with well-studied temperature responses revealed them clustering together as expected: heat shock proteins cluster together and separately from cold shock proteins, which also cluster together (Appendix D). The 56 clusters are listed in Appendix D with the top 30 variable vectors shown in Fig. 2.6 being highlighted.

2.4 Discussion

2.4.1 Energy Generation Pathway

K. olearia was reported to grow optimally using pyruvate (DiPippo et al. 2009), a substrate on which only a few other Thermotogae species have been reported to grow well (e.g. *Fervidobacterium* spp. (DiPippo et al. 2009)). Interestingly, two other *Kosmotoga* species (*K. pacifica* and *K. arenicorallina*) were reported not to grow on pyruvate (L'Haridon et al. 2014; Nunoura et al. 2010). To provide insight into the mechanism by which *K. olearia* is able to use this substrate the 65°C transcriptomes were used to examine gene expression during optimal growth, in addition to the temperature comparison. Genes used by *K. olearia* to metabolise pyruvate are expected to be highly expressed under optimal growth conditions, thereby allowing the transcriptomes to be used to refine metabolic pathways predicted from genomic information, based on the transcription level of each of the genes predicted to be involved.

K. olearia was reported to produce acetate, hydrogen, and carbon dioxide as the major end products of growth on maltose (DiPippo et al. 2009). Examination of its genome (Swithers et

al. 2011) revealed that the Embden-Meyerhof-Parnas (EMP) pathway and most of the pentose phosphate pathway are present, indicating that maltose metabolism likely occurs through one or both of these pathways. Since the EMP and pentose phosphate pathways produce pyruvate, growth on pyruvate was expected to yield the same end products. Therefore, predicted pathways for the conversion of pyruvate to acetate by *K. olearia* were sought. The model presented in Fig. 2.4 accounts for all of the major end products expected during growth at 65°C and contains genes with consistently high expression across all temperature treatments (Table 2.3). Furthermore, with the exception of Kole_0222 and Kole_1230, the protein products of all the genes in Fig. 2.4 were detected in the proteomes generated by Green (2013). Moreover, the enzyme pyruvate ferredoxin oxidoreductase was also found to be the enzyme responsible for conversion between pyruvate and acetyl-CoA in *Thermotoga neapolitana* during both sugar fermentation and conversion of exogenous acetate and carbon dioxide to lactic acid (D'Ippolito et al. 2014).

When the gaseous end products from growth on either maltose or pyruvate were measured (Table 2.4), carbon dioxide and hydrogen were indeed found in both, as predicted. Furthermore, the differences in the ratio of carbon dioxide to hydrogen produced by growth on maltose or pyruvate supports the model presented. During growth on maltose reduced electron carriers would be generated from the conversions of maltose to pyruvate. These would ultimately result in hydrogen production in addition to the hydrogen and carbon dioxide produced during pyruvate conversion to acetate. Therefore, growth on maltose is expected to yield a higher proportion of hydrogen to carbon dioxide, which was indeed observed (Table 2.4). Interestingly, the presence of thiosulfate in the medium did not significantly change the level of hydrogen in the headspace, suggesting that the mechanism by which (thio)sulfate improves growth of *K.*

olearia is not by removing this waste product. This examination of core function during optimal growth of *K. olearia* provides a starting point for further investigation of hydrogen production and sugar metabolism in this bacterium.

2.4.2 Transcriptional Responses to Temperature

The transcriptional responses of *K. olearia* to temperature were examined using RNA-Seq in order to gain insight into this organism's wide temperature range. The results of the transcriptome clustering show a difference between intermediate temperature and extreme temperature treatments, as well as between high temperature and low temperature treatments (Fig. 2.6). In addition, the gene variable vectors point to each quadrant (corresponding to each of the temperatures at which cells were grown) in an expected way based on the gene annotation. For example, the chaperone groEL and protease Kole_1599 point toward the 77°C transcriptomes, where their high expression is anticipated due to the known involvement of these proteins in the heat shock response of the related bacterium *Thermotoga maritima* (Pysz et al. 2004). The cold shock genes Kole_0109 and Kole_2064 point toward the 30°C and 25°C transcriptomes where they are also expected to have the increased expression that is observed. These data show that the transcriptomes are indeed representative of *K. olearia*'s growth at their respective temperatures, rather than being artifacts of the procedure. Therefore, the remaining variable vectors in Fig. 2.6, as well as the genes considered differentially expressed likely represent temperature-specific differences between the treatments. One possible exception is the alcohol dehydrogenase (Kole_0742) that was noted to increase in expression due to the RNA isolation method that included a phenol-ethanol stop solution (Section 2.3.2). In the K65-D-IT

and K65-E-IT transcriptomes (which are not part of the temperature analysis as their RNA were isolated using a different method) this gene was highly expressed (Fig. 2.3), implicating it as part of the production of the trace amounts of ethanol noted as end products during optimal growth (DiPippo et al. 2009). This gene is one of the top 30 variable vectors displayed in Fig 2.6 and points toward the 65°C transcriptomes. While this gene does not likely play a role in temperature response in *K. olearia*, the changes in its expression values between temperature treatments could instead be indicative of the overall level of cellular activity at each temperature. Expression of this gene is highest at 65°C, the optimal growth temperature of *K. olearia*, which is the temperature at which this organism would be most active and able to respond to changing conditions, such as the sudden increase in alcohol concentration, before cell death occurred due to the stop solution addition. The lower expression values at the other temperatures suggest that the cells were less able to respond to the alcohol addition before the concentration reached lethal levels. This could be due to lower metabolic activity, or perhaps a reduced need for this gene's protein product at other temperatures, resulting in a need for induction instead of up-regulation during the seconds it takes to inject the killing solution into the culture bottle. The metabolism gene glyceraldehyde-3-phosphate dehydrogenase (Kole_2020) pointing to the 65°C transcriptomes also suggests highest metabolic activity occurs at 65°C, as expected.

The genes ssrA, ffs and rnpB pointing towards the 77°C transcriptomes (Fig. 2.6) suggests there could be an increased rate of RNA turnover and protein export at supra-optimal temperature. However, the methods used to isolate and process the total RNA select against small RNA (<200 nucleotides) recovery. The ffs transcript is annotated as 115 nucleotides in length and may therefore not be fully represented in every transcriptome.

Ten of the top 30 variable vectors in Fig. 2.6 come from genes encoding ribosomal proteins and all of these point towards the 30°C and 25°C transcriptomes. The up-regulation of ribosomal proteins in response to cold stress has also been reported for other prokaryotes (Graumann and Marahiel 1999; Kaan et al. 2002; Watanabe et al. 2012; Williams et al. 2011). In particular, the genes Kole_1840 encoding 50S ribosomal protein L10 and Kole_1839 (rplL) encoding ribosomal protein L7/L12 showed dramatic increases in expression with decreasing temperature (Table 2.5). These have both been reported to increase in the *Bacillus subtilis* cold response (3 h after cold-shock) and were suggested to be part of its cold acclimatization mechanism (Graumann and Marahiel 1999). Finding strongly increased transcription of these genes under sustained low-temperature growth of the phylogenetically distant *K. olearia* suggests that the use of these proteins in the ribosome may be a general bacterial response to sub-optimal temperatures. Furthermore, the idea that ribosomal components need to be switched for growth at different temperatures may explain the increased lag phase length during temperature shifts (Fig. 2.1).

The transcriptome clustering (Fig. 2.5 and 2.6) also suggests that 65°C growth is most similar to 40°C growth. Indeed differential gene expression analysis found that 40°C compared to 65°C transcriptomes showed the lowest counts of genes with different expression (Appendix B.2 B). Furthermore, cultures grown under these two treatments reached similar maximum densities (Fig. 2.1). Yet the 40°C growth rate was only one-third of the 65°C growth rate (Fig 2.2), which raises the question: given the differences observed in growth kinetics, what is the basis for the transcriptional similarity between 65°C and 40°C? Examination of the 51 genes considered differentially expressed between the 65°C and 40°C treatments shows a decrease in expression of two transcriptional regulators (Kole_0685 and 1307) and sigma 70 (Kole_0343

rpoD subfamily, the housekeeping sigma factor) at 40°C. Additionally, decreased expression of chaperones dnaK (Kole_0886) and Hsp20 (Kole_0885) as well as protease Do (Kole_1599) at 40°C is observed (Table 2.5). Taken together these data could suggest an overall decrease in cellular activity at this lower temperature. However, the lack of a change in expression of the genes involved in energy generation (Table 2.3), and the increased expression of genes involved in lipid metabolism and amino acid metabolism and transport (Appendix B.2 A) point to an active temperature response rather than solely a passive consequence of decreased chemical reaction rates associated with decreased temperature. The increased transcription of genes like fabZ (Kole_0968), patatin (Kole_1960), and 3-oxoacyl-ACP synthase III (Kole_0969) points to the synthesis of fatty acids, perhaps to maintain or increase membrane fluidity (Table 2.5). Additionally, there are many genes annotated as hypothetical proteins that show increased expression at 40°C, confounding full elucidation of the differences between 65°C and 40°C growth (Appendix C). Despite these differences, among the temperatures examined, 65°C and 40°C represent an intermediate temperature range for *K. olearia*, suggesting 40°C may lie within a narrower, core growth range.

In contrast, near the upper temperature limit of its growth *K. olearia* is quite different transcriptionally. While having only slightly more genes with different expression than the 40°C to 65°C comparison, the supra-optimal comparison (77°C to 65°C) is the only one that shows more genes with decreased expression than increased (Appendix B.2 B). Notably, there are many genes annotated as hypothetical proteins that show changes in expression together with various transporters and metabolic enzymes (Appendix C). Strikingly, most of the genes that could be assigned to a functional category fall into either carbohydrate metabolism and transport or energy production and conversion. For carbohydrate metabolism and transport however, more of

the genes show increased expression than decreased expression at 77°C, which only occurs in this functional category and in the genes that couldn't be assigned to a category (Fig. 2.8, Appendix B.2). This finding is consistent with the findings of Wang et al. (2012) who also reported up-regulation of carbohydrate metabolism during higher temperature growth in *T. maritima*. Another study conducted on the immediate heat-shock response in *T. maritima* found induction of chaperones like dnaK, and groEL/groES, as well as the sigma factors rpoE and rpoD and the regulator hrcA (Pysz et al. 2004). With the current analysis however, none of these genes were found to have significantly different transcription at 77°C compared to 65°C. Interestingly, when removing the 65°C transcriptome (K65-1-ILL) that clustered with the 77°C transcriptomes from the analysis, the extreme heat stress sigma factor-24 (rpoE) (Kole_2150) and the positive regulator of rpoE ResC/MucC (Kole_1855) are significantly induced at 77°C. High expression of these two genes in the K65-1-ILL transcriptome prevents them from appearing as differentially expressed between 65°C and 77°C using the method outlined in section 2.2.3.3, and also suggests that the 65°C culture used to generate the K65-1-ILL transcriptome encountered stress before the cells were killed. However since many of the genes showing the largest increase in expression in K65-1-ILL compared to the other 65°C transcriptomes are hypothetical proteins, the source of the stress is cryptic. Even with the removal of K65-1-ILL, dnaK, groEL/ES, rpoD and hrcA do not show significantly different expression between the 65°C and 77°C groups. This likely reflects the differences between an immediate heat-shock response and a longer term supra-optimal temperature conditioned state. The distinction of the 77°C transcriptomes both in multivariate analyses and based on transcriptional differences suggests that this temperature represents a heat-stressed state.

Finally, the low temperature transcriptomes show the greatest number of genes with different expression compared to optimal. Due to the different method used to generate the 25°C transcriptomes, they should be interpreted with care because differences between these transcriptomes and the others could be a result of either the growth temperature or the fresh medium. For this reason, and due to their similarity illustrated in multivariate analyses (Fig. 2.5 and 2.6), the 25°C and 30°C transcriptomes, which both represent a cold-stressed state, are interpreted together here. Contrary to the 77°C transcriptomes, most genes in the Carbohydrate Metabolism and Transport COG category show decreased expression at the low end of *K. olearia*'s temperature range (Fig. 2.8). In particular, many genes in the gluconeogenic pathway show lower expression at these temperatures (Fig. 2.9), perhaps reflecting a decrease in peptidoglycan synthesis or a general decrease in demand for the intermediates of the pathway. Interestingly, many genes involved in amino acid metabolism and transport show increased expression at these temperatures (Fig. 2.8). Among these are all the genes that make up the predicted pathway from glutamate to ornithine (argC, argJ, Kole_0096 and 0097). Downstream metabolites of ornithine, such as spermidine, are known to accumulate as compatible solutes in some organisms (Higashibata et al. 2000). However, the gene predicted to encode the enzyme that would convert ornithine to putrescine (the first step in spermidine synthesis from ornithine) in *K. olearia* TBF 19.5.1 (Kole_1584) does not show a temperature response. Instead the increased demand on the pathway could reflect a greater need for these amino acids for proteins being synthesized, or perhaps a switch in metabolism to peptide catabolism. Remarkably, the gene predicted to convert acetyl-CoA to citrate (Kole_1230) also shows a dramatic increase in expression in both the 25°C and 30°C transcriptomes (Table 2.3), possibly causing a re-direction of acetyl-CoA utilization at low temperature. Additionally, a citrate transporter (Kole_1598)

shows a three-fold increase in expression (Table 2.5), while many of the subunits making up the F-type ATPase show decreased expression at low temperature. These data raise the possibility that central metabolism is changed at low temperature. Based on the pathway presented in Fig. 2.4, a re-direction of acetyl-CoA would change the amount of acetate produced as an end-product during growth on pyruvate, while leaving the amounts of carbon dioxide and reduced electron carriers the same. An analysis of both the head-space composition and water-soluble metabolites in the medium between cultures grown at 65°C and 30°C would confirm or refute this possibility. In addition to the metabolic changes noted, there is also a large increase in the expression of a set of genes located adjacent to each other annotated as a histidine kinase and two response regulators (Kole_1015 – 1017). This potential two-component regulatory system shows no significant change in expression at 77°C and a progressive increase in expression as temperatures decrease from 65°C (Table 2.5) and therefore could represent a cold-adaptation regulatory system. Wang et al. (2012) also found a sensor histidine kinase as well as several transcriptional regulators to be more abundant during growth of *T. maritima* at the low end of its temperature range. Interestingly, many of the types of genes observed to increase in expression in *K. olearia* at low temperature, such as ribosomal proteins, cold-shock proteins, and DEAD/DEAH box RNA helicases (Kole_0922) also increased in response to low temperature growth in the psychrotolerant mesophile *Sulfuricella denitrificans* (Watanabe et al. 2012). This could be an indication that *K. olearia* behaves more like a mesophile in the mesophilic part of its temperature range, or it could reflect a general microbial reaction to low temperature, although Wang et al. (2012) did not detect an increase in RNA helicases or cold-shock proteins at low temperature in their proteomic study of the thermophilic *T. maritima*.

While transcriptomic data provide information about one level of gene expression, additional proteomic and metabolomic data provide a more comprehensive understanding of *K. olearia*'s temperature response. Three proteomes were generated from *K. olearia* growing at 40°C, 50°C, and 65°C by Green (2013), though the author cautions that the quantification of relative protein abundances in the proteomes should not be considered reliable because they were unable to generate multiple biological replicates for proteomic sequencing. However, even without protein quantification these proteomes are very useful for confirming the presence of a protein product, which can support conclusions made from transcriptomic data, particularly for highly expressed genes. Indeed all but two of the genes presented in Fig. 2.4 were detected in at least one of the proteomes, in addition to showing high transcript abundance (Table 2.3), supporting their role in central metabolism in *K. olearia*.

If the protein abundances are taken into account, particularly from the 65°C proteome which shows the greatest sequencing depth, these data can be used to evaluate the postulation that transcriptomic data may indicate protein abundance in *K. olearia*, based on the finding that this was the case for *T. maritima* (Latif et al. 2013). The average expression values for the 65°C transcriptomes were plotted against the peptide count for all proteins detected in the 65°C proteome (523 proteins) (Fig. 2.10). A linear line of best fit had a Pearson correlation coefficient (r) of 0.61 which is similar to the value obtained for *T. maritima* (0.63; Latif et al., 2013). This suggests that there is also a linear relationship between transcription and translation in *K. olearia*, although the relationship is based on only 24% of the annotated protein coding genes in *K. olearia*, compared to 74% of the annotated proteins in *T. maritima* (Latif et al. 2013).

By contrasting the heat-stressed and cold-stressed states, it is clear that *K. olearia* does not have a single strategy for dealing with all temperatures in its growth-permissive range.

Rather, this organism has different ways of dealing with the extremes of its temperature range. These responses involve transcriptional changes in large numbers of genes involved in carbohydrate metabolism and transport, amino acid metabolism and transport, energy production, replication and repair, transcription, and inorganic ion transport and metabolism (Fig. 2.7), but these processes respond differently at either sub-optimal temperatures or supra-optimal temperatures. An overview of the cold-stressed, heat-stressed, and intermediate states compared to optimal temperature is given in Fig. 2.11.

2.5 Significance

Due to its wide temperature growth range *K. olearia* presents an excellent opportunity to study adaptations to temperature. Identifying the genes and processes that allow this organism to tolerate such a wide range of temperatures, which span both mesophilic and thermophilic ranges, provides a starting place for contrasting the temperature lifestyles of the *Kosmotoga* and *Mesotoga* lineages. By studying both lineages, inferences can be made about temperature adaptation in the Thermotogae, the nature of thermophiles and mesophiles, and the temperature lifestyle transition that has occurred in *Mesotoga*. The results presented here suggest that temperature affects global metabolism and may allow *K. olearia* to play distinct roles in thermally different environments. Additionally, the data allow for further examination of genomic predictions. For example, some of the hydrogenase complexes predicted by Schut et al. (2012) are highly expressed across all temperatures, implicating them as the active hydrogenases responsible for *K. olearia*'s hydrogen production under the conditions examined. Further study

of these two monomeric hydrogenases (Kole_0172 and Kole_1794) could also aid in the biotechnological exploitation of this organism or its hydrogenases for biohydrogen production across a range of temperatures.

Table 2.1. Inoculum volume used for different culture types.

Temperature (°C)	Culture Vessel	Volume of Medium in the Vessel (mL)	Volume of inoculum
65	Serum Bottle (125 mL)	50	0.5 mL
40	Serum Bottle (125 mL)	50	1 mL
77, 30	Serum Bottle (125 mL)	50	5 mL
25	Serum Bottle (125 mL)	25	25 mL of 30°C culture added to 25 mL KTM
65	Hungate Tube (17 mL)	9	0.1 mL
40	Hungate Tube (17 mL)	9	0.5 mL
Above 65	Hungate Tube (17 mL)	9	1 mL
Below 40	Hungate Tube (17 mL)	9	1 mL

Table 2.2. Transcriptome metadata. The cultivation temperature, sequencing platform, RNA extraction method, growth phase, and library size are shown for each transcriptome.

Transcriptome	Cultivation Temperature	Cell Harvest Method (followed by Total RNA Extraction using Zymo Kit)	Growth Phase in which Culture was Harvested (Growth Time)	Next-Generation Platform used for Sequencing	Reads Counted/ Total Reads in Final RNA-Seq (%)
K65-D-IT	65°C	Centrifugation	Late-log (24 h)	Ion Torrent PGM	1,011,260/1,409,539 (71%)
K65-E-IT	65°C	Centrifugation	Late-log (24 h)	Ion Torrent PGM	63,613/1,354,861 (4.7%)
K65-SK-IT	65°C	Stop Solution	Late-log (24 h)	Ion Torrent PGM	397,261/958,546 (41.44%)*
K65-ML-IT	65°C	Stop Solution	Mid-log (19 h)	Ion Torrent PGM	2,544,254/2,860,207 (88.95%)*
K40-SK-IT	40°C	Stop Solution	Late-log (50 h)	Ion Torrent PGM	2,049,204/2,396,962 (85.49%)*
K40-ML-IT	40°C	Stop Solution	Mid-log (34 h)	Ion Torrent PGM	1,612,179/1,888,061 (85.39%)*
K25-1-ILL	30°C then 25°C for 24 h	Stop Solution	Mid-log (405 h) then shifted (24 h)	Illumina MiSeq	3,623,271/4,093,530 (88.51%)
K25-2-ILL	30°C then 25°C for 24 h	Stop Solution	Mid-log (405 h) then shifted (24 h)	Illumina MiSeq	3,763,414/4,322,561 (87.06%)
K30-1-ILL	30°C	Stop Solution	Mid-log (405 h)	Illumina MiSeq	3,368,812/3,883,772 (86.74%)
K30-2-ILL	30°C	Stop Solution	Mid-log (405 h)	Illumina MiSeq	3,473,575/4,059,542 (85.57%)
K40-1-ILL	40°C	Stop Solution	Mid-log (34 h)	Illumina MiSeq	3,848,433/4,309,601 (89.30%)
K40-2-ML-ILL	40°C	Stop Solution	Mid-log (34 h)	Illumina MiSeq	2,465,467/2,713,275 (90.87%)
K65-1-ILL	65°C	Stop Solution	Mid-log (19 h)	Illumina MiSeq	3,937,942/4,557,881 (86.40%)

Transcriptome	Cultivation Temperature	Cell Harvest Method (followed by Total RNA Extraction using Zymo Kit)	Growth Phase in which Culture was Harvested (Growth Time)	Next-Generation Platform used for Sequencing	Reads Counted/ Total Reads in Final RNA-Seq (%)
K65-2-ML-ILL	65°C	Stop Solution	Mid-log (19 h)	Illumina MiSeq	4,158,449/4,797,788 (86.67%)
K77-1-ILL	77°C	Stop Solution	Late-log (25 h)	Illumina MiSeq	3,610,760/4,101,290 (88.04%)
K77-2-ILL	77°C	Stop Solution	Late-log (25 h)	Illumina MiSeq	2,924,852/3,188,603 (91.73%)

*Platform-specific differences seen between Ion Torrent and Illumina transcriptomes (particularly at low RPKM values) could be due to the lower sequencing depth of Ion Torrent transcriptomes

Table 2.3. Fifty-one genes predicted to be involved in energy production during growth on pyruvate (See section 2.2.4). The average expression values (RPKM) at each temperature are shown for every gene. The genes appearing in Fig. 2.4 (selected based on expression level) are written in bold and italics. Genes written in italics show high expression at certain temperatures which may reflect temperature-dependent changes in the pathway.

