Genomic insights and temperature-dependent transcriptional responses of *Kosmotoga olearia*, a deep-biosphere bacterium that can grow from 20ºC to 79ºC

Stephen M. J. Pollo[1], Abigail A. Adebusuyi[1#], Timothy J. Straub[2], Julia M. Foght[1], Olga Zhaxybayeva[2,3] and Camilla L. Nesbø[1,4*]

[1]Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

[2]Department of Biological Sciences, Dartmouth College, Hanover, NH, USA

[3]Department of Computer Science, Dartmouth College, Hanover, NH, USA

[4]Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, Oslo, Norway

*** Correspondence:**
Corresponding Author
nesbo@ualberta.ca

[#]**Current address:** Centre for Research & Innovation, Grande Prairie Regional College, Grande Prairie, AB

Keywords: Thermotogae, RNA-Seq, lateral gene transfer, temperature response, cold response, thermal adaptation

Word count for the text: 8755 words

4 Figures

Abstract

Temperature is one of the defining parameters of an ecological niche, and ambient temperature change is a physiological challenge faced by all living cells. Most organisms are adapted to growing within a temperature range that rarely exceeds ~ 30°C, but the anaerobic thermophilic deep subsurface bacterium *Kosmotoga olearia* is capable of growing over an extremely wide temperature range (20°C - 79°C). To identify genes with expression patterns correlated with this flexible phenotype, we compared transcriptomes of *K. olearia* cultures grown at its optimal 65°C to those at 30°C, 40°C, and 77°C. In addition to expected differences in growth rate we found that the temperature treatments significantly affected expression of 573 of 2,224 *K. olearia*'s genes. At different temperatures *K. olearia* remodels its metabolism dramatically, with increased expression of genes involved in energy and carbohydrate metabolism at high temperatures and up-regulation of amino acid metabolism at lower temperatures. At sub-optimal temperatures, many transcriptional changes were similar to those observed in mesophilic bacteria at physiologically low temperatures, including up-regulation of genes encoding enzymes for fatty acid synthesis, typical cold stress genes and ribosomal proteins. In comparison to other Thermotogae, *K. olearia* has multiple copies of some cold-associated genes, consistent with observations that increase in gene copy number is a strategy for adaptation to low temperature environments. Many of these cold response genes are predicted to have been laterally acquired and, particularly among the transferred genes shared with *Mesotoga*, a larger portion are up-regulated at low temperatures, suggesting that gene exchange plays a role in bacterial thermoadaptation. Notably, at 77°C one-third of the up-regulated genes encode proteins with hypothetical functions, indicating that many features of response to high temperature and stress may still be unknown. Our finding of coordinated temperature-specific gene expression patterns, and by extension temperature specific metabolism, suggests that *Kosmotoga* populations encounter variable environments, probably through migration. Therefore, we conjecture that deep subsurface microbial communities are more dynamic than currently perceived.

1 Introduction

Microorganisms are capable of growing over an impressive temperature range, at least from -15°C to 122°C (Takai et al., 2008; Mykytczuk et al., 2013), and temperature is one of the most important physical factors determining their distribution, diversity, and abundance (Schumann, 2009). However, individual microbial species grow only within a much narrower temperature interval. For example, *Escherichia coli* O157:H7 thrives in the laboratory between 19°C and 41°C (Raghubeer and Matches, 1990), while *Geobacillus thermoleovorans* has a growth range of 37°C to 70°C (Dinsdale et al., 2011). Microorganisms with temperature ranges >50°C are rare and, to date, research into the few having ranges >40°C has focused on psychrophiles (e.g., (Mykytczuk et al., 2013)). *Kosmotoga olearia* TBF 19.5.1 (hereafter referred to as *K. olearia*) is an anaerobic thermophile from the bacterial phylum Thermotogae with a growth range that spans almost 60°C (DiPippo et al., 2009). How does this organism achieve such physiological flexibility and what are the evolutionary advantages and implications of this capability?

Fluctuations in temperature induce broad physiological changes in cells, including growth rate, alterations to cell wall and membrane composition, translation, and energy metabolism (Barria et al., 2013; Pollo et al., 2015; Schumann, 2009). These physiological changes can be classified into two broad types of cellular response. Cold or heat *shock* designates the changes observed *immediately* after the shift of a culture to a lower or higher temperature, while *prolonged growth* at a specific lower or higher temperature elicits an *acclimated* low- or high-temperature response (Barria et al., 2013; Schumann, 2009). Most studies of prokaryotes have focused on temperature shock responses rather than acclimated growth. Among the Thermotogae, responses to both heat shock and prolonged growth at high temperatures have been studied in the hyperthermophile *Thermotoga maritima*, which can grow between 55°C and 90°C (Pysz et al., 2004; Wang et al., 2012). During prolonged high temperature growth *T. maritima* strongly up-regulates central carbohydrate metabolism genes and expresses a few typical heat shock protein genes (Wang et al., 2012). Little is known about how *T. maritima* responds to sub-optimal temperatures, although it encodes some genes implicated in cold shock response. For example, its family of cold shock proteins (Csp), which are nucleic acid chaperones known to be induced during cold shock and cold acclimation in mesophilic bacteria (Barria et al., 2013; Phadtare, 2004), exhibits nucleic acid melting activity at physiologically low temperatures (Phadtare et al., 2003). Similarly, responses to cold shock in a few other thermophiles involve many of the genes implicated in mesophilic cold shock response (e.g., (Boonyaratanakornkit et al., 2005; Mega et al., 2010)). In this study we systematically assessed bacterial physiological changes associated with response to prolonged growth at both high and low temperature using *K. olearia* as a model system. Such changes can reflect not only response to temperature itself, but also growth rate effects and general responses to stress. By conflating these factors, our study examined overall changes in *K. olearia*'s gene expression in the environments defined by a specific temperature.