Enzymatic Step	Locus Tag	Description	Average Expression at 25°C	Average Expression at 30°C	Average Expression at 40°C	Average Expression at 65°C	Average Expression at 77°C	Detected in the Proteomes[1]
Pyruvate to Acetyl-CoA	***Kole_0379***	***pyruvate/ketoisovalerate oxidoreductase, gamma subunit CDS***	***4661***	***4126***	***4304***	***4406***	***3210***	***Yes***
Pyruvate to Acetyl-CoA	***Kole_0380***	***pyruvate ferredoxin/flavodoxin oxidoreductase, delta subunit CDS***	***3931***	***3515***	***3360***	***3734***	***2913***	***Yes***
Pyruvate to Acetyl-CoA	***Kole_0381***	***pyruvate flavodoxin/ferredoxin oxidoreductase domain protein CDS***	***3355***	***2908***	***5206***	***5776***	***3085***	***Yes***
Pyruvate to Acetyl-CoA	***Kole_0382***	***thiamine pyrophosphate protein domain protein TPP-binding CDS***	***2911***	***2563***	***3904***	***4400***	***2899***	***Yes***
Acetyl-CoA to Acetyl-P	***Kole_0239***	***Propanediol utilization protein CDS***	***802***	***849***	***682***	***635***	***843***	***Yes***
Acetyl-P to Acetate	***Kole_1651***	***acetate kinase CDS***	***842***	***994***	***1925***	***1852***	***1336***	***Yes***
Predicted Hydrogenase*	***Kole_0172***	***hydrogenase, Fe-only CDS***	***1798***	***1891***	***2096***	***1895***	***2632***	***Yes***
Predicted Hydrogenase*	***Kole_1794***	***Ferredoxin hydrogenase CDS***	***424***	***481***	***1170***	***1246***	***488***	***Yes***
F0-type ATPase	***Kole_0218***	***F0F1 ATP synthase subunit beta CDS***	***679***	***635***	***2904***	***2101***	***876***	***Yes***
F0-type ATPase	***Kole_0219***	***ATP synthase F1, gamma subunit CDS***	***916***	***769***	***2649***	***1860***	***897***	***Yes***
F0-type ATPase	***Kole_0220***	***F0F1 ATP synthase subunit alpha CDS***	***974***	***729***	***2851***	***2124***	***1100***	***Yes***
F0-type ATPase	***Kole_0221***	***ATP synthase F1, delta subunit CDS***	***1365***	***1143***	***2829***	***2090***	***1555***	***Yes***

F0-type ATPase	***Kole_0222***	***ATP synthase F0, B subunit CDS***	***1166***	***957***	***3171***	***2271***	***1262***	***No***
F0-type ATPase	***Kole_0223***	***ATP synthase F0, C subunit CDS***	***992***	***775***	***1631***	***1484***	***1236***	***Yes***
F0-type ATPase	*Kole_0217*	*ATP synthase F1, epsilon subunit CDS*	*328*	*373*	*1022*	*779*	*352*	*Yes*
Acetyl-CoA to Citrate	*Kole_1230*	*2-methylcitrate synthase/citrate synthase II CDS*	*833*	*910*	*453*	*256*	*212*	*No*
Acetyl-CoA to Acetate	Kole_1124	Radical SAM domain protein CDS	67	109	70	50	148	No
Pyruvate to Acetyl-CoA	Kole_0319	pyruvate ferredoxin/flavodoxin oxidoreductase CDS	127	180	285	246	327	Yes
Pyruvate to Acetyl-CoA	Kole_2136	pyruvate ferredoxin/flavodoxin oxidoreductase, beta subunit CDS	134	176	204	95	231	No
Pyruvate to Acetyl-CoA	Kole_2158	pyruvate/ketoisovalerate oxidoreductase, gamma subunit CDS	360	252	167	220	418	Yes
Pyruvate to Acetyl-CoA	Kole_1390	dihydrolipoamide dehydrogenase CDS	185	267	427	332	312	Yes
Acetyl-P to Acetate	Kole_0583	acylphosphatase CDS	142	154	111	91	129	No
F0-type ATPase	Kole_0224	F0F1 ATP synthase subunit A CDS	281	303	426	434	408	No
Predicted Hydrogenase*	Kole_0560	hydrogenase maturation protease CDS	40	84	25	30	76	No
Predicted Hydrogenase*	Kole_0561	4Fe-4S ferredoxin iron-sulfur binding domain protein CDS	45	66	24	31	100	No
Predicted Hydrogenase*	Kole_0562	monovalent cation/proton antiporter, MnhG/PhaG subunit CDS	53	132	30	41	153	No
Predicted Hydrogenase*	Kole_0563	multiple resistance and pH regulation protein F CDS	37	93	23	29	90	No
Predicted Hydrogenase*	Kole_0564	cation antiporter CDS	34	104	23	36	87	No

Predicted Hydrogenase*	Kole_0565	NADH/Ubiquinone/plastoquinone (complex I) CDS	37	83	24	29	103	No
Predicted Hydrogenase*	Kole_0566	NADH-ubiquinone oxidoreductase chain 4L CDS	38	93	23	27	91	No
Predicted Hydrogenase*	Kole_0567	pseudogene	36	89	26	30	86	No
Predicted Hydrogenase*	Kole_0568	NADH-ubiquinone oxidoreductase chain 49kDa CDS	65	141	38	47	138	No
Predicted Hydrogenase*	Kole_0569	NADH dehydrogenase (ubiquinone) 30 kDa subunit CDS	50	95	32	31	89	No
Predicted Hydrogenase*	Kole_0570	NADH ubiquinone oxidoreductase 20 kDa subunit CDS	89	204	54	61	153	No
Predicted Hydrogenase*	Kole_0571	respiratory-chain NADH dehydrogenase subunit 1 CDS	68	123	36	46	134	No
Predicted Hydrogenase*	Kole_0572	4Fe-4S ferredoxin iron-sulfur binding domain protein CDS	33	53	22	19	58	No
Predicted Hydrogenase*	Kole_0573	oxidoreductase FAD/NAD(P)-binding domain protein CDS	52	104	27	37	122	No
Predicted Hydrogenase*	Kole_0574	4Fe-4S ferredoxin iron-sulfur binding domain protein CDS	48	84	27	27	95	No
Predicted Hydrogenase*	Kole_2051	oxidoreductase CDS	562	128	73	99	152	No
Predicted Hydrogenase*	Kole_2052	NADH dehydrogenase subunit D CDS	602	100	64	91	142	No
Predicted Hydrogenase*	Kole_2053	NADH dehydrogenase (ubiquinone) 30 kDa subunit CDS	609	111	61	87	147	No
Predicted Hydrogenase*	Kole_2054	NADH dehydrogenase subunit B CDS	555	94	47	61	115	No
Predicted Hydrogenase*	Kole_2055	respiratory-chain NADH dehydrogenase subunit 1 CDS	267	114	51	65	126	No
Predicted Hydrogenase*	Kole_2056	NADH dehydrogenase subunit M CDS	397	81	44	54	96	No
Predicted Hydrogenase*	Kole_2057	NADH/Ubiquinone/plastoquinone (complex I) CDS	502	86	56	68	104	No

Predicted Hydrogenase*	Kole_2058	NADH-ubiquinone oxidoreductase chain 4L CDS	404	73	33	31	68	No
Predicted Hydrogenase*	Kole_2059	monovalent cation/H+ antiporter subunit B CDS	698	102	53	68	133	No
Predicted Hydrogenase*	Kole_2060	monovalent cation/H+ antiporter subunit B CDS	597	81	33	45	69	No
Predicted Hydrogenase*	Kole_2061	monovalent cation/H+ antiporter subunit G CDS	782	119	54	61	140	No
Predicted Hydrogenase*	Kole_2062	monovalent cation/H+ antiporter subunit F CDS	856	209	76	77	147	No
Predicted Hydrogenase*	Kole_2063	cation antiporter CDS	962	150	72	74	161	No

1 Proteomes generated by (REF).

*Predicted by (Schut et al. 2012)

Table 2.4. Headspace composition of *K. olearia* cultures grown at 65°C on different substrates. Gaseous metabolites produced during growth are highlighted in bold and italics. The conditions used for the transcriptome generation are highlighted by the solid box. Measurements provided by Dr. Karen Budwill at Alberta Innovates – Technology Futures, Edmonton, Alberta.

			Normalised Composition (mole %)				
		CO_2	**He**	**H_2**	**O_2**	**N_2**	**Sum**
Kosmotoga olearia	KTM +vitamins +pyruvate	***19.63***	0.01	***29.73***	0.01	50.62	100
Kosmotoga olearia	KTM +vitamins +maltose	***9.96***	0.01	***27.57***	0.01	62.46	100
Kosmotoga olearia	KTM +vitamins +thiosulfate +pyruvate	***18.90***	0.01	***28.91***	0.02	52.17	100
Kosmotoga olearia	KTM +vitamins +thiosulfate +maltose	***7.74***	0.01	***24.23***	0.00	68.02	100

Table 2.5. Average expression values (RPKM) at each temperature of genes discussed in section 2.4.2.

Locus Tag	Description	Average Expression at 25°C	Average Expression at 30°C	Average Expression at 40°C	Average Expression at 65°C	Average Expression at 77°C
ssrA	transfer-messenger RNA; tmRNA	8674	6262	8011	18326	57798
ffs	4.5S RNA component of the signal recognition particle (SRP)	24575	13526	4866	11303	56080
Kole_0742	iron-containing alcohol dehydrogenase CDS	775	1288	23107	39155	17706
Kole_1501	hypothetical protein CDS	14912	15122	18338	23110	31664
rnpB	catalytic subunit of RNase P, RNA component	1761	1174	1588	3081	15654
Kole_0109	cold-shock DNA-binding domain protein CDS	12369	5602	521	326	119
groEL	groEL - chaperonin CDS	8800	4612	4352	11700	14251
Kole_1840	50S ribosomal protein L10 CDS	8460	6079	2108	2663	1160
rplL	rplL - 50S ribosomal subunit protein L7/L12 CDS	8835	6068	3365	3800	2246
Kole_0972	acyl carrier protein CDS	1766	6531	8823	4129	4241
Kole_2064	cold-shock DNA-binding domain protein CDS	7022	4513	1814	1659	283
Kole_0258	ribosomal protein L34 CDS	7326	4070	1960	801	716
Kole_1599	protease Do CDS	1716	1061	557	1380	5929
Kole_0303	extracellular solute-binding protein family 1 CDS	1063	1250	2793	4338	3247
Kole_1907	30S ribosomal protein S12 CDS	7904	7715	3659	3349	3651
Kole_1511	extracellular solute-binding protein family 5 CDS	10880	10302	13606	11991	13183
Kole_1322	hypothetical protein CDS	1088	1057	3461	3421	2587
Kole_1906	30S ribosomal protein S7 CDS	7726	7761	4384	4155	4175
Kole_2020	glyceraldehyde-3-phosphate dehydrogenase, type I CDS	2546	2919	4771	6206	4595
Kole_1500	S-layer domain protein CDS	14229	14442	17432	13576	16535

Kole_1876	30S ribosomal protein S11 CDS	5293	4410	2456	2415	1820
Kole_1394	extracellular solute-binding protein family 1 CDS	1236	1825	6118	3765	2540
Kole_1431	hypothetical protein CDS	1389	442	439	618	3705
Kole_0973	3-oxoacyl-ACP synthase II CDS	1570	4815	9313	5266	4498
Kole_1554	basic membrane lipoprotein CDS	1993	2230	3776	4433	3602
rpsD	rpsD - 30S ribosomal subunit protein S4 CDS	5022	4091	2654	2726	2043
Kole_1707	translation elongation factor Ts CDS	3552	2587	2450	1168	866
Kole_1873	ribosomal protein L17 CDS	4067	3696	2151	1872	1471
Kole_1902	50S ribosomal protein L3 CDS	5823	5473	3576	4446	4171
rpmA	rpmA - 50S ribosomal subunit protein L27 CDS	4041	3390	1411	1556	1334
Kole_0685	transcriptional regulator CDS	28	57	33	85	89
Kole_1307	XRE family transcriptional regulator CDS	114	123	50	146	205
Kole_0343	RNA polymerase, sigma 70 subunit, RpoD subfamily CDS	325	325	123	331	389
Kole_0885	heat shock protein Hsp20 CDS	977	484	337	2255	1869
dnaK	dnaK - Chaperone protein (Hsp70) CDS	1009	541	331	1631	1275
Kole_1960	Patatin CDS	624	583	411	167	202
fabZ	fabZ - 3R-hydroxymyristoyl acyl carrier protein (ACP) dehydratase (fatty acid biosynthesis) CDS	318	397	983	244	344
Kole_0969	3-oxoacyl-ACP synthase III CDS	892	2386	2644	1023	1424
Kole_2150	RNA polymerase, sigma-24 subunit, ECF subfamily CDS	544	480	179	433	776
Kole_1855	positive regulator of sigma E, RseC/MucC CDS	150	197	71	148	247
argC	argC - N- N-acetyl-gamma-glutamyl-phosphate reductase CDS	492	382	183	95	136
argJ	argJ - N-acetylglutamate synthase CDS	483	426	200	110	116
Kole_0096	acetylglutamate kinase CDS	300	267	136	81	85
Kole_0097	acetylornithine aminotransferase CDS	257	239	139	93	96

Kole_1598	Citrate transporter CDS	228	207	88	64	101
Kole_1015	response regulator receiver modulated metal dependent phosphohydrolase CDS	1279	1289	421	146	130
Kole_1016	response regulator receiver protein CDS	894	783	249	111	72
Kole_1017	histidine kinase CDS	693	697	276	91	90
Kole_0922	DEAD/DEAH box helicase domain protein CDS	1135	755	224	106	102

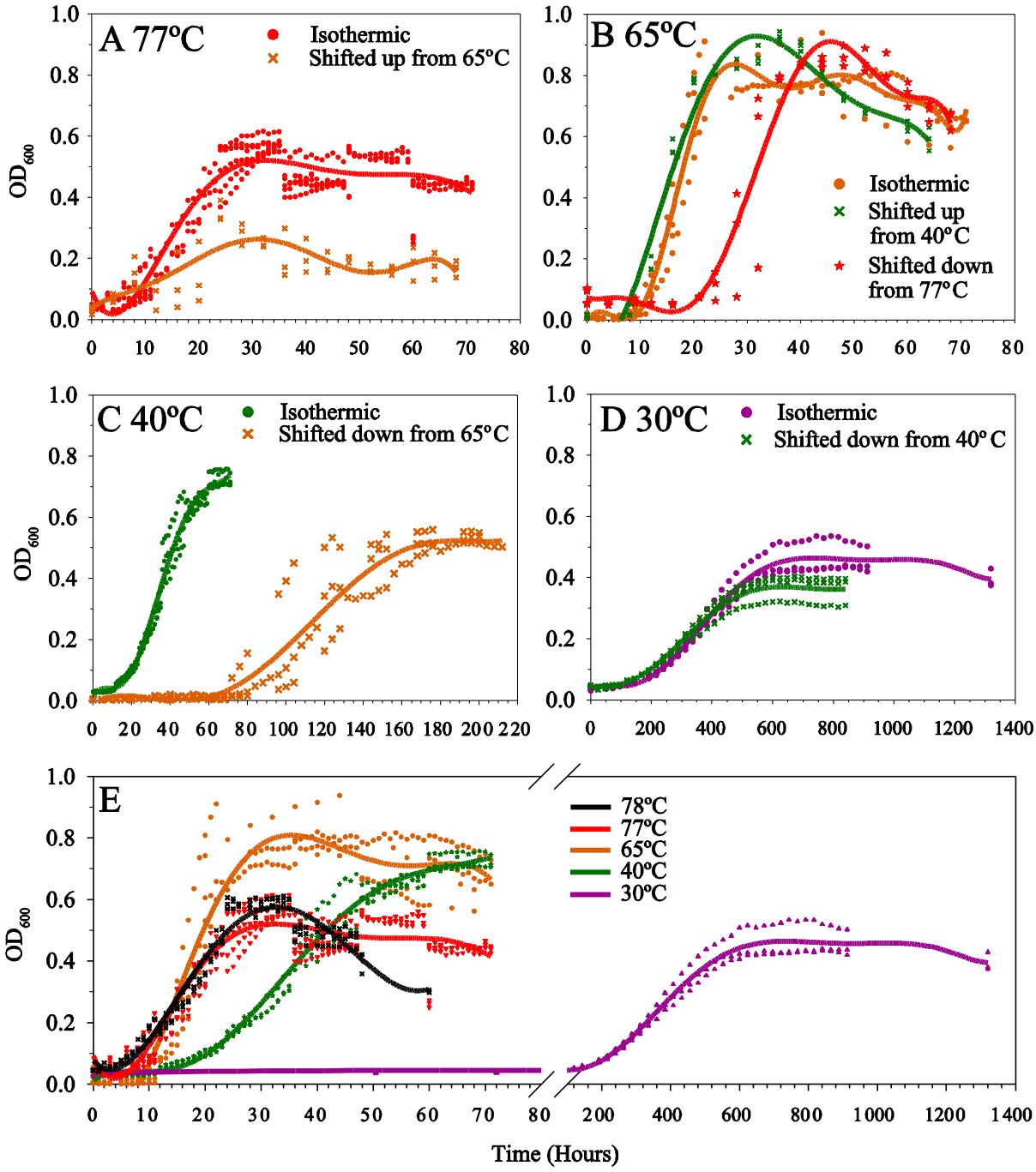


Fig. 2.1. Growth curves of *Kosmotoga olearia* TBF 19.5.1 at different temperatures in the same medium. Curves were individually fitted to data from replicate cultures (See section 2.2.1.1) with a regression line to show the overall trend. Data points and trend lines are coloured according to the growth temperature of the culture used for inoculum: purple, 30°C; green, 40°C; brown, 65°C; red, 77°C; black, 78°C. A) Cultures grown at 77°C, B) cultures grown at 65°C, C) cultures grown at 40°C, D) Cultures grown at 30°C, E) all isothermic growth curves from every temperature.

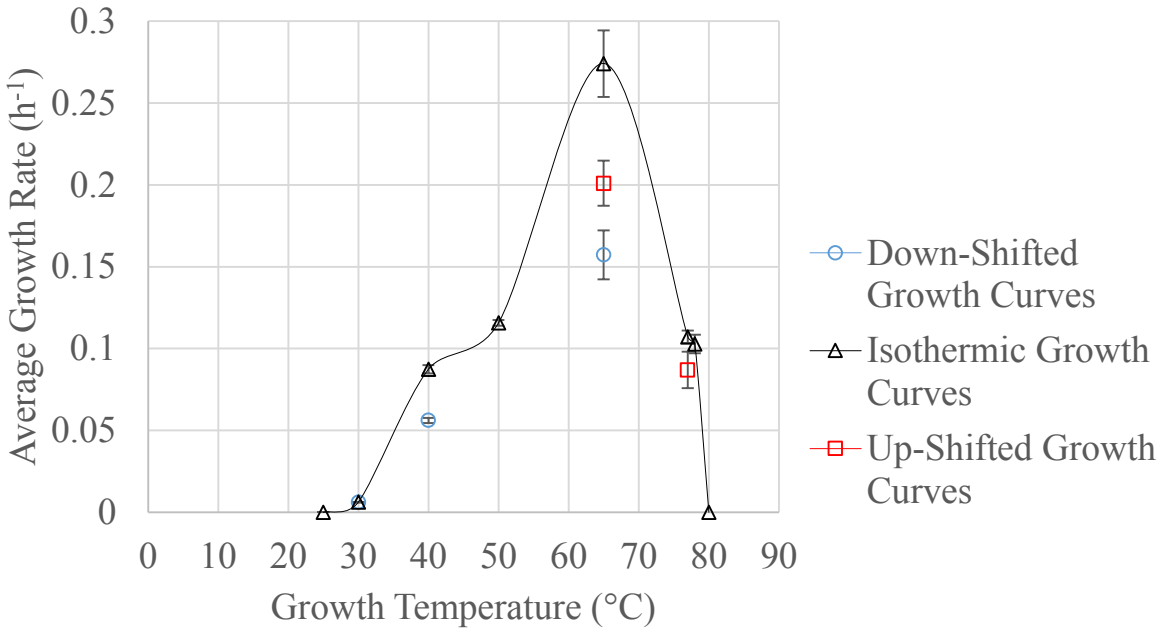


Fig. 2.2. Growth rate of *Kosmotoga olearia* TBF 19.5.1 compared to temperature. Isothermic growth curves were generated at each temperature from an inoculum that had grown at that temperature for at least three transfers (except 25°C and 80°C for which an inoculum from the same temperature could not be generated). Up-shifted growth curves were generated from an inoculum that was grown at lower temperatures while down-shifted growth curves were generated from an inoculum that was grown at higher temperatures. Data points represent the mean of replicate cultures (see Materials and Methods 2.2.1.1); error bars represent standard error.

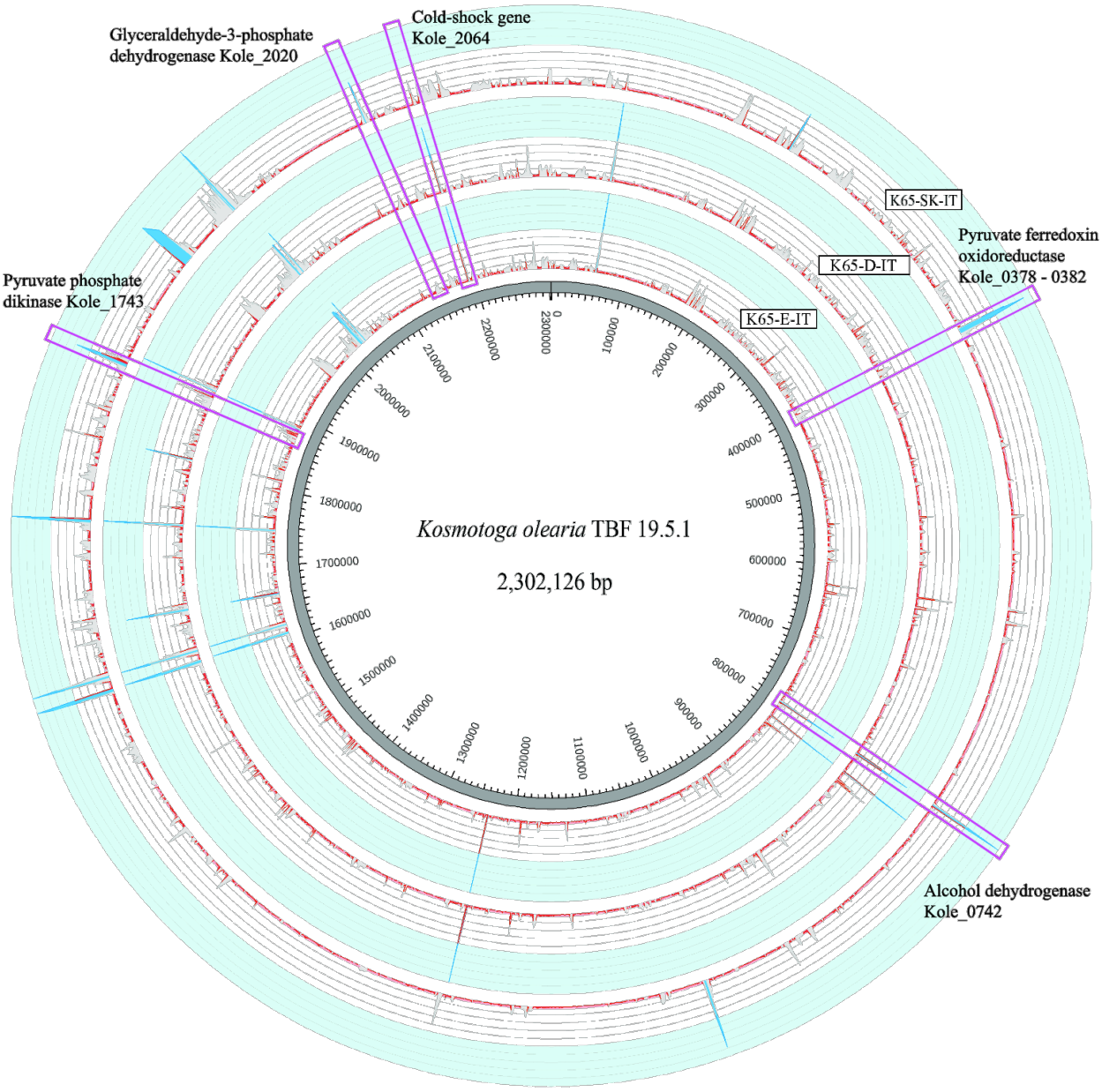


Fig. 2.3. Circos plot of K65-E-IT, K65-D-IT, and K65-SK-IT transcriptomes displaying high-level gene expression. Blue peaks indicate areas with RPKM > 5000. Red indicates RPKM < 250. Each circle ranges from 0 – 10 000. Any genes with RPKM >10 000 will reach the edge of the circle and appear as RPKM = 10 000. Peaks representing genes responding to the RNA isolation procedure and metabolism genes that are expected to be highly expressed are highlighted in boxes.

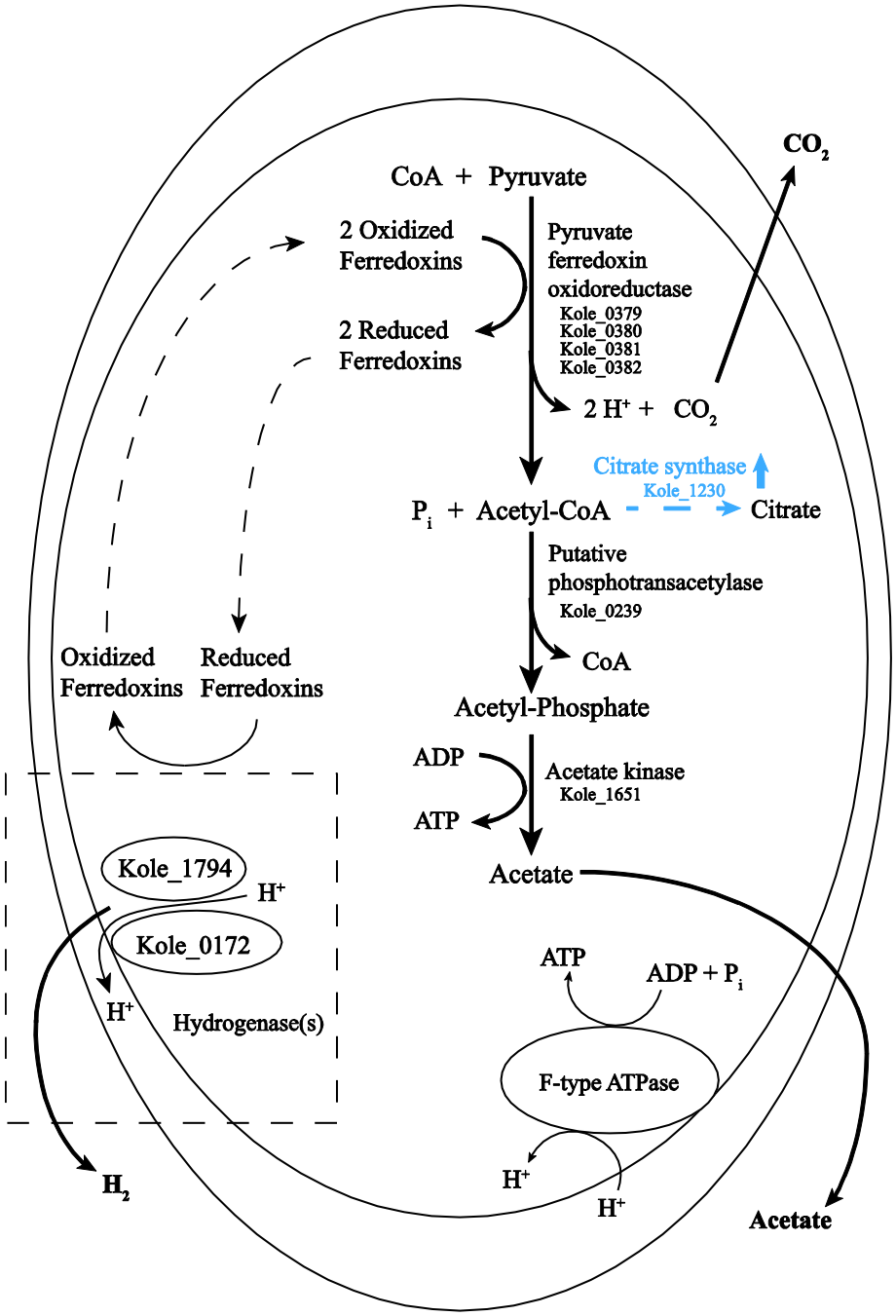


Fig. 2.4. Model of energy generation pathway in *Kosmotoga olearia* during growth on pyruvate. The genes shown here were selected from those predicted to be involved, based on their similarly high expression levels across all temperatures (Table 2.4). Transporters are not shown. The dashed box indicates hydrogenase activity. The two highly expressed hydrogenases are shown but their potential interactions with each other or the membrane are not known. Increased expression of citrate synthase at low temperature, which could redirect acetyl-CoA away from acetate production, is shown in blue.

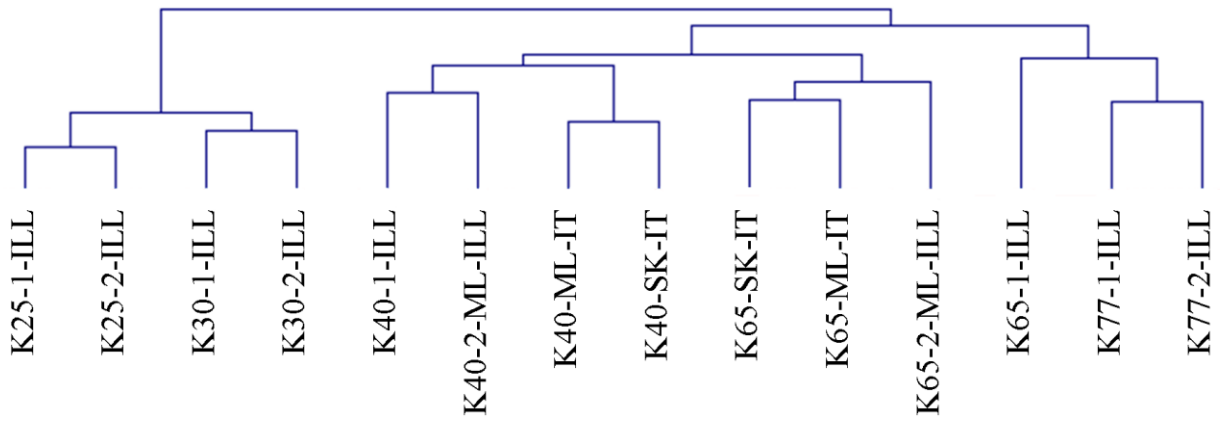
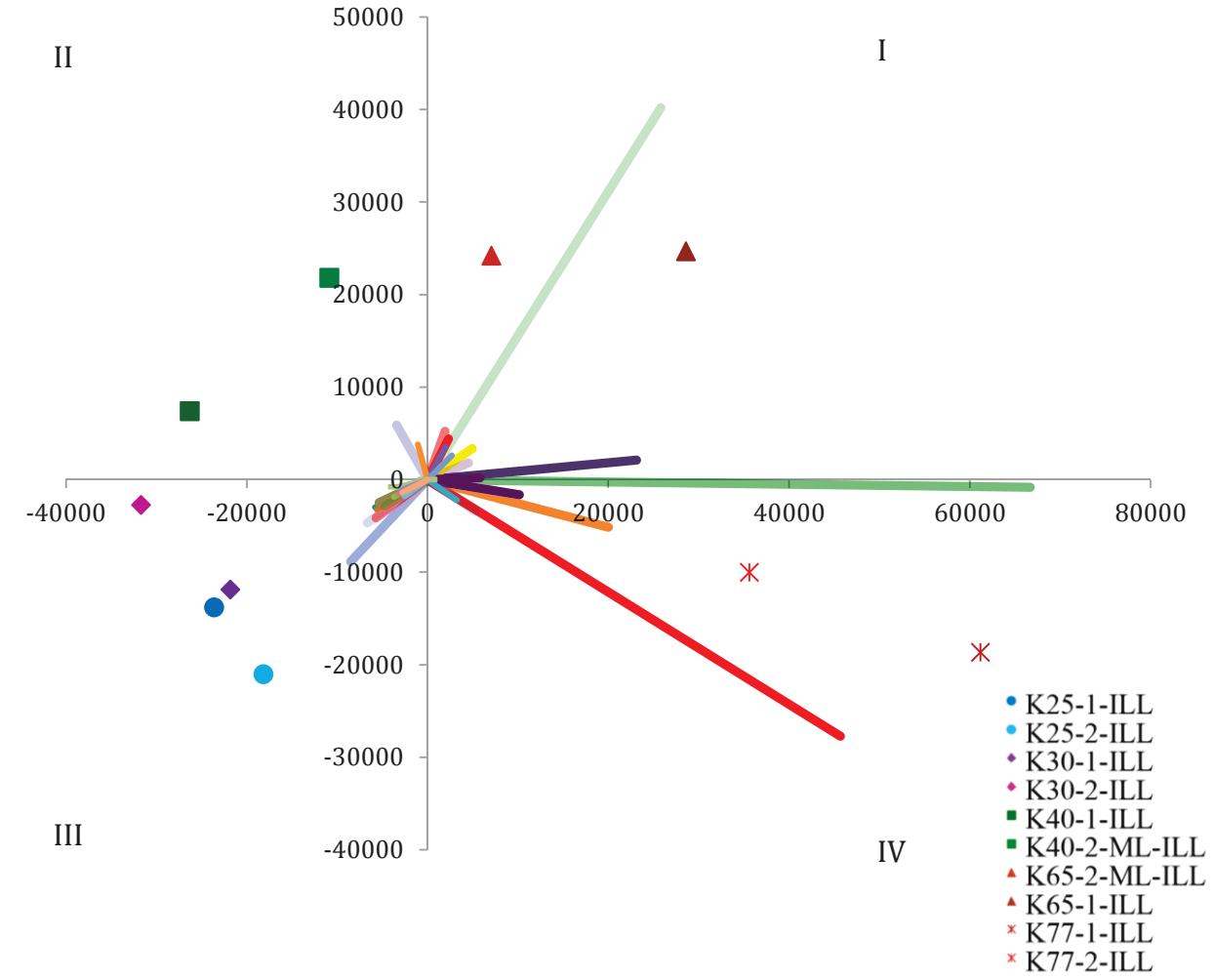


Fig. 2.5. MeV hierarchical clustering of transcriptomes by Manhattan distance.



ssrA	transfer-messenger RNA; tmRNA	IV	Kole_1511	extracellular solute-binding protein family 5 CDS	I
ffs	4.5S RNA component of the signal recognition particle (SRP)	IV	Kole_1322	hypothetical protein CDS	I
Kole_0742	iron-containing alcohol dehydrogenase CDS	I	Kole_1906	30S ribosomal protein S7 CDS	III
Kole_1501	hypothetical protein CDS	I	Kole_2020	glyceraldehyde-3-phosphate dehydrogenase, type I CDS	I
rnpB	catalytic subunit of RNase P, RNA component	IV	Kole_1500	S-layer domain protein CDS	I
Kole_0109	cold-shock DNA-binding domain protein CDS	III	Kole_1876	30S ribosomal protein S11 CDS	III
groEL	groEL - chaperonin CDS	IV	Kole_1394	extracellular solute-binding protein family 1 CDS	I
Kole_1840	50S ribosomal protein L10 CDS	III	Kole_1431	hypothetical protein CDS	IV
rplL	rplL - 50S ribosomal subunit protein L7/L12 CDS	III	Kole_0973	3-oxoacyl-ACP synthase II CDS	II
Kole_0972	acyl carrier protein CDS	II	Kole_1554	basic membrane lipoprotein CDS	I
Kole_2064	cold-shock DNA-binding domain protein CDS	III	rpsD	rpsD - 30S ribosomal subunit protein S4 CDS	III
Kole_0258	ribosomal protein L34 CDS	III	Kole_1707	translation elongation factor Ts CDS	III
Kole_1599	protease Do CDS	IV	Kole_1873	ribosomal protein L17 CDS	III
Kole_0303	extracellular solute-binding protein family 1 CDS	I	Kole_1902	50S ribosomal protein L3 CDS	III
Kole_1907	30S ribosomal protein S12 CDS	III	rpmA	rpmA - 50S ribosomal subunit protein L27 CDS	III

Fig. 2.6. Biplot of Illumina transcriptomes. Each transcriptome is shown as a point from a principal component analysis (See section 2.2.3.2). Additionally the 30 longest vectors calculated from every gene are shown. The vectors are presented in the legend in order of length. Additionally the quadrant into which the vector points is shown beside the label. CDS = Coding DNA Sequence.

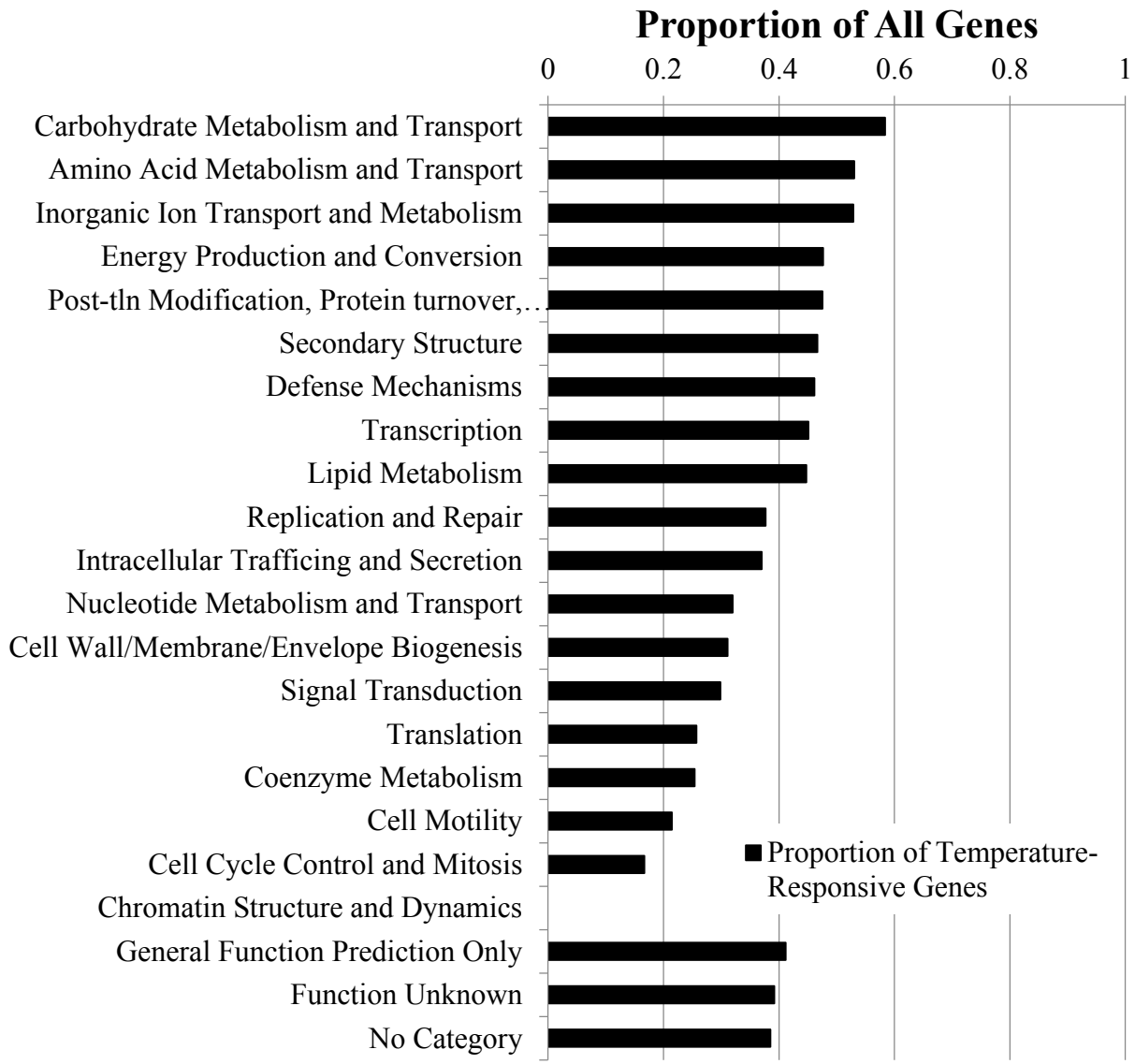


Fig. 2.7. Proportion of temperature-responsive genes out of all genes in *K. olearia* in each COG category. The categories are presented in order of the proportion of temperature-responsive genes they contain.

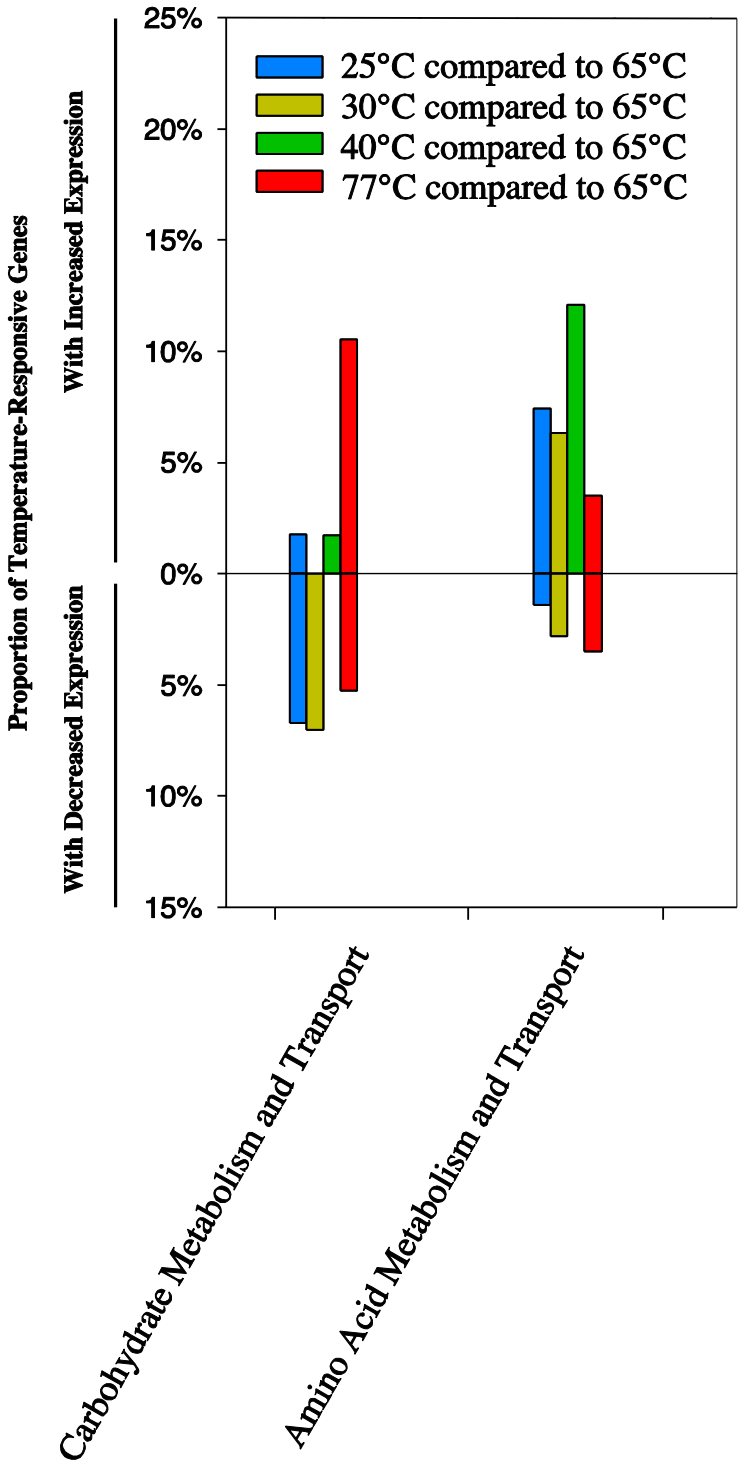


Fig. 2.8. Changes in the COG categories Carbohydrate Metabolism and Transport and Amino Acid Metabolism and Transport at each temperature compared to 65°C. For each temperature comparison the significantly differently expressed genes were sorted into their COG categories and further divided depending on whether they showed higher or lower expression at that temperature.

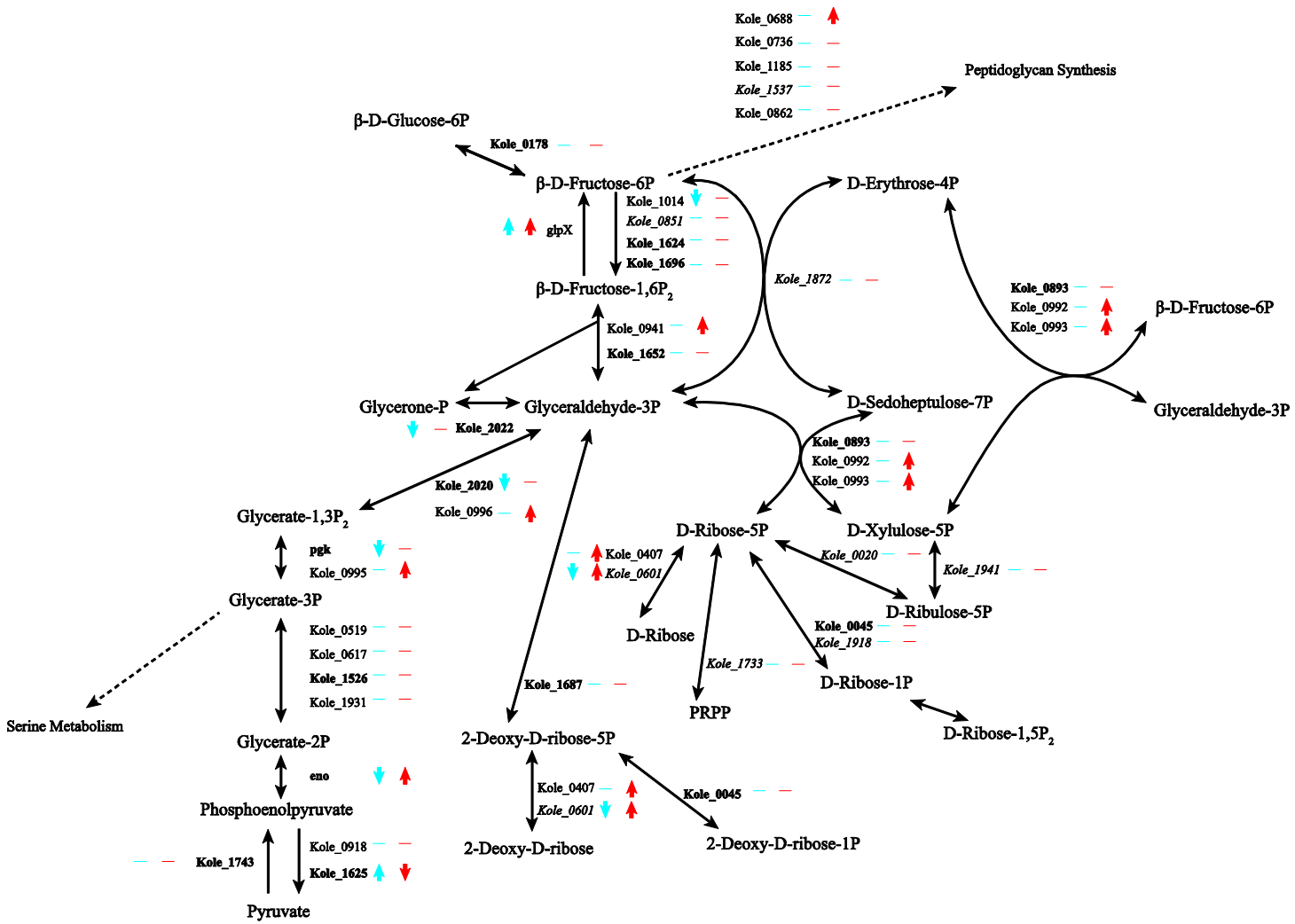


Fig. 2.9. Combined glycolysis and non-oxidative pentose phosphate pathways in *K. olearia*. Gene names in bold have very high expression (RPKM >1000) in one or more transcriptomes while those in italics have moderately high expression (RPKM >300) in one or more transcriptomes. Dashed arrows indicate interaction with other pathways. Blue indicates relative expression at sub-optimal temperatures (< 65°C) while red shows relative expression at supra-optimal temperature (> 65°C). The direction of the arrows indicates the expression relative to optimal conditions. Lines denote that no significant temperature response was observed.

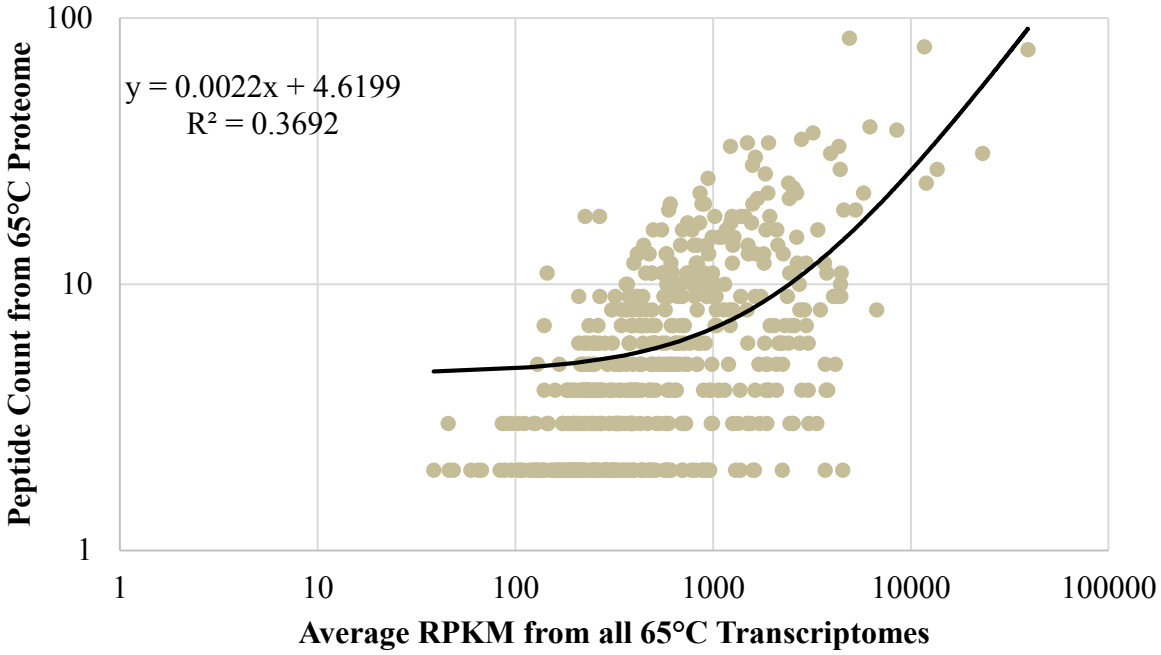


Fig. 2.10. Correlation of mRNA expression and protein abundance. Only the 65˚C proteome was used because it had the greatest number of detected peptides.

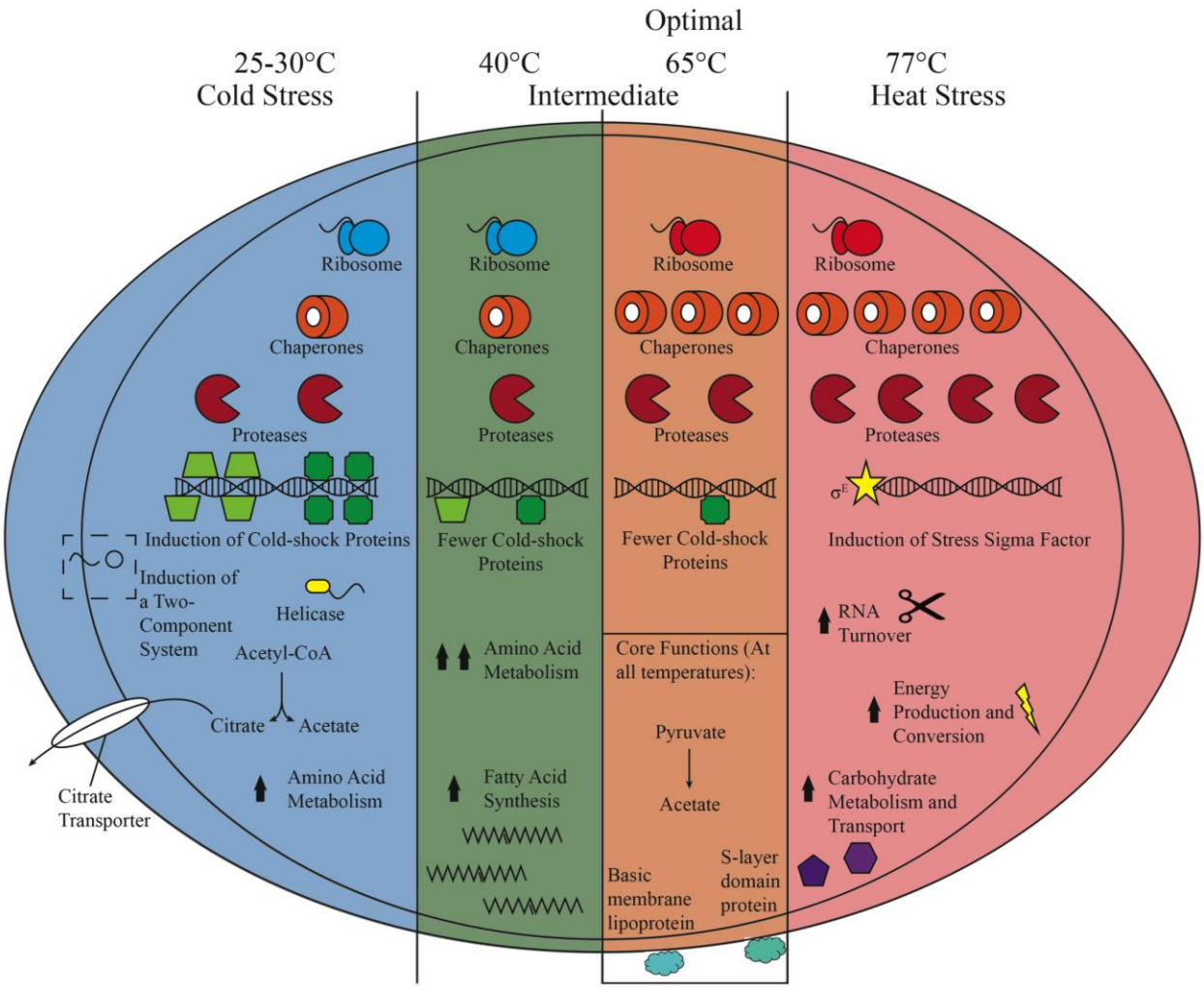


Fig. 2.11. Model of transcriptional changes in *K. olearia* TBF 19.5.1 due to temperature. The major changes outlined in Section 2.4.2 occurring in the three states observed are shown. The different coloured ribosomes indicate changes in ribosomal protein composition at sub-optimal temperatures. The number of chaperone, protease and cold-shock symbols reflects their relative expression at each temperature. Chaperones include groEL, groES and dnaK, proteases include protease Do (Kole_1599) and protease La (Kole_0536), and cold-shock genes include Kole_2064 (dark green squares) and Kole_0109 (green trapezoids). The putative two-component regulatory system (Kole_1015 – 1017) that had a dramatic increase in expression under cold stress is shown. Core functions occurring at all temperatures are shown in the box in the 65°C column (Basic membrane lipoprotein Kole_1554 and S-layer domain protein Kole_1500). Prominent COG category changes (Fig. 2.8) for each temperature condition are also shown. Arrows indicate relative expression compared to optimal temperature.

2.6 References

Balch, W.E., Fox, G.E., Magrum, L.J., Woese, C.R., and Wolfe, R.S. 1979. Methanogens: Reevaluation of a unique biological group. Microbiol. Rev. **43**(2): 260-296.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Weerasinghe, D., Zhang, P., and Karp, P.D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **42**: D459-D471.

DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N., and Noll, K.M. 2009. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int. J. Syst. Evol. Microbiol. **59**: 2991-3000.

D'Ippolito, G., Dipasquale, L., and Fontana, A. 2014. Recycling of carbon dioxide and acetate as lactic acid by the hydrogen-producing bacterium *Thermotoga neapolitana*. ChemSusChem **in press**: 1-7.

Febrer, M., McLay, K., Caccamo, M., Twomey, K.B., and Ryan, R.P. 2011. Advances in bacterial transcriptome and transposon insertion-site profiling using second-generation sequencing. Trends Biotechnol. **29**(11): 586-594.

Frock, A.D., Gray, S.R., and Kelly, R.M. 2012. Hyperthermophilic Thermotoga species differ with respect to specific carbohydrate transporters and glycoside hydrolases. Appl. Environ. Microbiol. **78**(6): 1978-1986.

Graumann, P.L., and Marahiel, M.A. 1999. Cold shock response in *Bacillus subtilis*. J. Mol. Microbiol. Biotechnol. **1**(2): 203-209.

Green, A.G. 2013. Evolution of optimal growth temperature and temperature tolerance range in thermophilic bacteria. B. Sc. thesis, University of Connecticut.

Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. 2011. Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nature reviews.Microbiology **9**(9): 658-669.

Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., and Livny, J. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics **13**(1): 734-745.

Higashibata, H., Fujiwara, S., Ezaki, S., Takagi, M., Fukui, K., and Imanaka, T. 2000. Effect of polyamines on histone-induced DNA compaction of hyperthermophilic Archaea.
J. Biosci. Bioeng. **89**(1): 103-106.

Kaan, T., Homuth, G., Mäder, U., Bandow, J., and Schweder, T. 2002. Genome-wide transcriptional profiling of the *Bacillus subtilis* cold-shock response. Microbiology **148**: 3441-3455.

Latif, H., Lerman, J.A., Portnoy, V.A., Tarasova, Y., Nagarajan, H., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Lee, D., Qiu, Y., and Zengler, K. 2013. The genome organization of *Thermotoga maritima* reflects its lifestyle. PLoS genetics **9**(4): e1003485.

L'Haridon, S., Jiang, L., Alain, K., Chalopin, M., Rouxel, O., Beauverger, M., Xu, H., Shao, Z., and Jebbar, M. 2014. *Kosmotoga pacifica* sp. nov., a thermophilic chemoorganoheterotrophic bacterium isolated from an East Pacific hydrothermal sediment. Extremophiles **18**(1): 81-88.

Markowitz, V.M., Chen, I.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., Anderson, I., Billis, K., Varghese, N., Mavromatis, K., Pati, A., Ivanova, N.N., and Kyrpides, N.C. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. **42**: D560-7.

Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V.V., Osipov, G.A., Belyaev, S.S., and Ivanov, M.V. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*. Int. J. Syst. Evol. Microbiol. **51**: 433-446.

Nunoura, T., Hirai, M., Imachi, H., Miyazaki, M., Makita, H., Hirayama, H., Furushima, Y., Yamamoto, H., and Takai, K. 2010. *Kosmotoga arenicorallina* sp. nov. a thermophilic and obligately anaerobic heterotroph isolated from a shallow hydrothermal system occurring within a coral reef, southern part of the Yaeyama Archipelago, Japan, reclassification of *Thermococcoides shengliensis* as *Kosmotoga shengliensis* comb. nov., and emended description of the genus *Kosmotoga*. Arch. Microbiol. **192**(10): 811-819.

Oshlack, A., and Wakefield, M.J. 2009. Transcript length bias in RNA-seq data confounds systems biology. Biology direct **4**: 14.

Oshlack, A., Robinson, M.D., and Young, M.D. 2010. From RNA-seq reads to differential expression results. Genome Biol. **11**(12): 220.

Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J., Maskell, D.J., Parkhill, J., Choudhary, J., Thomson, N.R., and Dougan, G. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella* Typhi. PLoS genetics **5**(7): e1000569.

Pysz, M.A., Ward, D.E., Shockley, K.R., Montero, C.I., Conners, S.B., Johnson, M.R., and Kelly, R.M. 2004. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8**(3): 209-217.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing .

Raghubeer, E.V., and Matches, J.R. 1990. Temperature range for growth of *Escherichia coli* serotype 0157:H7 and selected coliforms in *E. coli* medium. J. Clin. Microbiol. **28**(4): 803-805.

Robinson, M.D., and Smyth, G.K. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics **9**(2): 321-332.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A.,

Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. 2003. TM4: A free, open-source system for microarray data management and analysis. BioTechniques **34**: 374-378.

Schut, G.J., Boyd, E.S., Peters, J.W., and Adams, M.W.W. 2012. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. FEMS Microbiol. Rev. **37**(2): 182-203.

Simon, D.F., Domingos, R.F., Hauser, C., Hutchins, C.M., Zerges, W., and Wilkinson, K.J. 2013. Transcriptome sequencing (RNA-Seq) analysis of the effects of metal nanoparticle exposure on the transcriptome of *Chlamydomonas reinhardtii*. Appl. Environ. Microbiol. **79**(16): 4774-4785.

Swithers, K.S., DiPippo, J.L., Bruce, D.C., Detter, C., Tapia, R., Han, S., Goodwin, L.A., Han, J., Woyke, T., Pitluck, S., Pennacchio, L., Nolan, M., Mikhailova, N., Land, M.L., Nesbø, C.L., Gogarten, J.P., and Noll, K.M. 2011. 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. **193**(19): 5566-5567.

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., and Horikoshi, K. 2008. Cell proliferation at 122°C and isotopically heavy CH_4 production by a hyperthermophilic methanogen under high-pressure cultivation. Proc. Natl. Acad. Sci. U. S. A. **105**(31): 10949-10954.

Tuorto, S.J., Darias, P., McGuinness, L.R., Panikov, N., Zhang, T., Häggblom, M.M., and Kerkhof, L.J. 2014. Bacterial genome replication at subzero temperatures in permafrost. The ISME journal **8**(1): 139-149.

van Vliet, A.H.M. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol. Lett. **302**(1): 1-7.

Wang, Z., Tong, W., Wang, Q., Bai, X., Chen, Z., Zhao, J., Xu, N., and Liu, S. 2012. The temperature dependent proteomic analysis of *Thermotoga maritima*. PloS one **7**(10): e46463.

Watanabe, T., Kojima, H., and Fukui, M. 2012. Draft genome sequence of a psychrotolerant sulfur-oxidizing bacterium, *Sulfuricella denitrificans* skB26, and proteomic insights into cold adaptation. Appl. Environ. Microbiol. **78**(18): 6545-6549.

Williams, T.J., Lauro, F.M., Ertan, H., Burg, D.W., Poljak, A., Raftery, M.J., and Cavicchioli, R. 2011. Defining the response of a microorganism to temperatures that span its complete growth temperature range (-2°C to 28°C) using multiplex quantitative proteomics. Environ. Microbiol. **13**(8): 2186-2203.

3 COMPARATIVE GENOMIC ANALYSIS OF *KOSMOTOGA* ISOLATES AND IDENTIFICATION OF GENES INVOLVED IN GROWTH AT EXTREME TEMPERATURES

3.1 Introduction

The genus *Kosmotoga* currently has four characterized species: *K. olearia* TBF 19.5.1 (DiPippo et al. 2009), *K. shengliensis* 2SM-2 (formerly *Thermococcoides shengliensis*) (Feng et al. 2010; Nunoura et al. 2010), *K. arenicorallina* S304 (Nunoura et al. 2010), and *K. pacifica* SLHLJ1 (L'Haridon et al. 2014). Table 3.1 shows the reported growth temperature ranges and optima of all four species. Notably, *K. arenicorallina* was reported to have a much smaller temperature range (15°C) than the other members (30 - 60°C). Given its phylogenetic relatedness to *K. olearia* TBF 19.5.1 a comparative genomic approach could identify genes that allow for wide temperature tolerance by *K. olearia* TBF 19.5.1, particularly if an additional genome was included in the comparison. Here a new strain of *K. olearia* was isolated, *K. olearia* DU53, and both its genome and the genome of *K. arenicorallina* S304 were sequenced to facilitate this genomic comparison. Additionally, prediction of laterally acquired genes in *K. olearia* was used to evaluate the role lateral gene transfer (LGT) may have played in the acquisition of wide temperature tolerance in this lineage.

3.2 Materials and Methods

3.2.1 Cultivation of *K. arenicorallina* and confirmation of its growth temperature range

K. arenicorallina was grown in *Kosmotoga* Medium (KTM) since it is a reduced medium that matches *K. arenicorallina*'s optimal NaCl concentration, contains yeast extract [which was determined to be required for growth (Nunoura et al. 2010)], and is well within *K. arenicorallina*'s pH range (6.2-8.0) (Nunoura et al. 2010). KTM consists of (L^{-1} glass-distilled water): 30 g NaCl, 0.335 g KCl, 1.395 g $MgCl_2 \cdot 2H_2O$, 1.4 g $MgSO_4 \cdot 7H_2O$, 0.25 g NH_4Cl, 0.14 g $CaCl_2 \cdot 2H_2O$, 0.45 g KH_2PO_4, 10 mL trace mineral solution (Balch et al. 1979) adjusted to pH 7.0, 5.0 g Bacto yeast extract (BD), 0.5 g cysteine, 6.7 g PIPES, 1 mg resazurin. The pH was adjusted to 6.8 with 10 M NaOH. The medium was sparged with N_2 for 30 min with gentle heating and dispensed anaerobically under a stream of N_2. Serum bottles (Wheaton, 125 mL; Fisher Scientific) containing 50 mL KTM, or Hungate tubes (17 mL; Fisher Scientific) containing 9 mL KTM were sealed with butyl rubber stoppers and aluminum crimps and autoclaved at 121°C for 20 min. All bottles were then supplemented from sterile, anaerobic stocks with 2.5 mL 10% maltose, 0.5 mL 1 M sodium thiosulfate, and 0.5 mL vitamin solution (Balch et al. 1979), and all tubes were supplemented with 0.5 mL 10% maltose, 0.1 mL 1 M sodium thiosulfate, and 0.1 mL vitamin solution. All cultures were inoculated using a sterile syringe and needle and were incubated stationary in the dark. One millilitre of culture that had grown in the same medium was used as inoculum for all cultures. Growth was determined by culture turbidity and the temperature range of *K. arenicorallina* growing on maltose was confirmed. If growth was found at a given temperature a new culture was set up from that culture

and incubated at a new temperature further from the optimal growth temperature of *K. arenicorallina* (60°C).

3.2.2 Isolation, cultivation and determination of growth temperature range of *Kosmotoga olearia* DU53

K. olearia DU53 was isolated by Lauren Bradford from free water knock out (FWKO) water collected from "oil field D" in Alberta, Canada as previously described (Hulecki et al. 2010; Nesbø et al. 2010). Two millilitres of FWKO water that had been stored under anoxic conditions at room temperature for 4 years were injected into 50 mL of KTM supplemented with 0.5 mL vitamin solution, 0.5 mL 1 M sodium thiosulfate, and 2 mL of 10% maltose. The culture bottle was incubated at 55°C for 5 days. The turbid culture was stored at room temperature for 5 weeks. A dilution series was made by transferring 0.2 mL of the enrichment culture into a Hungate tube containing 9 mL KTM supplemented with 0.05 mL vitamin solution, 0.05 mL 1 M sodium thiosulfate, and 0.2 mL 10% maltose and performing serial 10-fold dilutions to 10^{-5}. Three millilitres of molten 3% Gelrite (Sigma G1910) in water was added to each tube to make "bottle plates". Four white, round colonies were picked from the 10^{-5} dilution after incubation at 55°C for 3 weeks, and the 16S rRNA gene sequences of the two that survived were determined after amplification by PCR using primers Eub8F 5' – AGAGTTTGATCMTGGCTCAG – 3' and Univ1392R 5' – ACGGGCGGTGTGTRC – 3'. Isolate DU53 was selected by Lauren for genome sequencing and was therefore used in all subsequent work by me.

After isolation, *K. olearia* DU53 was initially maintained in tubes of KTM supplemented with 0.1 mL vitamin solution, 0.1 mL 1 M sodium thiosulfate, and 0.5 mL 10% maltose. It was transferred to tubes containing KTM supplemented with 0.1 mL vitamin solution, 0.1 mL 1 M sodium thiosulfate, and 0.5 mL 10% sodium pyruvate. It passed through three transfers on this new substrate before determination of its growth temperature range began. Growth was determined by culture turbidity. One millilitre of culture was used to inoculate all cultures.

3.2.3 DNA extraction and whole-genome sequencing

DNA was extracted from 50-mL cultures of both *Kosmotoga arenicorallina* S304 (done by Rhianna Charchuk) and *K. olearia* DU53 (done by Lauren Bradford) grown on maltose. The entire culture contents were pelleted in either a 50-mL conical polypropylene centrifuge tube (Corning 430290) or 250-mL polypropylene centrifuge bottle (Nalgene PPCO) and the supernatant was removed. Cell pellets were resuspended in 3.2 mL of TNE buffer (50 mM Tris, 140 mM NaCl, 5 mM EDTA, pH 7.4) to which 400 µL of 10% N-lauroyl-sarcosine, 400 µL of 10% SDS, and 200 µL 20 mg/mL Proteinase K (Thermo Scientific EO0491; recombinant, PCR grade >600 U/mL) were added. The mixture was incubated overnight at 45°C then transferred in equal volume into eight 2-mL microfuge tubes. One microlitre of 10 mg/mL RNase A (Thermo Scientific EN0531; DNase and protease free) was added to each and the tubes were incubated at 37°C for 1 h. A commercial phenol:chloroform solution (Amnesco 0883; 25:24:1 phenol:chloroform:isoamyl alcohol) (650 µL) was added to each tube and the tubes were vortexed gently for 10 min. The tubes were centrifuged at max. speed (15000 rpm; 21130 g) for 10 min and the top layer from each was collected into new tubes. The phenol:chloroform

extraction step was repeated and one chloroform extraction was performed in the same way using 650 µL chloroform. DNA was precipitated by adding 1 mL of 98% ethanol and incubating the tubes at -20°C overnight before being centrifuged at max. speed. The supernatant was removed and the pellet was washed twice with 500 µL of cold 70% ethanol. The tubes were spun at max. speed for 5 min and any remaining supernatant was removed with a micropipette. The pellets were dried for 15 min at room temperature by leaving the tubes open to the air and resuspended in 50 µL nuclease-free water. The purity and quantity of the DNA were measured using Nanodrop and Qubit instruments before the DNA was used to construct libraries for whole-genome sequencing.

K. olearia DU53 DNA was enzymatically sheared using the Ion Shear Plus kit (Life Technologies) then cloned using the Ion Plus Fragment Library kit (Life Technologies) following the manufacturer's instructions. The library was then sequenced on an Ion Torrent PGM using a 316D chip and 500 flows (done by Rhianna Charchuk).

K. arenicorallina S304 library construction used the Nextera XT kit from Illumina following the manufacturer's instructions. The library was then run as one of ten pooled, barcoded libraries on an Illumina MiSeq using 500 cycles generating 2x250 bp paired-end reads (done by The Applied Genomics Core at the University of Alberta).

3.2.4 Genome assembly and annotation

The genome of *K. arenicorallina* S304 was assembled *de novo* using CLC Genomics Workbench version 7.0.4. Sequencing reads were trimmed using default settings (quality score limit, 0.05;

maximum number of ambiguities, 2; discard reads below 50 nt in length) and assembled with automatic word size, a bubble size corresponding to the average length of the input reads, and a minimum contig length of 1000 bp. Reads were then mapped back to the contigs using default parameters (mismatch cost, 1; insertion cost, 3; deletion cost, 3; length fraction, 0.9; similarity fraction, 0.96, update contigs). Contigs were exported and annotated in RAST (Overbeek et al. 2014). The four contigs containing parts of the two non-identical 16S genes were located and manually resolved into two contigs using its published 16S gene sequences (AB530678 and AB530679). The genome was then re-annotated in RAST and also annotated in the Integrated Microbial Genomes database (IMG; Markowitz et al. 2014) as a draft genome.

The genome of *K. olearia* DU53 was assembled *de novo* (by Camilla Nesbø) using MIRA 3. Ion Torrent default settings were used (http://sourceforge.net/projects/mira-assembler/). The resulting contigs were annotated in RAST and in IMG as a draft genome.

3.2.5 Calculation of Average Nucleotide Identity (ANI)

Average % nucleotide identity was calculated for each pair of *Kosmotoga* genomes (using the website http://enve-omics.ce.gatech.edu/ani/index) following the method of Goris et al. (2007). The query genome is cut into 1000 bp consecutive fragments (every 200 bp) which are used to search against the reference genome using the BLASTN algorithm. The best match is saved for calculation of the mean identity of all matches. Parameters for the calculation were the default: alignment minimum length of 700 bp; alignment minimum identity of 70%; 50 minimum alignments; fragment window size of 1000 bp; fragment step size of 200 bp.

3.2.6 Genome comparison in IMG and correlation to *K. olearia* TBF 19.5.1 transcriptomes

IMG was used to retrieve information for all three *Kosmotoga* genomes and to locate groups of genes that were predicted to be acquired laterally in each. During the microbial genome functional annotation and data integration in IMG, prediction of horizontally transferred genes is one of the standard outputs. All genes from a query genome are searched against the IMG database. A gene is considered putatively horizontally transferred if a block of the top hits (every hit that is ≥0.95 x the maximum bit score) all match a domain, phylum, class, or order that is different from that of the query genome (Markowitz et al. 2014). These genes represent potential lineage-specific transfers from distantly related taxa.

The phylogenetic profiler for single genes was used to examine differences in gene content between the *K. olearia* genomes and to find genes in *K. olearia* TBF 19.5.1 that have homologs in *K. olearia* DU53, but do not have homologs in *K. arenicorallina*, using the following settings: maximum expected-value (e-value) at $1e^{-5}$; minimum percent identity at 30%; including pseudogenes; using the present/absent homologs algorithm. The locus tags of genes in *K. olearia* TBF 19.5.1 were then used to locate their expression in the transcriptomes (Chapter 2) to evaluate their temperature response.

The phylogenetic profiler for single genes was also used to find genes shared between *K. olearia* TBF 19.5.1 and *K. olearia* DU53, but not with the thermophilic or hyperthermophilic Thermotogae. Table 3.2 lists all Thermotogae genomes used in this study. This search used the same parameters listed above.

3.3 Results and Discussion

3.3.1 Growth temperature range of *Kosmotoga olearia* DU53 and *K. arenicorallina* S304

Growth of *K. olearia* DU53 was not obviously different with maltose versus pyruvate. The cultures reached similar maximal OD_{600} values (0.213 for maltose, 0.209 for pyruvate) after overnight growth at 65°C. *K. olearia* DU53 grew from 25 - 79°C. Growth below 25°C was not tested due to time constraints. *K. arenicorallina* grew from 35 - 70°C. No growth was observed at 71°C after 5.5 weeks, nor at 30°C after 35 weeks. Table 3.3 shows the approximate time required for each of the three *Kosmotoga* isolates to reach $OD_{600} \geq 0.1$ at each temperature tested.

3.3.2 Description of three *Kosmotoga* genomes

Table 3.4 shows the properties of the newly sequenced *K. arenicorallina* S304 and *K. olearia* DU53 genomes compared to *K. olearia* TBF 19.5.1 (Swithers et al. 2011). The genome of *K. arenicorallina* is the smallest, with the fewest predicted protein coding genes and highest proportion of coding DNA. The genome of *K. olearia* DU53 on the other hand, is the largest, with the most predicted protein coding genes, and the lowest proportion of coding DNA. All three isolates have similar GC content, two ribosomal operons and similar numbers of tRNAs. Additionally, the IMG metabolism pathway assertion pipeline predicts that all three organisms

are auxotrophic for many amino acids, selenocysteine, biotin, and coenzyme A. The requirement of *K. olearia* TBF 19.5.1 and *K. arenicorallina* S304 for yeast extract in the medium to support growth (DiPippo et al. 2009; Nunoura et al. 2010) may reflect a need for one or more of these components.

Genes predicted by IMG to have been horizontally transferred from outside the Thermotogae were examined for each genome. This prediction for *K. arenicorallina* S304 yielded 83 LGT genes, mostly from Firmicutes (51%) and Archaea (12%), consistent with the strong lateral connection noted between these groups and other Thermotogae (Nelson et al. 1999; Zhaxybayeva et al. 2009; Zhaxybayeva et al. 2012). Likewise, this prediction for *K. olearia* DU53 yielded 41 LGT genes, mostly from Firmicutes (54%) and Archaea (20%) and yielded 311 genes for *K. olearia* TBF 19.5.1, mostly from Firmicutes (45%) and Archaea (9%). The higher number of predicted LGT genes in *K. olearia* TBF 19.5.1 is due to the fact that the automated analysis did not include the newly added genomes of *K. olearia* DU53 and *K. arenicorallina*. Genes shared between either of these genomes and *K. olearia* TBF 19.5.1 would not be considered putative LGT genes using this tool (See Materials and Methods 3.2.6). Therefore a phylogenetic profile for single genes search was conducted (max. e-value, $1e^{-5}$; min. percent identity, 30) to find genes in *K. olearia* TBF 19.5.1 with no homolog in any Thermotogae genome, which resulted in 59 presumptive LGT genes.

The ANI calculated for *K. olearia* TBF 19.5.1 and DU53 of > 99% across 9689 fragments (Table 3.5) supports their assignment to the same species based on the species cutoff value of >95% ANI (Goris et al. 2007). The % ANI between *K. arenicorallina* and either of the *K. olearia* genomes suggests *K. arenicorallina* is part of a distinct lineage. Based on the example from the online tool (http://enve-omics.ce.gatech.edu/ani/index) between *Escherichia* and

Yersinia (84.34% ANI), *K. arenicorallina* could be assigned to a different genus. Differences in 16S rRNA gene sequence between *K. arenicorallina* and *K. olearia* TBF 19.5.1 were not deemed quite enough to assign a novel genus within the Thermotogae by Nunoura et al. (2010), however, adding into consideration the % ANI between *K. arenicorallina* and the two *K. olearia* strains, as well as the relationship to other Thermotogae based on 16S (Chapter 1 Fig. 1.3), *K. arenicorallina* should probably be assigned to a novel genus.

3.3.3 Comparison of *K. olearia* TBF 19.5.1 and *K. olearia* DU53 genomes

Examination of the genes found in *K. olearia* TBF 19.5.1 that have no homologs in *K. olearia* DU53 (max. e-value, $1e^{-5}$; min. percent identity, 30) revealed 89 genes, 62 of which are annotated as encoding hypothetical proteins, and seven of which are annotated as pseudogenes. The remaining 20 genes include transcriptional regulators (Kole_1170, 1276, 1306), a potential restriction endonuclease system (Kole_0063 – 0064), an ammonium transporter (Kole_0200), and proteins of unknown function. Conversely, examination of the genes found in *K. olearia* DU53 with no homolog in *K. olearia* TBF 19.5.1 revealed 169 genes, 76 of which are annotated as hypothetical proteins, and 17 of which are CRISPR-related. Within the remaining 76 genes are many sugar metabolism-related genes such as transporters, kinases, and isomerases, particularly for fucose. Together with the larger genome size of *K. olearia* DU53 (Table 3.4) the presence of these genes suggests this strain may be able to metabolize a larger variety of substrates than *K. olearia* TBF 19.5.1. Moreover, a Fe^{3+} transport system (du53_00807 – 00808) and peroxiredoxin (du53_01426) were present. Additionally, a gene annotated as the antitoxin of a toxin-antitoxin stability system (du53_02416) was found along with two adjacent genes (du53_02417 – 02418),

either of which could be the toxin of the system as both are annotated as hypothetical proteins. The complete list of genes found by these two comparisons is tabulated in Appendix E.

3.3.4 Genes in *K. olearia* TBF 19.5.1 with homologs in *K. olearia* DU53 that have no homologs in *K. arenicorallina*

Since *K. olearia* TBF 19.5.1 and *K. olearia* DU53 were both able to grow across the same wide temperature range and *K. arenicorallina* was found to have a narrower range, genes found in both *K. olearia* strains that were absent in *K. arenicorallina* were sought in order to identify genes that could be important for the extraordinary breadth of the *K. olearia* temperature range. Using the phylogenetic profiler for single genes in IMG (max. e-value of $1e^{-5}$; min. percent identity at 30%) 489 such genes were identified. To further narrow down the genes important for temperature, these 489 genes were compared to the transcriptomes in Chapter 2 to see how many were among the 890 temperature-responsive genes. Of the 489 genes found here, 243 were also responsive in at least one of the temperature comparisons done in Chapter 2. The complete lists of the 489 genes found in *K. olearia* but not in *K. arenicorallina*, and the 243 genes that are also temperature-responsive in *K. olearia* TBF 19.5.1 are found in Appendix F. Among the 243 temperature-responsive genes, 189 have higher expression at 30°C than at 65°C; 192 have higher expression at 77°C than at 65°C; and 26 show highest expression in the intermediate temperatures (40°C and 65°C). Many of the genes show higher expression at both temperature extremes, while others display more cryptic (or variable) expression patterns across the temperature treatments.

While the proteins must have wide temperature ranges of activity to allow an organism to have a wide growth range, this alone does not explain the temperature range of *K. olearia* TBF 19.5.1. The up-regulation at low temperatures (such as 30°C at which *K. arenicorallina* cannot grow) of complete metabolic pathways (such as the pathway from glutamate to ornithine), metabolic genes (e.g., citrate synthase Kole_1230, isocitrate dehydrogenase Kole_1227, 3-isopropylmalate dehydratase Kole_1229), and numerous regulatory genes like transcriptional regulators that are absent in *K. arenicorallina* (Appendix Table F.2), suggests that *K. olearia* TBF 19.5.1 has functional capabilities lacking in *K. arenicorallina*, and that these may play a role in the differences in growth temperature range between the two species. Therefore, the 489 genes found in *K. olearia* but not in *K. arenicorallina* and the 243 temperature-responsive genes in *K. olearia* TBF 19.5.1 were sorted into COG categories assigned by IMG to assess the functions present in *K. olearia* and absent in *K. arenicorallina* (Fig. 3.1). All COG categories to which the genes in *K. olearia* TBF 19.5.1 were assigned in Chapter 2 (Fig. 2.7) are represented among the 489 genes not found in *K. arenicorallina*. Indeed the categories to which most of these 489 genes were assigned, namely Carbohydrate Metabolism and Transport, and Amino Acid Metabolism and Transport, are the categories in which other Thermotogae also show large strain- or lineage-specific variability (Mongodin et al. 2005; Zhaxybayeva et al. 2012). The 243 genes that showed a temperature response in *K. olearia* TBF 19.5.1 also showed a similar distribution across the COG categories, suggesting that temperature affects global metabolism and that no single cellular process or pathway found in *K. olearia* but not in *K. arenicorallina* can explain the differences in growth temperature range between these two species.

Within the 243 temperature-responsive genes not found in *K. arenicorallina* 33 genes showed increased transcription at high temperatures, 70 had increased transcription at low

temperatures, 26 had greatest transcription at intermediate temperatures, and 114 displayed either cryptic expression patterns, or increased transcription at both high and low temperature extremes in *K. olearia* TBF 19.5.1. Some of the genes with increased transcription at low temperatures, such as XRE family transcriptional regulator (Kole_0827), PadR family transcriptional regulator (Kole_1031), and GCN5-related N-acetyltransferase (Kole_1814), may have global effects through regulation of cellular pathways and processes. For example, proteins in the same family as GCN5-related N-acetyltransferases (the GNAT family) are known to acetylate proteins, thereby regulating their activity (Hu et al. 2010). This could be important for activating or changing the expression of pathways needed for growth at temperature extremes. However, classifying the genes into COG categories, 123 of the 243 genes (and 14 of the top 25 genes that show the largest up-regulation at low temperatures) fall into either the General Function Only category, Function Unknown category, or no category. Furthermore, 73 of the 243 genes are annotated as hypothetical proteins. All these genes of unknown, or poorly understood function may represent important components of both the temperature response of *K. olearia* TBF 19.5.1 and the difference between the growth temperature ranges of *K. olearia* and *K. arenicorallina*.

BLASTX and conserved domain searches for 15 of the 73 genes annotated as hypothetical proteins yielded high similarity (BLASTX hits exp. value < $1e^{-4}$; conserved domain hits e-value < $1e^{-3}$) to annotated proteins or protein domains (Table 3.6, Appendix F). Given the search results of the five of these proteins that are up-regulated at 30°C, they may be affiliated with ABC transporters (Kole_0080, 1115), ATPases (Kole_0231), permeases (Kole_1158), and CRISPR-associated proteins (Kole_0674).

The top 25 genes with increased expression at 30°C versus 65°C in *K. olearia* TBF 19.5.1 that are not found in *K. arenicorallina* (Table 3.7) consist of genes associated with mobile

genetic elements, ABC transporters, regulatory genes, and hypothetical proteins. Although finding genes associated with mobile genetic elements among the genes not shared between lineages may not be surprising, an RNA-directed DNA polymerase (Kole_1336), an Integrase (Kole_0628), two CRISPR-associated proteins (Kole_0673 and 0675), and two Transposases (Kole_1798, 1799) are among the 25 genes with the largest transcriptional increases at low temperature. Whether this reflects a response to the growth temperature, or stressful conditions triggering systems unrelated to growth is unclear. Interestingly however, a LVIVD repeat protein (Kole_0192) and PilT protein (Kole_1295) were also found among these 25 genes. LVIVD repeats are a domain found in bacterial and archaeal cell surface proteins (Adindla et al. 2007), while PilT is associated with supplying energy for motility and uptake of foreign DNA (Okamoto and Ohmori 2002; Wilharm et al. 2013). Modifications to the cell surface as well as altered interactions with the surrounding environment, combined with increased activity of genes that could integrate foreign DNA, could suggest induction of natural competence by low temperature in *K. olearia* TBF 19.5.1. More experiments are needed to understand the functions of these genes that show increased transcription at 30°C to determine if they could be involved in this process. Alternatively, the changes in cell surface proteins could be involved in biofilm formation at low temperature.

3.3.5 Prediction of laterally transferred genes in *Kosmotoga olearia*

The putatively laterally transferred genes predicted by IMG for *K. olearia* DU53 and *K. arenicorallina* likely represent recent acquisitions after the *Kosmotoga* and *Mesotoga* lineages (or even the *K. olearia* TBF 19.5.1 and DU53 strains) diverged. Therefore, prediction of

horizontally acquired genes in both strains of *K. olearia* was done using the phylogenetic profiler for single genes in IMG. Genes were sought in *K. olearia* TBF 19.5.1 having homologs in *K. olearia* DU53 but no homologs in any other Thermotogae (not including members of the genus *Mesotoga*). Using this approach, the predicted potentially laterally transferred genes in *K. olearia* may or may not be shared with *Mesotoga* spp. but will not be shared with any of the thermophilic or hyperthermophilic Thermotogae. The resulting 187 genes were queried against the transriptomes in Chapter 2 to see how many were among the 890 temperature-responsive genes. The 187 genes predicted to be laterally acquired by *K. olearia* and 84 of these genes that were temperature-responsive are listed in Appendix G. Sorting both of these sets of genes into the COG categories assigned by IMG (data not shown) yielded distributions similar to those described above and noted for all of the genes in *K. olearia* TBF 19.5.1 (Chapter 2 Fig. 2.7). This supports the notion that no single cellular process acquired by *K. olearia* by LGT can explain the wide temperature range of this lineage.

Within the 187 putative LGT genes in *K. olearia* are groups of genes that are adjacent in the *K. olearia* TBF 19.5.1 genome. Most of these are groups of two adjacent genes, with 26 such pairs being found. There are also six groups of three adjacent genes, two groups of four adjacent genes (Kole_0827-0830, 1808-1811), one group of five adjacent genes (Kole_0682-0686) and one of six adjacent genes (Kole_0469-0474). Notably, the Kole_0827-0830 gene cluster contains an XRE family transcriptional regulator, a hypothetical protein, a ThiJ/PfpI domain protein, and a MazG nucleotide pyrophosphohydrolase. MazG causes depletion of the signalling molecule ppGpp in *E. coli* and was proposed to limit deleterious activity of the MazF toxin, thereby regulating programmed cell death in this bacterium (Gross et al. 2006), while ThiJ and PfpI were found to be a kinase and a protease respectively (Halio et al. 1996; Mizote et al. 1996). All four

genes (Kole_0827-0830) were up-regulated at both temperature extremes in *K. olearia* TBF 19.5.1 (Appendix Table G.2), suggesting a role in promoting cell survival under stressful conditions.

The 187 genes found in *K. olearia* but not in any of the thermophilic or hyperthermophilic Thermotogae were also queried against all the *Mesotoga* genomes in IMG (Table 3.2) using BLAST cutoffs of 30% identity and a maximum e-value of $1e^{-5}$. Forty-six of the genes were shared with one or more of the *Mesotoga* genomes (Appendix Table G.3). Of these 46 genes, 18 were found in all of the genomes of *Mesotoga* isolates (not single-cell amplified genomes) and one was found in all but one of the *Mesotoga* isolate genomes (missing in *Mesotoga sp.* HF07.pep.5.2.highcov). Together these 19 genes may represent a "core" set of cold-related Thermotogae genes. Of the 46 genes shared with at least one *Mesotoga* spp., 16 were temperature-responsive in *K. olearia* TBF 19.5.1, five of which are among the 19 "core" cold-related Thermotogae genes. These five genes display various expression patterns from up-regulated at low temperature (Xaa-Pro aminopeptidase-like protein Kole_0331 and hypothetical protein Kole_1674), to up-regulated at both temperature extremes (von Willebrand factor type A protein Kole_0418), to cryptic (oxidoreductase domain protein Kole_0689 and HAD-superfamily hydrolase, subfamily IA, variant 1 Kole_1189). The von Willebrand factor type A protein (Kole_0418), as a membrane-associated protein, could be involved in maintaining membrane fluidity at extreme temperatures. Alternatively, it could be involved in biofilm formation together with other membrane-associated proteins noted to increase in expression at low temperatures in *K. olearia* TBF 19.5.1 such as LVIVD repeat protein (Kole_0192). Interestingly, an InterProScan (Jones et al. 2014) search of Kole_1674 for protein domains found

sigma3 and sigma4 domains, suggesting this gene may have a global regulatory role at low temperatures as a sigma factor.

All but one (Kole_1321) of the 16 temperature-responsive genes shared with *Mesotoga* spp. showed increased transcription at 30°C compared to 65°C in *K. olearia* TBF 19.5.1. While the temperature-responsive genes represent less than half of the total putative LGT genes shared between *K. olearia* and *Mesotoga* spp., the large proportion of these 16 genes exhibiting increased expression at low temperature suggests that LGT may have enhanced adaptation to low temperatures. The 84 temperature-responsive putative LGT genes in *K. olearia* TBF 19.5.1 contain 18 of the 232 genes that have greater transcription at 25°C and 30°C than at 40°C or 65°C from among the total of 890 temperature-responsive genes identified in Chapter 2. Five of these are shared with *Mesotoga* spp. The numbers of putative LGT genes among the total temperature-responsive genes suggests that LGT has augmented changes to temperature tolerance that were also occurring through other mechanisms such as changes to core genes shared among all Thermotogae.

3.4 Significance

This study combines transcriptomics with comparative genomics to investigate whether genes identified by comparative genomics are involved in temperature adaptation. Through this approach 19 genes were identified as a potential "core" set of cold-related Thermotogae genes from their detection in only *K. olearia* and at least nine of ten *Mesotoga* isolates, and five of these were found to be potentially involved in temperature adaptation. Furthermore, increased

transcription at low temperatures of 15 of the 46 genes shared between *K. olearia* and *Mesotoga* spp. but not found in any of the (hyper)thermophilic Thermotogae suggests that these genes may be important for the low temperature growth of these lineages.

Table 3.1. Growth temperature range and optimum of every characterized *Kosmotoga* species.

***Kosmotoga* species**	**Reported Growth Temperature Range (°C)**	**Reported Growth Temperature Optimum (°C)**	**Reference**
***K. olearia* TBF 19.5.1**	20 – 80	65	(Dipippo et al. 2009)
***K. shengliensis* 2SM-2**	45 - 75	65	(Feng et al. 2010)
***K. arenicorallina* S304**	50 - 65	60	(Nunoura et al. 2010)
***K. pacifica* SLHLJ1**	33 - 78	70	(L'Haridon et al. 2014)

Table 3.2. Thermotogae genomes, as they appear in IMG, used in this study. Single cell genomes have an SC preceding the strain name. [F] = finished genome status, [D] = draft genome status, [P] = permanent draft. The number preceding the genome name is the IMG Genome ID.

640753026	Fervidobacterium nodosum Rt17-B1 [F]
2507149014	Fervidobacterium pennivorans Ven 5, DSM 9078 [F]
2558309038	Kosmotoga arenicorallina S304 [D]
2528311059	Kosmotoga olearia DU53 (Draft genome of Kosmotoga olearia DU53) [D]
644736379	Kosmotoga olearia TBF 19.5.1 [F]
2508501120	Marinitoga piezophila KA3 [F]
2558309017	Marinitoga sp. 1155 [D]
2558309018	Marinitoga sp. 1197 [D]
2510065086	Mesotoga prima MesG1AG4.2 [F]
2558309022	Mesotoga sp. BH458_6_3_2_1 [D]
2556793016	Mesotoga sp. Brook.08.105.5.1 [D]
2563366530	Mesotoga sp. Brook.08.YT.4.2.5.1 [D]
2568526018	Mesotoga sp. Brook.08.YT.4.2.5.2. [D]
2528768015	Mesotoga sp. Brook.08.YT.4.2.5.4. (Mesotoga sp. Brook.08.YT.4.2.5.4_MIRAassembly) [D]
2554235750	Mesotoga sp. H07.pep.5.3 (Mesotoga sp. H07.pep.5.3) [D]
2568526015	Mesotoga sp. HF07.pep.5.2.highcov [D]
2558309021	Mesotoga sp. HF07pep_5_4 [D]
2568526017	Mesotoga sp. PhosAc3 [D]
2579779153	Mesotoga sp. SC_3PWM13N19 [D]
2579779155	Mesotoga sp. SC_4PWA21 [D]
2579779154	Mesotoga sp. SC_4PWL113PWK15 [D]
2582580737	Mesotoga sp. SC_NapDC [D]
2582580735	Mesotoga sp. TolDC [D]
2556793017	Petrotoga halophila DSM 16923 [D]
2556793012	Petrotoga mexicana DSM 14811 [D]
2556793007	Petrotoga miotherma ATCC 51224 [D]
641228500	Petrotoga mobilis SJ95 [F]
2556793011	Petrotoga olearia DSM 13574 [D]
2556793018	Petrotoga sibirica DSM 13575 [D]
2571042355	Petrotoga sp. 8T1HF07.NaAc.6.1 (CLC corrected assembly) [D]
2574179775	Petrotoga sp. 9PW.55.5.1 [D]
2558309008	Petrotoga sp. 9PWA.NaAc.5.4 [D]
2558309007	Petrotoga sp. 9T1HF07.CasAA.8.2 [D]
2571042354	Petrotoga sp. HKA.pet.4.5 [D]
2562617031	Petrotoga sp. HWH.PT.55.6.1 [D]
2558309006	Petrotoga sp. HWHPT.55.6.3 [D]
2558309015	Petrotoga sp. SL27 [D]
2571042353	Petrotoga sp. Shatin.DS.tank11.9.2.9.3 [D]
2531839125	Thermosipho africanus H17ap60334 [P]
2517572017	Thermosipho africanus Ob7, DSM 5309 [D]

643348583 Thermosipho africanus TCF52B [F]
640753057 Thermosipho melanesiensis BI429 [F]
2558309103 Thermotoga elfii DSM 9442 [P]
2558309118 Thermotoga elfii NBRC 107921 [F]
2565956526 Thermotoga hypogea DSM 11164 [P]
2558309119 Thermotoga hypogea NBRC 106472 [F]
641228511 Thermotoga lettingae TMO [F]
2554235387 Thermotoga maritima MSB8 [F]
646311964 Thermotoga naphthophila RKU-10 [F]
643348584 Thermotoga neapolitana DSM 4359 [F]
640427150 Thermotoga petrophila RKU-1 [F]
2517572018 Thermotoga sp. 2812B [D]
2563366758 Thermotoga sp. A7A [P]
2517572015 Thermotoga sp. Cell2 [D]
2531839610 Thermotoga sp. EMP [P]
2517572021 Thermotoga sp. Mc24 [D]
642487181 Thermotoga sp. RQ2 [F]
2517572020 Thermotoga sp. TBGT1765 [D]
2517572083 Thermotoga sp. TBGT1766 [D]
2519103003 Thermotoga sp. Xyl54 [D]
2503508007 Thermotoga thermarum LA3, DSM 5069 [F]
2579779163 Mesoaciditoga lauensis cd-1655R, DSM 25116 [P]
2264867249 Thermotogae bacterium JGI 0000106-O11 (GBS-C_001_293) [P]

Table 3.3. Qualitative growth temperature data from *Kosmotoga* spp. used in this study.

Temperature (°C)	**Time for culture to become turbid (days)**		
	Kosmotoga arenicorallina S304	*Kosmotoga olearia* DU2.5.3	*Kosmotoga olearia* TBF 19.5.1
80	N/A	ND	No growth after 3 weeks
79	N/A	1	1
78	N/A	1	1
77	N/A	1	1
71	No growth after 5.5 weeks	ND	ND
70	3	ND	ND
69	4	ND	ND
68	1	ND	ND
67	2	ND	ND
66	2	ND	ND
65	2	1	1
60	1	1	1
55	ND	1	ND
50	1	1	ND
45	6	1	ND
40	7	ND	2
35	30	3	ND
30	No growth after 34 weeks	24	21
28	N/A	ND	20
25	N/A	>4 months*	>4 months*

ND = Not Determined, N/A = Not Tested because it is beyond the growth range

*Due to the length of time it took for cultures at 25°C to grow, they were not checked every day. When signs of growth were evident it was not clear exactly when growth had begun.

Table 3.4. Comparison of *Kosmotoga* genomes. The three members of the genus *Kosmotoga* that have sequenced genomes are shown. Properties of all three genomes were retrieved from IMG (Markowitz et al. 2014). *K. olearia* TBF 19.5.1 has a closed genome (Swithers et al. 2011) and is considered finished, while *K. olearia* DU2.5.3 and *K. arenicorallina* S304 have draft genomes that may have gaps.

	***Kosmotoga arenicorallina* S304**	***Kosmotoga olearia* DU2.5.3**	***Kosmotoga olearia* TBF 19.5.1**
Genome Size (bp)	> 2,113,910	> 2,375,260	2,302,126
DNA Scaffolds	39	97	1
Average G+C Content (%)	41.0	41.4	41.6
Predicted Protein-coding Genes	1980	2430	2169
Ribosomal Operons	2	2	2
tRNAs	46	44	46
Coding DNA (%)	92.9	89.2	90.2
CRISPR Count	1	8	3
Biosynthetic Clusters	10	14	9
Protein-coding Genes Encoding Signal Peptides	48	43	44
Protein-coding Genes Encoding Transmembrane Proteins	590	623	588

Table 3.5. Average % nucleotide identity (ANI) between the three *Kosmotoga* genomes in this study. Two-way ANI (%) ± standard deviation is shown along with the number of fragments used in the calculation.

Genome	**K. *olearia* DU2.5.3**	**K. *arenicorallina* S304**
***K. olearia* TBF 19.5.1**	99.15 ± 2.18 from 9689 fragments	84.00 ± 8.73 from 174 fragments
K. *olearia* DU2.5.3		83.85 ± 9.19 from 157 fragments

Table 3.6. Fifteen of 73 hypothetical proteins for which a function could be inferred (based on BLASTX and conserved domain search) from among the 243 temperature-responsive genes not found in *K. arenicorallina*.

Locus Tag	Suggested Function	Conserved Domain Information (Marchler-Bauer et al. 2011)
Kole_0080	ABC transporter	Type I periplasmic ligand-binding domain of uncharacterized ABC-type transport systems that are predicted to be involved in the uptake of amino acids, peptides, or inorganic ions; This group includes the type I periplasmic ligand-binding domain of uncharacterized ABC (ATPase Binding Cassette)-type transport systems that are predicted to be involved in the uptake of amino acids, peptides, or inorganic ions. This subgroup has high sequence similarity to members of the family of hydrophobic amino acid transporters (HAAT), such as leucine/isoleucine/valine binding protein (LIVBP); its ligand specificity has not been determined experimentally.
Kole_0231	ATPase	Predicted ATPase (AAA+ superfamily) [General function prediction only]
Kole_0482	Glycoside hydrolase	Glycoside hydrolase family 99, an endo-alpha-1,2-mannosidase; This family of glycoside hydrolases 99 (following the CAZY nomenclature) includes endo-alpha-1,2-mannosidase (EC 3.2.1.130), which is an important membrane-associated eukaryotic enzyme involved in the maturation of N-linked glycans. Specifically, it cleaves mannoside linkages internal to N-linked glycan chains by hydrolyzing an alpha-1,2-mannosidic bond between a glucose-substituted mannose and the remainder of the chain. The biological function and significance of the soluble bacterial orthologs, which may have obtained the genes via horizontal transfer, is not clear.
Kole_0507	MarR family transcriptional regulator	MarR family; The Mar proteins are involved in the multiple antibiotic resistance, a non-specific resistance system. The expression of the mar operon is controlled by a repressor, MarR. A large number of compounds induce transcription of the mar operon. This is thought to be due to the compound binding to MarR, and the resulting complex stops MarR binding to the DNA. With the MarR repression lost, transcription of the operon proceeds. The structure of MarR is known and shows MarR as a dimer with each subunit containing a winged-helix DNA binding motif.
Kole_0674	CRISPR-associated protein	CRISPR/Cas system-associated protein Cas8a1; CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and associated Cas proteins comprise a system for heritable host defense by prokaryotic cells against phage and other foreign DNA; Large proteins, some contain Zn-finger domain; signature gene for I-A subtype; also known as TM1802 family
Kole_0786	XRE family transcriptional regulator	Helix-turn-helix XRE-family like proteins. Prokaryotic DNA binding proteins belonging to the xenobiotic response element family of transcriptional regulators.
Kole_0981	Carbohydrate kinase	ribokinase/pfkB superfamily: Kinases that accept a wide variety of substrates, including carbohydrates and aromatic small molecules, all are phosphorylated at a hydroxyl group. The superfamily includes ribokinase, fructokinase, ketohexokinase, 2-dehydro-3-deoxygluconokinase, 1-phosphofructokinase, the minor 6-phosphofructokinase (PfkB), inosine-guanosine kinase, and adenosine kinase. Even though there is a high degree of structural conservation within this superfamily, their multimerization level varies widely, monomeric (e.g. adenosine kinase), dimeric (e.g. ribokinase), and trimeric (e.g THZ kinase).

Kole_1115	ABC transporter	ABC-2 family transporter protein; This family acts as the transmembrane domain (TMD) of ABC transporters. The family includes proteins responsible for the transport of herbicides.
Kole_1158	Permease	Predicted permease; This family of integral membrane proteins are predicted to be permeases of unknown specificity.
Kole_1266	CRISPR-associated protein	CRISPR/Cas system-associated RAMP superfamily protein; CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and associated Cas proteins comprise a system for heritable host defense by prokaryotic cells against phage and other foreign DNA; RAMP superfamily proteins
Kole_1270	CRISPR-associated protein	CRISPR/Cas system-associated RAMP superfamily protein; CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and associated Cas proteins comprise a system for heritable host defense by prokaryotic cells against phage and other foreign DNA; RAMP superfamily proteins
Kole_1790	Hydrogenase	Thioredoxin (TRX)-like [2Fe-2S] Ferredoxin (Fd) family; composed of [2Fe-2S] Fds with a TRX fold (TRX-like Fds) and proteins containing domains similar to TRX-like Fd including formate dehydrogenases, NAD-reducing hydrogenases and the subunit E of NADH:ubiquinone oxidoreductase (NuoE). TRX-like Fds are soluble low-potential electron carriers containing a single [2Fe-2S] cluster. The exact role of TRX-like Fd is still unclear. It has been suggested that it may be involved in nitrogen fixation. Its homologous domains in large redox enzymes (such as Nuo and hydrogenases) function as electron carriers.
Kole_1820	Peptidase	Peptidase Gluzincin family (thermolysin-like proteinases, TLPs) includes peptidases M1, M2, M3, M4, M13, M32 and M36 (fungalysins); Gluzincin family (thermolysin-like peptidases or TLPs) includes several zinc-dependent metallopeptidases such as the M1, M2, M3, M4, M13, M32, M36 peptidases (MEROPS classification), and contain HEXXH and EXXXD motifs as part of their active site. All peptidases in this family bind a single catalytic zinc ion which is tetrahedrally co-ordinated by three amino acid ligands and a water molecule that forms the nucleophile on activation during catalysis. **Abbreviated description.**
Kole_1991	Hydrolase	Cell wall-associated hydrolases (invasion-associated proteins) [Cell envelope biogenesis, outer membrane]
Kole_1996	Beta-galactosidase	A4 beta-galactosidase middle domain: a type 1 glutamine amidotransferase (GATase1)-like domain; A4 beta-galactosidase middle domain: a type 1 glutamine amidotransferase (GATase1)-like domain. This group includes proteins similar to beta-galactosidase from Thermus thermophilus. Beta-Galactosidase hydrolyzes the beta-1,4-D-galactosidic linkage of lactose, as well as those of related chromogens, o-nitrophenyl-beta-D-galactopyranoside (ONP-Gal) and 5-bromo-4-chloro-3-indolyl-beta-D-galactoside (X-gal). This A4 beta-galactosidase middle domain lacks the catalytic triad of typical GATase1 domains. The reactive Cys residue found in the sharp turn between a beta strand and an alpha helix termed the nucleophile elbow in typical GATase1 domains is not conserved in this group.

Table 3.7. Twenty-five genes with the most increased expression at 30°C compared to 65°C in *K. olearia* TBF 19.5.1 but not found in *K. arenicorallina*. Descriptions in square brackets are from conserved domain searches (Marchler-Bauer et al. 2011). Descriptions in curly brackets are from InterProScan 5 protein domain searches (Jones et al. 2014). Average expression (RPKM) in *K. olearia* TBF 19.5.1 at each temperature (from Chapter 2) is shown.

Locus Tag	Description	Average Expression at 25°C	Average Expression at 30°C	Average Expression at 40°C	Average Expression at 65°C	Average Expression at 77°C	Fold increase in expression at 30°C
Kole_1151	pseudo [Staphylococcal nuclease-like]	297	247	15	22	14	11.07
Kole_1799	transposase IS204/IS1001/IS1096/IS1165 family protein	391	404	62	41	108	9.95
Kole_1798	transposase	324	289	61	30	78	9.51
Kole_1336	RNA-directed DNA polymerase (Reverse transcriptase)	165	222	66	34	45	6.46
Kole_1814	GCN5-related N-acetyltransferase	797	513	78	81	110	6.37
Kole_0674	hypothetical protein [CRISPR-associated protein]	655	577	193	91	123	6.31
Kole_0719	hypothetical protein {transmembrane protein}	581	424	114	70	135	6.06
Kole_0675	CRISPR-associated protein Cas6	353	306	67	51	87	6.05
Kole_0192	LVIVD repeat protein	144	251	65	45	151	5.54
Kole_0673	CRISPR-associated protein, Csh2 family	733	750	313	140	184	5.36
Kole_0364	Integrase catalytic region	207	377	104	71	157	5.27
Kole_0080	hypothetical protein [ABC transporter]	1571	1468	218	282	366	5.21

Kole_0082	ABC transporter	859	889	170	173	221	5.12
Kole_0258	ribosomal protein L34	7326	4070	1960	801	716	5.08
Kole_0335	Integrase catalytic region	118	202	56	40	98	4.98
Kole_0764	hypothetical protein	157	213	104	43	88	4.95
Kole_0093	argininosuccinate lyase	478	367	186	74	78	4.95
Kole_1295	PilT protein	321	314	119	67	76	4.72
Kole_0827	XRE family transcriptional regulator	558	177	30	38	112	4.65
Kole_2091	hypothetical protein {transmembrane protein}	1336	1051	538	231	967	4.55
Kole_0628	Integrase catalytic region	224	314	59	71	192	4.43
Kole_1245	RNA-directed DNA polymerase (Reverse transcriptase)	52	135	42	31	94	4.30
Kole_1339	hypothetical protein	32	109	28	27	95	4.07
argC	argC - N- N-acetyl-gamma-glutamyl-phosphate reductase	492	382	183	95	136	4.03
Kole_0497	hypothetical protein {peptidase}	278	372	285	93	112	4.00

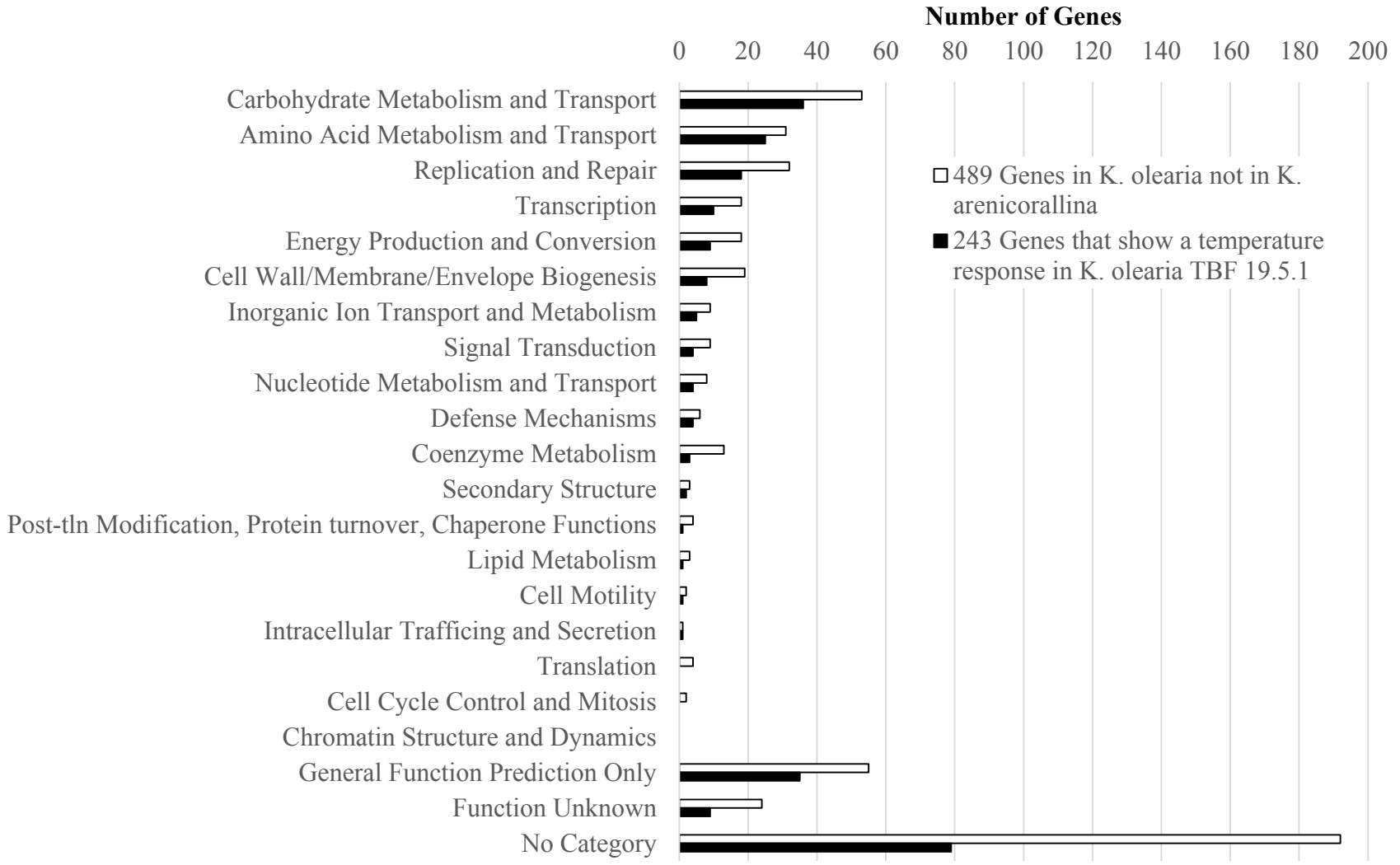


Fig. 3.1. COG categories of all 489 genes found in *K. olearia* and not in *K. arenicorallina* and all 243 of these that were temperature-responsive in Chapter 2. The categories are presented in order of the number of temperature-responsive genes they contain.

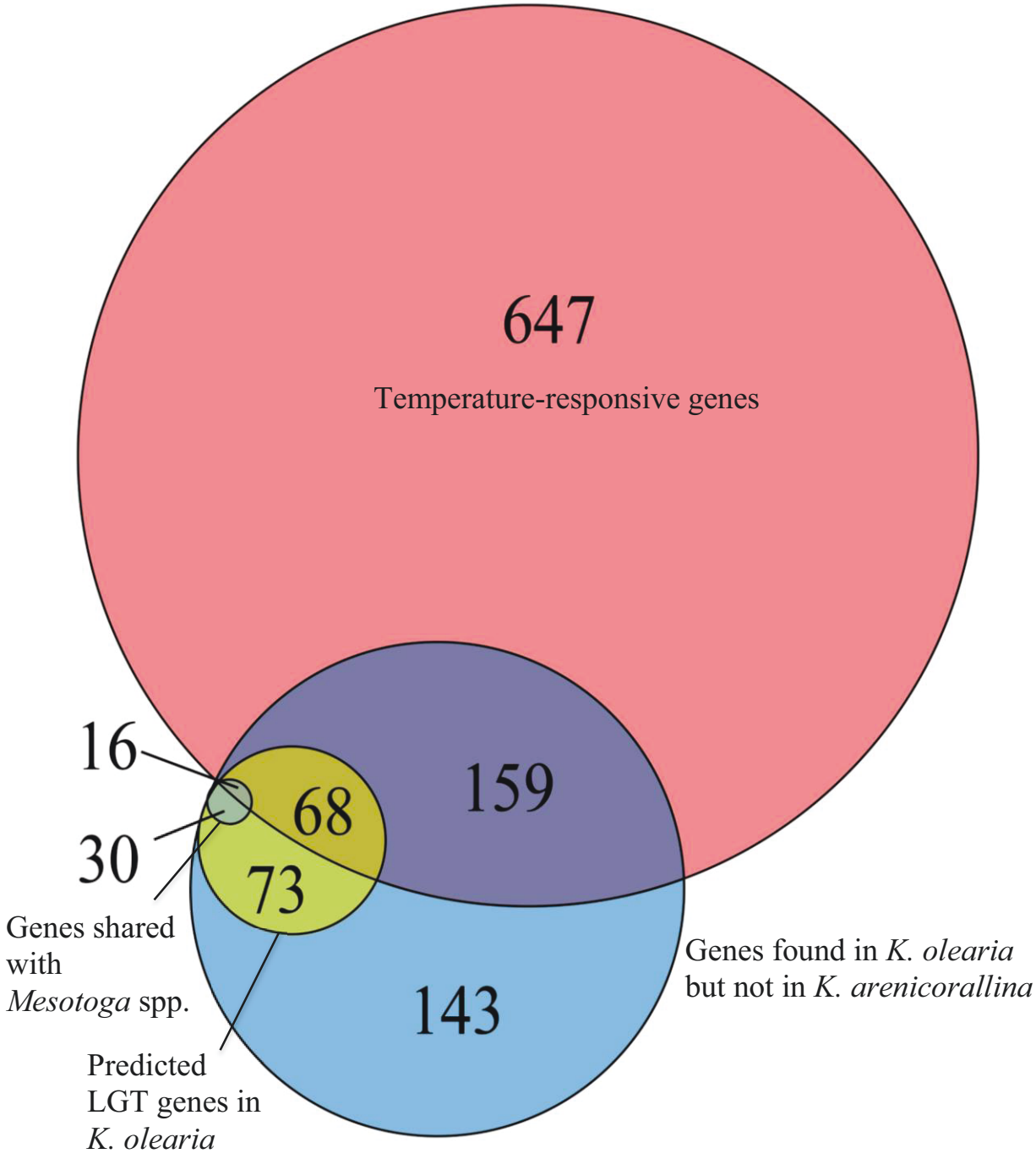


Fig. 3.2. Venn diagram of genes in *K. olearia* TBF 19.5.1. The red circle represents the 890 temperature-responsive genes (Chapter 2). The blue circle represents the 489 genes found in *K. olearia* but not in *K. arenicorallina*. The yellow circle represents the 187 genes predicted to have been acquired laterally by *K. olearia*. The grey circle represents the 46 genes that are predicted to have been acquired laterally by *K. olearia* that are also found in at least one *Mesotoga* genome. The numbers denote how many genes fall into each area of a single colour.

3.5 References

Adindla, S., Inampudi, K.K., and Guruprasad, L. 2007. Cell surface proteins in archaeal and bacterial genomes comprising LVIVD, RIVW and LGxL tandem sequence repeats are predicted to fold as beta-propeller. Int. J. Biol. Macromol. **41**(4): 454-468.

Balch, W.E., Fox, G.E., Magrum, L.J., Woese, C.R., and Wolfe, R.S. 1979. Methanogens: Reevaluation of a unique biological group. Microbiol. Rev. **43**(2): 260-296.

DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N., and Noll, K.M. 2009. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int. J. Syst. Evol. Microbiol. **59**: 2991-3000.

Feng, Y., Cheng, L., Zhang, X., Li, X., Deng, Y., and Zhang, H. 2010. *Thermococcoides shengliensis* gen. nov., sp. nov., a new member of the order *Thermotogales* isolated from oil-production fluid. Int. J. Syst. Evol. Microbiol. **60**: 932-937.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. **57**: 81-91.

Gross, M., Marianovsky, I., and Glaser, G. 2006. MazG - a regulator of programmed cell death in *Escherichia coli*. Mol. Microbiol. **59**(2): 590-601.

Halio, S.B., Blumentals, I.I., Short, S.A., Merrill, B.M., and Kelly, R.M. 1996. Sequence, expression in *Escherichia coli*, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **178**(9): 2605-2612.

Hu, L.I., Lima, B.P., and Wolfe, A.J. 2010. Bacterial protein acetylation: the dawning of a new age. Mol. Microbiol. **77**(1): 15-21.

Hulecki, J.C., Foght, J.M., and Fedorak, P.M. 2010. Storage of oil field-produced waters alters their chemical and microbiological characteristics. J. Ind. Microbiol. Biot. **37**(5): 471-481.

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S., Lopez, R., and Hunter, S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics **30**(9): 1236-1240.

L'Haridon, S., Jiang, L., Alain, K., Chalopin, M., Rouxel, O., Beauverger, M., Xu, H., Shao, Z., and Jebbar, M. 2014. *Kosmotoga pacifica* sp. nov., a thermophilic chemoorganoheterotrophic bacterium isolated from an East Pacific hydrothermal sediment. Extremophiles **18**(1): 81-88.

Markowitz, V.M., Chen, I.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., Anderson, I., Billis, K., Varghese, N., Mavromatis, K., Pati, A., Ivanova, N.N., and Kyrpides, N.C. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. **42**: D560-7.

Mizote, T., Tsuda, M., Nakazawa, T., and Nakayama, H. 1996. The *thiJ* locus and its relation to phosphorylation of hydroxymethylpyrimidine in *Escherichia coli*. Microbiology **142**: 2969-2974.

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. 2005. Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. J. Bacteriol. **187**(14): 4935-4944.

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. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature **399**(6734): 323-329.

Nesbø, C.L., Kumaraswamy, R., Dlutek, M., Doolittle, W.F., and Foght, J. 2010. Searching for mesophilic *Thermotogales* bacteria: "Mesotogas" in the wild. Appl. Environ. Microbiol. **76**(14): 4896-4900.

Nunoura, T., Hirai, M., Imachi, H., Miyazaki, M., Makita, H., Hirayama, H., Furushima, Y., Yamamoto, H., and Takai, K. 2010. *Kosmotoga arenicorallina* sp. nov. a thermophilic and obligately anaerobic heterotroph isolated from a shallow hydrothermal system occurring within a coral reef, southern part of the Yaeyama Archipelago, Japan, reclassification of *Thermococcoides shengliensis* as *Kosmotoga shengliensis* comb. nov., and emended description of the genus *Kosmotoga*. Arch. Microbiol. **192**(10): 811-819.

Okamoto, S., and Ohmori, M. 2002. The cyanobacterial PilT protein responsible for cell motility and transformation hydrolyzes ATP. Plant Cell Physiology **43**(10): 1127-1136.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., and Stevens, R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. **42**: D206-D214.

Swithers, K.S., DiPippo, J.L., Bruce, D.C., Detter, C., Tapia, R., Han, S., Goodwin, L.A., Han, J., Woyke, T., Pitluck, S., Pennacchio, L., Nolan, M., Mikhailova, N., Land, M.L., Nesbø, C.L., Gogarten, J.P., and Noll, K.M. 2011. 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. **193**(19): 5566-5567.

Wilharm, G., Piesker, J., Laue, M., and Skiebe, E. 2013. DNA uptake by the nosocomial pathogen *Acinetobacter baumannii* occurs during movement along wet surfaces. J. Bacteriol. **195**(18): 4146-4153.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. U. S. A. **106**(14): 5865-5870.

Zhaxybayeva, O., Swithers, K.S., Foght, J., Green, A.G., Bruce, D., Detter, C., Han, S., Teshima, H., Han, J., Woyke, T., Pitluck, S., Nolan, M., Ivanova, N., Pati, A., Land, M.L., Dlutek, M., Doolittle, W.F., Noll, K.M., and Nesbø, C.L. 2012. Genome sequence of the mesophilic

Thermotogales bacterium *Mesotoga prima* MesG1.Ag.4.2 reveals the largest Thermotogales genome to date. Genome Biol. Evol. **4**(8): 700-708.

4 CONCLUSION AND SYNTHESIS

4.1 Straddling the divide between thermophily and mesophily

Temperature has a profound effect on cellular structures and function. For this reason, every organism has adapted to grow within a certain temperature range at which it can maintain all of the cellular processes essential for life. Based on this range or on optimal growth temperature, microbes are classified as (hyper)thermophilic, mesophilic or psychrophilic. Furthermore, microbes within each group have specific adaptations to tolerate their corresponding temperature ranges and these adaptations are often not found in members of the other groups (Chapter 1).

While individual organisms may have a growth temperature range that spans either side of the cut-off between thermophily and mesophily (50°C) (Pikuta et al. 2007), typical growth temperature ranges span 20 - 40°C (Chapter 2). *Kosmotoga olearia* TBF 19.5.1, however, has a growth temperature range spanning 54°C (~25 - 79°C) that includes the majority of both mesophilic and thermophilic temperatures (Chapter 2). Examination of its transcriptional responses to sustained growth at various temperatures spanning this range suggests that metabolism in *K. olearia* TBF 19.5.1 is greatly affected by temperature, with many genes showing response to temperature (Chapter 2). This may allow *K. olearia* to play distinct roles in thermally different environments, implying that changes in growth temperature may allow for adoption of a new niche. One such niche could be incorporation into biofilms at low temperatures due to the lower growth rate and increased transcription of cell surface-related proteins at low temperatures (e.g., pilT Kole_1295, von Willebrand factor type A Kole_0418).

Deletion of the slow-growth-activated gene *rpoS* in *E. coli* caused a decrease in biofilm density which was used to support the supposition that slow growth is a major aspect of biofilms (Adams and McLean 1999). Additionally, the increased transcription of CRISPR-related proteins in *K. olearia* TBF 19.5.1 could indicate that there is a higher challenge from viruses at low temperatures.

Since the genus *Kosmotoga* contains other members that have both wide (*K. pacifica*) and narrow (*K. arenicorallina*) temperature ranges, lateral gene transfer (LGT) could have been involved in the acquisition of the breadth of the *K. olearia* temperature range. Comparative genomics between *K. arenicorallina* and both *K. olearia* TBF 19.5.1 and the newly isolated *K. olearia* DU2.5.3 suggests that both LGT and also modification of genes common to all Thermotogae genomes have allowed *K. olearia* to grow at low temperatures (Chapter 3). The changes to these common genes may be evident in the CvP values of the proteins. Furthermore, 15 of the 46 genes found to be shared only between *K. olearia* and *Mesotoga* spp. had increased transcription at low temperatures in *K. olearia* 19.5.1. Further highlighting similarities between the low temperature growth of *Mesotoga* spp. and *K. olearia*, 19 genes were identified as a putative "core" set of cold-related Thermotogae genes, five of which were temperature-responsive in *K. olearia* TBF 19.5.1, implicating them in temperature adaptation.

4.2 Future studies of *K. olearia* TBF 19.5.1

The transcriptomes generated here provide information on one level of the overall response of *K. olearia* TBF 19.5.1 to temperature. Combined with other 'omics strategies, these data could be

used to determine how gene products are regulated, and what effect each has on allowing the cell to cope with different temperatures. Indeed, the combination of proteomic and transcriptomic data was used here to show a correlation between transcript and protein abundance (Chapter 2). Additionally, the transcriptomic data were used to further refine predictions generated by comparative genomic analysis of *K. arenicorallina*, *K. olearia* DU2.5.3, *K. olearia* TBF 19.5.1, and *Mesotoga* spp. The transcriptomic data can also be combined with future proteomic and metabolomic experiments to test predictions made here about the temperature responses of *K. olearia*, such as the production of citrate (or downstream metabolites of citrate) at low temperature (Chapter 2). Moreover, the indications from the genomic data about the component of yeast extract required by *K. olearia* TBF 19.5.1 and *K. arenicorallina* for growth (Chapter 3) can be tested using a culture-based method to potentially find a defined minimal medium upon which these, and perhaps other Thermotogae isolates, can grow. This can aid in the characterization of new isolates as well as provide a means other than antibiotics for selection, which can support physiological and genetic experiments.

4.3 The Thermotogae as a Model System for Temperature Lifestyle Transition

The thermophilic *Kosmotoga* lineage is most closely related to the only known mesophilic Thermotogae lineage *Mesotoga*. Indeed these two lineages are both capable of growth at temperatures lower than any of the other currently known Thermotogae. They also share genes that are not found in any of the other Thermotogae, implicating these genes in the low temperature growth of these two lineages. Among the genes shared between low temperature-

adapted Thermotogae are those likely to have regulatory functions such as transcriptional regulators (Kole_1298 and 1315), a GCN5-related N-acetyltransferase (Kole_1814) and a hypothetical protein with sigma3 and sigma4 domains (Kole_1674). Additionally there is the possibly membrane-associated von Willebrand factor type A (Kole_0418) and a potentially CRISPR-associated protein (Kole_0674) (Chapter 3). Since some genes shared between *K. olearia* and *Mesotoga* spp. are up-regulated at low temperatures in *K. olearia* TBF 19.5.1, comparisons to the transcriptional responses of *Mesotoga* spp. at either their optimal growth temperature or the low end of their growth temperature range could provide insight into both the communal and individual aspects of the low temperature growth of these lineages. This could further inform inferences about the transition from thermophily to mesophily that likely occurred in *Mesotoga* (Zhaxybayeva et al. 2009; Zhaxybayeva et al. 2012). These two closely related lineages could then serve as a model for such transitions in Bacteria. Understanding how Bacteria respond and adapt to changes in temperature in general, on both immediate and evolutionary time scales, will be increasingly important as global temperatures change and microbial populations respond to that change. This project provides the *Kosmotoga* side for direct *Kosmotoga-Mesotoga* temperature comparisons since data have been generated under optimal temperature growth and growth at the low end of the temperature range of *K. olearia* TBF 19.5.1. In addition to the analysis described here, the transcriptomes from *K. olearia* TBF 19.5.1 are available for investigation of any of the genes in this organism, and can be used to confirm whether a gene is transcribed and at what level under different growth temperatures.

4.4 References

Adams, J.L., and McLean, R.J. 1999. Impact of *rpoS* deletion on *Escherichia coli* biofilms. Appl. Environ. Microbiol. **65**(9): 4285-4287.

Pikuta, E.V., Hoover, R.B., and Tang, J. 2007. Microbial extremophiles at the limits of life. Crit. Rev. Microbiol. **33**(3): 183-209.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. U. S. A. **106**(14): 5865-5870.

Zhaxybayeva, O., Swithers, K.S., Foght, J., Green, A.G., Bruce, D., Detter, C., Han, S., Teshima, H., Han, J., Woyke, T., Pitluck, S., Nolan, M., Ivanova, N., Pati, A., Land, M.L., Dlutek, M., Doolittle, W.F., Noll, K.M., and Nesbø, C.L. 2012. Genome sequence of the mesophilic Thermotogales bacterium *Mesotoga prima* MesG1.Ag.4.2 reveals the largest Thermotogales genome to date. Genome Biol. Evol. **4**(8): 700-708.

REFERENCES

Achenbach-Richter, L., Gupta, R., Stetter, K.O., and Woese, C.R. 1987. Were the original eubacteria thermophiles? Syst. Appl. Microbiol. **9**(1-2): 34-39.

Adams, J.L., and McLean, R.J. 1999. Impact of *rpoS* deletion on *Escherichia coli* biofilms. Appl. Environ. Microbiol. **65**(9): 4285-4287.

Adindla, S., Inampudi, K.K., and Guruprasad, L. 2007. Cell surface proteins in archaeal and bacterial genomes comprising LVIVD, RIVW and LGxL tandem sequence repeats are predicted to fold as beta-propeller. Int. J. Biol. Macromol. **41**(4): 454-468.

Atomi, H., Matsumi, R., and Imanaka, T. 2004. Reverse gyrase is not a prerequisite for hyperthermophilic life. J. Bacteriol. **186**(14): 4829-4833.

Balch, W.E., Fox, G.E., Magrum, L.J., Woese, C.R., and Wolfe, R.S. 1979. Methanogens: Reevaluation of a unique biological group. Microbiol. Rev. **43**(2): 260-296.

Ben Hania, W., Ghodbane, R., Postec, A., Brochier-Armanet, C., Hamdi, M., Fardeau, M., and Ollivier, B. 2011. Cultivation of the first mesophilic representative (mesotoga) within the order *Thermotogales*. Syst. Appl. Microbiol. **34**(8): 581-585.

Ben Hania, W., Postec, A., Aüllo, T., Ranchou-Peyruse, A., Erauso, G., Brochier-Armanet, C., Hamdi, M., Ollivier, B., Saint-Laurent, S., Magot, M., and Fardeau, M. 2013. *Mesotoga infera*

sp. nov., a mesophilic member of the order *Thermotogales*, isolated from an underground gas storage aquifer. Int. J. Syst. Evol. Microbiol. **63**: 3003-3008.

Bjørk, A., Dalhus, B., Mantzilas, D., Sirevåg, R., and Eijsink, V.G.H. 2004. Large improvement in the thermal stability of a tetrameric malate dehydrogenase by single point mutations at the dimer-dimer interface. J. Mol. Biol. **341**(5): 1215-1226.

Bjørk, A., Dalhus, B., Mantzilas, D., Eijsink, V.G.H., and Sirevåg, R. 2003. Stabilization of a tetrameric malate dehydrogenase by introduction of a disulfide bridge at the dimer-dimer interface. J. Mol. Biol. **334**(4): 811-821.

Blocher, D., Gutermann, R., Henkel, B., and Ring, K. 1984. Physicochemical characterization of tetraether lipids from *Thermoplasma acidophilum*. Biochemica et Biophysica Acta **778**: 74-80.

Borges, N., Matsumi, R., Imanaka, T., Atomi, H., and Santos, H. 2010. *Thermococcus kodakarensis* mutants deficient in di-*myo*-inositol phosphate use aspartate to cope with heat stress. J. Bacteriol. **192**(1): 191-197.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbø, C.L., Case, R.J., and Doolittle, W.F. 2003. Lateral gene transfer and the origins of prokaryotaic groups. Annu. Rev. Genet. **37**: 283-328.

Boussau, B., Blanquart, S., Necsulea, A., Lartillot, N., and Gouy, M. 2008. Parallel adaptations to high temperatures in the Archaean eon. Nature **456**(7224): 942-945.

Brochier-Armanet, C., and Forterre, P. 2006. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. Archaea **2**(2): 83-93.

Cambillau, C., and Claverie, J. 2000. Structural and genomic correlates of hyperthermostability. The Journal of biological chemistry **275**(42): 32383-32386.

Canaves, J.M. 2004. Predicted role for the Archease protein family based on structural and sequence analysis of TM1083 and MTH1598, two proteins structurally characterized through structural genomics efforts. Proteins **56**(1): 19-27.

Carballeira, N.M., Reyes, M., Sostre, A., Huang, H., Verhagen, M.F., and Adams, M.W. 1997. Unusual fatty acid compositions of the hyperthermophilic archaeon *Pyrococcus furiosus* and the bacterium *Thermotoga maritima*. J. Bacteriol. **179**(8): 2766-2768.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C.A., Holland, T.A., Keseler, I.M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D.S., Weerasinghe, D., Zhang, P., and Karp, P.D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. **42**: D459-D471.

Chang, E.L. 1994. Unusual thermal stability of liposomes made from bipolar tetraether lipids. Biochem. Biophys. Res. Commun. **202**(2): 673-679.

Chhabra, S.R., Shockley, K.R., Ward, D.E., and Kelly, R.M. 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.

Clark, A.T., McCrary, B.S., Edmondson, S.P., and Shriver, J.W. 2004. Thermodynamics of core hydrophobicity and packing in the hyperthermophile proteins Sac7d and Sso7d. Biochemistry **43**: 2840-2853.

Conners, S.B., Mongodin, E.F., Johnson, M.R., Montero, C.I., Nelson, K.E., and Kelly, R.M. 2006. Microbial biochemistry, physiology, and biotechnology of hyperthermophilic *Thermotoga* species. FEMS Microbiol. Rev. **30**(6): 872-905.

Dahle, H., Hannisdal, B., Steinsbu, B.O., Ommedal, H., Einen, J., Jensen, S., Larsen, Ø, Øvreås, L., and Norland, S. 2011. Evolution of temperature optimum in *Thermotogaceae* and the prediction of trait values of uncultured organisms. Extremophiles **15**(4): 509-516.

Damsté, J.S.S., Rijpstra, W.I.C., Hopmans, E.C., Schouten, S., Balk, M., and Stams, A.J.M. 2007. Structural characterization of diabolic acid-based tetraester, tetraether and mixed ether/ester, membrane-spanning lipids of bacteria from the order *Thermotogales*. Arch. Microbiol. **188**(6): 629-641.

DiDonato, M., Deacon, A.M., Klock, H.E., McMullan, D., and Lesley, S.A. 2004. A scaleable and integrated crystallization pipeline applied to mining the *Thermotoga maritima* proteome. J. Struct. Funct. Genomics **5**: 133-146.

DiPippo, J.L., Nesbø, C.L., Dahle, H., Doolittle, W.F., Birkland, N., and Noll, K.M. 2009. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. Int. J. Syst. Evol. Microbiol. **59**: 2991-3000.

D'Ippolito, G., Dipasquale, L., and Fontana, A. 2014. Recycling of carbon dioxide and acetate as lactic acid by the hydrogen-producing bacterium *Thermotoga neapolitana*. ChemSusChem **in press**: 1-7.

Doerrler, W.T., Sikdar, R., Kumar, S., and Boughner, L.A. 2013. New functions for the ancient DedA membrane protein family. J. Bacteriol. **195**(1): 3-11.

Endo, A., Sasaki, M., Maruyama, A., and Kurusu, Y. 2006. Temperature adaptation of *Bacillus subtilis* by chromosomal *groEL* replacement. Biosci. Biotechnol. Biochem. **70**(10): 2357-2362.

Facchiano, A., Colonna, G., and Ragone, R. 1998. Helix-stabilizing factors and stabilization of thermophilic proteins: an X-ray based study. Protein Eng. **11**(9): 753-760.

Febrer, M., McLay, K., Caccamo, M., Twomey, K.B., and Ryan, R.P. 2011. Advances in bacterial transcriptome and transposon insertion-site profiling using second-generation sequencing. Trends Biotechnol. **29**(11): 586-594.

Feng, Y., Cheng, L., Zhang, X., Li, X., Deng, Y., and Zhang, H. 2010. *Thermococcoides shengliensis* gen. nov., sp. nov., a new member of the order *Thermotogales* isolated from oil-production fluid. Int. J. Syst. Evol. Microbiol. **60**: 932-937.

Forterre, P., Bouthier De La Tour C, Philippe, H., and Duguet, M. 2000. Reverse gyrase from hyperthermophiles: probable transfer of a thermoadaptation trait from Archaea to Bacteria. Trends in genetics : TIG **16**(4): 152-154.

Forterre, P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. Trends in genetics : TIG **18**(5): 236-237.

Frock, A.D., Gray, S.R., and Kelly, R.M. 2012. Hyperthermophilic Thermotoga species differ with respect to specific carbohydrate transporters and glycoside hydrolases. Appl. Environ. Microbiol. **78**(6): 1978-1986.

Galtier, N., and Lobry, J.R. 1997. Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. **44**(6): 632-636.

Gao, J., and Wang, W. 2012. Analysis of structural requirements for thermo-adaptation from orthologs in microbial genomes. Annals of Microbiology **62**(4): 1635-1641.

Gonçalves, L.G., Borges, N., Serra, F., Fernandes, P.L., Dopazo, H., and Santos, H. 2012. Evolution of the biosynthesis of di-*myo*-inositol phosphate, a marker of adaptation to hot marine environments. Environ. Microbiol. **14**(3): 691-701.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. **57**: 81-91.

Graumann, P.L., and Marahiel, M.A. 1999. Cold shock response in *Bacillus subtilis*. J. Mol. Microbiol. Biotechnol. **1**(2): 203-209.

Green, A.G. 2013. Evolution of optimal growth temperature and temperature tolerance range in thermophilic bacteria. B. Sc. thesis, University of Connecticut.

Green, A.G., Swithers, K.S., Gogarten, J.F., and Gogarten, J.P. 2013. Reconstruction of ancestral 16S rRNA reveals mutation bias in the evolution of optimal growth temperature in the Thermotogae phylum. Mol. Biol. Evol. **30**(11): 2463-2474.

Gross, M., Marianovsky, I., and Glaser, G. 2006. MazG - a regulator of programmed cell death in *Escherichia coli*. Mol. Microbiol. **59**(2): 590-601.

Gu, J., and Hilser, V.J. 2009. Sequence-based analysis of protein energy landscapes reveals nonuniform thermal adaptation within the proteome. Mol. Biol. Evol. **26**(10): 2217-2227.

Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. 2011. Bacterial transcriptomics: what is beyond the RNA horiz-ome? Nature reviews.Microbiology **9**(9): 658-669.

Haas, B.J., Chin, M., Nusbaum, C., Birren, B.W., and Livny, J. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? BMC Genomics **13**(1): 734-745.

Halio, S.B., Blumentals, I.I., Short, S.A., Merrill, B.M., and Kelly, R.M. 1996. Sequence, expression in *Escherichia coli*, and analysis of the gene encoding a novel intracellular protease (PfpI) from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. **178**(9): 2605-2612.

Higashibata, H., Fujiwara, S., Takagi, M., and Imanaka, T. 1999. Analysis of DNA compaction profile and intracellular contents of archaeal histones from *Pyrococcus kodakaraensis* KOD1. Biochem. Biophys. Res. Commun. **258**(2): 416-424.

Higashibata, H., Fujiwara, S., Ezaki, S., Takagi, M., Fukui, K., and Imanaka, T. 2000. Effect of polyamines on histone-induced DNA compaction of hyperthermophilic Archaea. J. Biosci. Bioeng. **89**(1): 103-106.

Hobbs, J.K., Shepherd, C., Saul, D.J., Demetras, N.J., Haaning, S., Monk, C.R., Daniel, R.M., and Arcus, V.L. 2012. On the origin and evolution of thermophily: Reconstruction of functional precambrian enzymes from ancestors of *Bacillus*. Mol. Biol. Evol. **29**(2): 825-835.

Holder, T., Basquin, C., Ebert, J., Randel, N., Jollivet, D., Conti, E., Jékely, G., and Bono, F. 2013. Deep transcriptome-sequencing and proteome analysis of the hydrothermal vent annelid *Alvinella pompejana* identifies the CvP-bias as a robust measure of eukaryotic thermostability. Biology direct **8**(2): 1-16.

Hu, L.I., Lima, B.P., and Wolfe, A.J. 2010. Bacterial protein acetylation: the dawning of a new age. Mol. Microbiol. **77**(1): 15-21.

Huber, R., and Hannig, M. 2006. Thermotogales. *In* The Prokaryotes. *Edited by* M. Dworkin, S. Falkow, E. Rosenberg, K. Schleifer, Stackebr and E. t. Springer New York pp. 899-922.

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

Hulecki, J.C., Foght, J.M., and Fedorak, P.M. 2010. Storage of oil field-produced waters alters their chemical and microbiological characteristics. J. Ind. Microbiol. Biot. **37**(5): 471-481.

Hurst, L.D., and Merchant, A.R. 2001. High guanine-cytosine content is not an adaptation to high temperature: a comparative analysis amongst prokaryotes. Proc. R. Soc. Lond. B **268**(1466): 493-497.

Imanaka, T. 2011. Molecular bases of thermophily in hyperthermophiles. Proc. Japan Acad. , Ser. B **87**(9): 587-602.

Jolivet, E., Matsunaga, F., Ishino, Y., Forterre, P., Prieur, D., and Myllykallio, H. 2003. Physiological responses of the hyperthermophilic archaeon "*Pyrococcus abyssi*" to DNA damage caused by ionizing radiation. J. Bacteriol. **185**(13): 3958-3961.

Jones, P., Binns, D., Chang, H., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S., Lopez, R., and Hunter, S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics **30**(9): 1236-1240.

Jorge, C.D., Lamosa, P., and Santos, H. 2007. Alpha-D-mannopyranosyl-(1->2)-alpha-D-glucopyranosyl-(1->2)-glycerate in the thermophilic bacterium *Petrotoga miotherma* - structure, cellular content and function. FEBS J. **274**(12): 3120-3127.

Kaan, T., Homuth, G., Mäder, U., Bandow, J., and Schweder, T. 2002. Genome-wide transcriptional profiling of the *Bacillus subtilis* cold-shock response. Microbiology **148**: 3441-3455.

Kimura, H., Mori, K., Yamanaka, T., and Ishibashi, J. 2013. Growth temperatures of archaeal communities can be estimated from the guanine-plus-cytosine contents of 16S rRNA gene fragments. Environ. Microbiol. Rep. **5**(3): 468-474.

Knapp, S., Vos, W.M.D., Rice, D., and Ladenstein, R. 1997. Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. J. Mol. Biol. **267**: 916-932.

Koga, Y. 2012. Thermal adaptation of the archaeal and bacterial lipid membranes. Archaea : doi:10.1155/2012/789652.

Koga, Y., and Morii, H. 2005. Recent advances in structural research on ether lipids from Archaea including comparative and physiological aspects. Biosci. Biotechnol. Biochem. **69**(11): 2019-2034.

Koga, Y., Katsumi, R., You, D., Matsumura, H., Takano, K., and Kanaya, S. 2008. Crystal structure of highly thermostable glycerol kinase from a hyperthermophilic archaeon in a dimeric form. FEBS J. **275**(10): 2632-2643.

Large, A.T., Goldberg, M.D., and Lund, P.A. 2009. Chaperones and protein folding in the Archaea. Biochem. Soc. Trans. **37**: 46-51.

Latif, H., Lerman, J.A., Portnoy, V.A., Tarasova, Y., Nagarajan, H., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Lee, D., Qiu, Y., and Zengler, K. 2013. The genome organization of *Thermotoga maritima* reflects its lifestyle. PLoS genetics **9**(4): e1003485.

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., Wooley, J., Taylor, S.S., Hodgson, K.O., Wilson, I.A., Schultz, P.G., and Stevens, R.C. 2002. Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. Proc. Natl. Acad. Sci. U. S. A. **99**(18): 11664-11669.

L'Haridon, S., Jiang, L., Alain, K., Chalopin, M., Rouxel, O., Beauverger, M., Xu, H., Shao, Z., and Jebbar, M. 2014. *Kosmotoga pacifica* sp. nov., a thermophilic chemoorganoheterotrophic bacterium isolated from an East Pacific hydrothermal sediment. Extremophiles **18**(1): 81-88.

Li, W.F., Zhou, X.X., and Lu, P. 2005. Structural features of thermozymes. Biotechnol. Adv. **23**(4): 271-281.

Mansilla, M.C., Cybulski, L.E., Albanesi, D., and Mendoza, D.D. 2004. Control of membrane lipid fluidity by molecular thermosensors. J. Bacteriol. **186**(20): 6681-6688.

Markowitz, V.M., Chen, I.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., Anderson, I., Billis, K., Varghese, N., Mavromatis, K., Pati, A., Ivanova, N.N., and Kyrpides, N.C. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. **42**: D560-7.

McCloskey, J.A., Graham, D.E., Zhou, S., Crain, P.F., Ibba, M., Konisky, J., Söll, D., and Olsen, G.J. 2001. Post-transcriptional modification in archaeal tRNAs: identities and phylogenetic

relations of nucleotides from mesophilic and hyperthermophilic *Methanococcales*. Nucleic Acids Res. **29**(22): 4699-4706.

Mikulik, K., and Anderova, M. 1994. Role of polyamines in the binding of initiator tRNA to the 70S ribosomes of extreme thermophilic bacterium *Calderobacterium hydrogenophilum*. Arch. Microbiol. **161**: 508-513.

Mizote, T., Tsuda, M., Nakazawa, T., and Nakayama, H. 1996. The *thiJ* locus and its relation to phosphorylation of hydroxymethylpyrimidine in *Escherichia coli*. Microbiology **142**: 2969-2974.

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. 2005. Gene transfer and genome plasticity in *Thermotoga maritima*, a model hyperthermophilic species. J. Bacteriol. **187**(14): 4935-4944.

Mukherjee, A., Sokunbi, A.O., and Grove, A. 2008. DNA protection by histone-like protein HU from the hyperthermophilic eubacterium *Thermotoga maritima*. Nucleic Acids Res. **36**(12): 3956-3968.

Nazina, T.N., Tourova, T.P., Poltaraus, A.B., Novikova, E.V., Grigoryan, A.A., Ivanova, A.E., Lysenko, A.M., Petrunyaka, V.V., Osipov, G.A., Belyaev, S.S., and Ivanov, M.V. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus, Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G.*

kaustophilus, G. thermoglucosidasius and *G. thermodenitrificans*. Int. J. Syst. Evol. Microbiol. **51**: 433-446.

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. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. Nature **399**(6734): 323-329.

Nesbø, C.L., Dlutek, M., Zhaxybayeva, O., and Doolittle, W.F. 2006. Evidence for existence of Mesotogas, members of the order *Thermotogales* adapted to low-temperature environments. Appl. Environ. Microbiol. **72**(7): 5061-5068.

Nesbø, C.L., Kumaraswamy, R., Dlutek, M., Doolittle, W.F., and Foght, J. 2010. Searching for mesophilic *Thermotogales* bacteria: "Mesotogas" in the wild. Appl. Environ. Microbiol. **76**(14): 4896-4900.

Nesbø, C.L., Bradnan, D.M., Adebusuyi, A., Dlutek, M., Petrus, A.K., Foght, J., Doolittle, W.F., and Noll, K.M. 2012. *Mesotoga prima* gen. nov., sp. nov., the first described mesophilic species of the Thermotogales. Extremophiles **16**(3): 387-393.

Nesbø, C.L., Bapteste, E., Curtis, B., Dahle, H., Lopez, P., Macleod, D., Dlutek, M., Bowman, S., Zhaxybayeva, O., Birkeland, N., and Doolittle, W.F. 2009. The genome of *Thermosipho*

africanus TCF52B: lateral genetic connections to the *Firmicutes* and *Archaea*. J. Bacteriol. **191**(6): 1974-1978.

Nishibori, N., Niitsu, M., Fujihara, S., Sagara, T., Nishio, S., and Imai, I. 2009. Occurrence of the polyamines caldopentamine and homocaldopentamine in axenic cultures of the red tide flagellates *Chattonella antiqua* and *Heterosigma akashiwo (Raphidophyceae)*. FEMS Microbiol. Lett. **298**(1): 74-78.

Nunoura, T., Hirai, M., Imachi, H., Miyazaki, M., Makita, H., Hirayama, H., Furushima, Y., Yamamoto, H., and Takai, K. 2010. *Kosmotoga arenicorallina* sp. nov. a thermophilic and obligately anaerobic heterotroph isolated from a shallow hydrothermal system occurring within a coral reef, southern part of the Yaeyama Archipelago, Japan, reclassification of *Thermococcoides shengliensis* as *Kosmotoga shengliensis* comb. nov., and emended description of the genus *Kosmotoga*. Arch. Microbiol. **192**(10): 811-819.

Okamoto, S., and Ohmori, M. 2002. The cyanobacterial PilT protein responsible for cell motility and transformation hydrolyzes ATP. Plant Cell Physiology **43**(10): 1127-1136.

Ollivier, B., and Cayol, J. 2005. Fermentative, iron-reducing, and nitrate-reducing microorganisms. *In* Petroleum Microbiology. *Edited by* B. Ollivier and M. Magot. ASM Press, Washington, DC. pp. 71-88.

Oshima, T., Moriya, T., and Terui, Y. 2011. Identification, chemical synthesis, and biological functions of unusual polyamines produced by extreme thermophiles. Methods Mol. Biol. **720**: 81-111.

Oshlack, A., and Wakefield, M.J. 2009. Transcript length bias in RNA-seq data confounds systems biology. Biology direct **4**: 14.

Oshlack, A., Robinson, M.D., and Young, M.D. 2010. From RNA-seq reads to differential expression results. Genome Biol. **11**(12): 220.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., and Stevens, R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. **42**: D206-D214.

Pace, N.R. 1991. Origin of life- Facing up to the physical setting. Cell **65**: 531-533.

Perkins, T.T., Kingsley, R.A., Fookes, M.C., Gardner, P.P., James, K.D., Yu, L., Assefa, S.A., He, M., Croucher, N.J., Pickard, D.J., Maskell, D.J., Parkhill, J., Choudhary, J., Thomson, N.R., and Dougan, G. 2009. A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus *Salmonella* Typhi. PLoS genetics **5**(7): e1000569.

Pikuta, E.V., Hoover, R.B., and Tang, J. 2007. Microbial extremophiles at the limits of life. Crit. Rev. Microbiol. **33**(3): 183-209.

Prakash, O., and Jaiswal, N. 2010. Alpha-amylase: An ideal representative of thermostable enzymes. Appl. Biochem. Biotechnol. **160**(8): 2401-2414.

Pysz, M.A., Ward, D.E., Shockley, K.R., Montero, C.I., Conners, S.B., Johnson, M.R., and Kelly, R.M. 2004. Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. Extremophiles **8**(3): 209-217.

R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing .

Raghubeer, E.V., and Matches, J.R. 1990. Temperature range for growth of *Escherichia coli* serotype 0157:H7 and selected coliforms in *E. coli* medium. J. Clin. Microbiol. **28**(4): 803-805.

Robinson, M.D., and Smyth, G.K. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics **9**(2): 321-332.

Rodionov, D.A., Rodionova, I.A., Li, X., Ravcheev, D.A., Tarasova, Y., Portnoy, V.A., Zengler, K., and Osterman, A.L. 2013. Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. Front. Microbiol. **4**: 244.

Rodionova, I.A., Leyn, S.A., Burkart, M.D., Boucher, N., Noll, K.M., Osterman, A.L., and Rodionov, D.A. 2013. Novel inositol catabolic pathway in *Thermotoga maritima*. Environ. Microbiol. **15**(8): 2254-2266.

Rodrigues, M.V., Borges, N., Almeida, C.P., Lamosa, P., and Santos, H. 2009. A unique beta-1,2-mannosyltransferase of *Thermotoga maritima* that uses di-*myo*-inositol phosphate as the mannosyl acceptor. J. Bacteriol. **191**(19): 6105-6115.

Rothschild, L.J., and Mancinelli, R.L. 2001. Life in extreme environments. Nature **409**(6823): 1092-1101.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A.,

Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., and Quackenbush, J. 2003. TM4: A free, open-source system for microarray data management and analysis. BioTechniques **34**: 374-378.

Santos, H., Lamosa, P., Borges, N., Gonçalves, L.G., Pais, T., and Rodrigues, M.V. 2011. Organic compatible solutes of prokaryotes that thrive in hot environments: The importance of ionic compounds for thermostabilization. *In* Extremophiles Handbook. *Edited by* K. Horikoshi. Springer Japan, Tokyo. pp. 498-516.

Schut, G.J., Boyd, E.S., Peters, J.W., and Adams, M.W.W. 2012. The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. FEMS Microbiol. Rev. **37**(2): 182-203.

Shockley, K.R., Ward, D.E., Chhabra, S.R., Conners, S.B., Montero, C.I., Robert, M., and Kelly, R.M. 2003. Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl. Environ. Microbiol. **69**(4): 2365-2371.

Simon, D.F., Domingos, R.F., Hauser, C., Hutchins, C.M., Zerges, W., and Wilkinson, K.J. 2013. Transcriptome sequencing (RNA-Seq) analysis of the effects of metal nanoparticle exposure on the transcriptome of *Chlamydomonas reinhardtii*. Appl. Environ. Microbiol. **79**(16): 4774-4785.

Suhre, K., and Claverie, J. 2003. Genomic correlates of hyperthermostability, an update. J. Biol. Chem. **278**(19): 17198-17202.

Swithers, K.S., DiPippo, J.L., Bruce, D.C., Detter, C., Tapia, R., Han, S., Goodwin, L.A., Han, J., Woyke, T., Pitluck, S., Pennacchio, L., Nolan, M., Mikhailova, N., Land, M.L., Nesbø, C.L., Gogarten, J.P., and Noll, K.M. 2011. 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. **193**(19): 5566-5567.

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., and Horikoshi, K. 2008. Cell proliferation at 122°C and isotopically heavy CH_4 production by a hyperthermophilic methanogen under high-pressure cultivation. Proc. Natl. Acad. Sci. U. S. A. **105**(31): 10949-10954.

Taylor, T.J., and Vaisman, I.I. 2010. Discrimination of thermophilic and mesophilic proteins. BMC Struct. Biol. **10**: S5.

Techtmann, S.M., and Robb, F.T. 2010. Archaeal-like chaperonins in bacteria. Proc. Natl. Acad. Sci. U. S. A. **107**(47): 20269-20274.

Thompkins, K., Chattopadhyay, B., Xiao, Y., Henk, M.C., and Doerrler, W.T. 2008. Temperature sensitivity and cell division defects in an *Escherichia coli* strain with mutations in *yghB* and *yqjA*, encoding related and conserved inner membrane proteins. J. Bacteriol. **190**(13): 4489-4500.

Tuorto, S.J., Darias, P., McGuinness, L.R., Panikov, N., Zhang, T., Häggblom, M.M., and Kerkhof, L.J. 2014. Bacterial genome replication at subzero temperatures in permafrost. The ISME journal **8**(1): 139-149.

Ulrih, N.P., Gmajner, D., and Raspor, P. 2009. Structural and physicochemical properties of polar lipids from thermophilic archaea. Appl. Microbiol. Biotechnol. **84**(2): 249-260.

van Vliet, A.H.M. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol. Lett. **302**(1): 1-7.

Wang, Z., Tong, W., Wang, Q., Bai, X., Chen, Z., Zhao, J., Xu, N., and Liu, S. 2012. The temperature dependent proteomic analysis of *Thermotoga maritima*. PloS one **7**(10): e46463.

Watanabe, T., Kojima, H., and Fukui, M. 2012. Draft genome sequence of a psychrotolerant sulfur-oxidizing bacterium, *Sulfuricella denitrificans* skB26, and proteomic insights into cold adaptation. Appl. Environ. Microbiol. **78**(18): 6545-6549.

Wilharm, G., Piesker, J., Laue, M., and Skiebe, E. 2013. DNA uptake by the nosocomial pathogen *Acinetobacter baumannii* occurs during movement along wet surfaces. J. Bacteriol. **195**(18): 4146-4153.

Williams, T.J., Lauro, F.M., Ertan, H., Burg, D.W., Poljak, A., Raftery, M.J., and Cavicchioli, R. 2011. Defining the response of a microorganism to temperatures that span its complete growth temperature range (-2°C to 28°C) using multiplex quantitative proteomics. Environ. Microbiol. **13**(8): 2186-2203.

Zeldovich, K.B., Berezovsky, I.N., and Shakhnovich, E.I. 2007. Protein and DNA sequence determinants of thermophilic adaptation. PLoS computational biology **3**(1): e5.

Zellner, G., and Kneifel, H. 1993. Caldopentamine and caldohexamine in cells of *Thermotoga* species, a possible adaptation to the growth at high temperatures. Arch. Microbiol. **159**(3753): 472-476.

Zhaxybayeva, O., Swithers, K.S., Lapierre, P., Fournier, G.P., Bickhart, D.M., DeBoy, R.T., Nelson, K.E., Nesbø, C.L., Doolittle, W.F., Gogarten, J.P., and Noll, K.M. 2009. On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. Proc. Natl. Acad. Sci. U. S. A. **106**(14): 5865-5870.

Zhaxybayeva, O., Swithers, K.S., Foght, J., Green, A.G., Bruce, D., Detter, C., Han, S., Teshima, H., Han, J., Woyke, T., Pitluck, S., Nolan, M., Ivanova, N., Pati, A., Land, M.L., Dlutek, M., Doolittle, W.F., Noll, K.M., and Nesbø, C.L. 2012. Genome sequence of the mesophilic Thermotogales bacterium *Mesotoga prima* MesG1.Ag.4.2 reveals the largest Thermotogales genome to date. Genome Biol. Evol. **4**(8): 700-708.

Appendix A

Trace Mineral Solution

Nitrilotriacetic acid	1.50 g
$MgSO_4 \cdot 7H_2O$	3.00 g
$MnSO_4 \cdot H_2O$	0.50 g
NaCl	1.00 g
$FeSO_4 \cdot 7H_2O$	0.10 g
$CoSO_4 \cdot 7H_2O$	0.18 g
$CaCl_2 \cdot 2H_2O$	0.10 g
$ZnSO_4 \cdot 7H_2O$	0.18 g
$CuSO_4 \cdot 5H_2O$	0.01 g
$KAl(SO_4)_2 \cdot 12H_2O$	0.02 g
H_3BO_3	0.01 g
$Na_2MoO_4 \cdot 2H_2O$	0.01 g
$NiCl_2 \cdot 6H_2O$	0.03 g
$Na_2SeO_3 \cdot 5H_2O$	0.30 mg
Distilled Water	Final Volume: 1 L

First dissolve Nitrilotriacetic acid and adjust pH to 6.5 with KOH, then add minerals.

Final pH 7.0 (with KOH).

Vitamin Solution

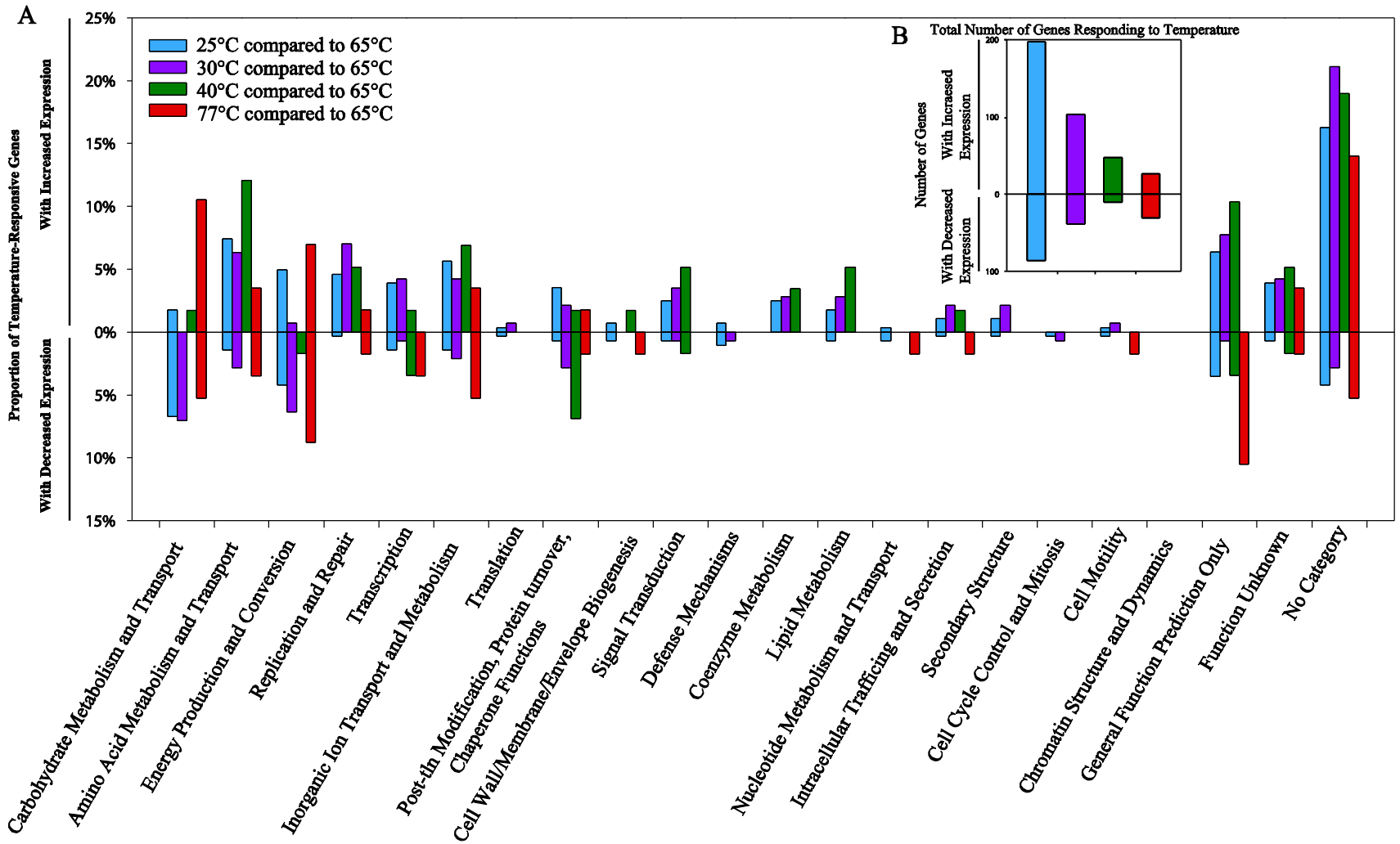
Biotin	2.00 mg
Folic acid	2.00 mg
Pyridoxine-HCl	10.00 mg
Thiamine-HCl·$2H_2O$	5.00 mg
Riboflavin	5.00 mg
Nicotinic acid	5.00 mg
D-Ca-pantothenate	5.00 mg
Vitamin B_{12}	0.10 mg
p-Aminobenzoic acid	5.00 mg
Lipoic acid	5.00 mg
Distilled Water	Final Volume: 1 L

Appendix B.1. Seventeen genes considered to be differentially detected between K65-D-IT and K65-E-IT transcriptomes. The same criteria were used as for differential expression analysis of temperature treatments (Section 2.2.3.3).

Description	Locus Tag	Fold Change in K65-D-IT compared to K65-E-IT*	K65-E-IT Expression values (RPKM)	K65-E-IT Total gene reads	K65-D-IT Expression values (RPKM)	K65-D-IT Total gene reads
ABC transporter CDS	Kole_1836	-3.74	245	24	66	102
degV family protein CDS	Kole_0423	-3.39	418	23	124	108
transfer-messenger RNA; tmRNA	ssrA	-2.98	43900	983	14718	5239
hypothetical protein CDS	Kole_0994	-2.84	608	49	214	274
catalytic subunit of RNase P, RNA component	rnpB	-2.49	37508	680	15065	4342
diguanylate cyclase with PAS/PAC sensor CDS	Kole_0660	-2.33	517	60	222	410
hypothetical protein CDS	Kole_1322	2.28	947	32	2160	1160
hypothetical protein CDS	Kole_1678	2.27	1601	33	3629	1189
hypothetical protein CDS	Kole_2026	-2.22	370	20	167	143
CBS domain containing membrane protein CDS	Kole_0347	-2.15	1353	39	631	289
hypothetical protein CDS	Kole_1698	-2.11	1707	116	807	872
thioesterase superfamily protein CDS	Kole_0966	-2.09	800	20	382	152
peptidylprolyl isomerase FKBP-type CDS	Kole_1745	2.08	2345	64	4887	2120
engB - probably GTP-binding protein CDS	engB	-2.08	631	23	304	176
hypothetical protein CDS	Kole_0808	-2.05	571	35	279	272
ABC transporter CDS	Kole_1149	-2.05	364	22	178	171
nusB - transcription antitermination protein CDS	nusB	-2.02	844	24	418	189

CDS = Coding DNA Sequence

*Fold Change of Expression Values



Appendix B.2. COG category changes at each temperature compared to 65°C. For each temperature comparison the significantly differently expressed genes were sorted into their COG categories and further divided depending on whether they showed higher or lower expression at that temperature. (A) The proportion of genes with increased and decreased expression in each COG category at each temperature compared to 65°C. (B) The total number of genes with increased and decreased expression at each temperature compared to 65°C.

Appendices C – G

Included in this thesis are electronic copies of the full lists of genes discussed. These can be accessed through ERA: Education and Research Archive at the University of Alberta by using the following link: http://hdl.handle.net/10402/era.39639