The *K. olearia* genome (NC_012785) has 2,302,126 bp and is predicted to encode 2,224 genes (Swithers et al., 2011). Within the Thermotogae, genome size, intergenic region size, and number of predicted coding regions correlate with the optimal growth temperature of an isolate (Zhaxybayeva et al., 2012), with hyperthermophilic Thermotogae genomes being the most compact. Phylogenetically, the Thermotogae order Kosmotogales comprises the genera *Kosmotoga* and *Mesotoga* spp., the latter being the only described mesophilic Thermotogae lineage (Pollo et al., 2015). Assuming a hyperthermophilic last common ancestor of the Thermotogae (Zhaxybayeva et al., 2009), the Kosmotogales can be hypothesized to have acquired wide growth temperature

tolerance secondarily by expanding its gene repertoire. Moreover, it is likely that the ability of the Kosmotogales common ancestor to grow at low temperatures enabled the evolution of mesophily in *Mesotoga* (Pollo et al., 2015).

Such adaptations of lineages to new environments can be greatly facilitated by lateral gene transfer (LGT), since genes already "adapted" to the new conditions are readily available in the microbial communities of the new environment (Boucher et al., 2003). For instance, LGT has been implicated in adaptation to high temperature growth in hyperthermophilic bacteria, including *Thermotoga* spp., and to low temperature growth in archaea (López-García et al., 2015; Pollo et al., 2015; Boucher et al., 2003). Genome analysis of the mesophilic *Mesotoga prima* revealed that it laterally acquired 32% of its genes after it diverged from other Thermotogae lineages (Zhaxybayeva et al., 2012). Many of the predicted gene donors are mesophiles, supporting the importance of lateral acquisition of genes already adapted to mesophilic conditions in the evolution of *Mesotoga*.

To further gain insights into mechanisms of bacterial temperature response we sequenced 19 transcriptomes from isothermal and temperature-shifted cultures of *K. olearia* and examined transcriptional differences at temperatures spanning its wide growth range. Additionally, through comparative genomic and phylogenetic analyses of identified temperature responsive genes and their homologs in two newly sequenced *Kosmotoga* isolates, as well as in genomes of other thermophilic and mesophilic Thermotogae, we investigated the importance of within-lineage evolution through LGT and gene duplication for adaptation of *K. olearia* to growth over a wide temperature range.

2 Material and Methods

2.1 Bacterial culturing

K. olearia TBF 19.5.1 (DSM 21960(T), ATCC BAA-1733(T), Genbank accession number NC_012785) was grown at different temperatures (4°C, 25°C, 30°C, 40°C, 65°C, and 77°C), but otherwise optimal conditions, in *Kosmotoga* medium (KTM) using pyruvate as growth substrate as described in (DiPippo et al., 2009). Cultures used as inocula were stored at 4°C, except those used for experiments at ≥77°C. Actively growing cultures at temperatures ≥77°C had to be used directly as inoculum because the cultures would not grow from inocula stored at either 4°C or room temperature (~22°C). Replicate cultures received the same volume of inoculum; however, variable inoculum volume was used at different temperatures (Table S1 in Dataset S1), as larger inoculum volumes were required to achieve growth at the non-optimal temperature treatments. Due to the very slow growth rate at 25°C and lack of growth at 4°C, cultures were first grown at 30°C, and then transferred into new medium (25 mL) as a 50% inoculum (25 mL) and incubated at 25°C or 4°C, respectively, for 24 h.

2.2 Measurement of *K. olearia* growth at different temperatures

Growth curves were constructed from optical density measurements at 600 nm (OD_{600}) using an Ultrospec 3100 pro. For cultures grown at 40°C, 65°C, 77°C, and 78°C two sets of triplicate bottles, inoculated from the same inoculum 12 h apart, were monitored for a 12 h period per day to generate the growth curves. The cultures for isothermic growth at 40°C, 65°C, 77°C, and 78°C were monitored hourly, while the cultures for shifted growth at these temperatures were monitored every 1-5 hours. At 30°C one set of triplicate bottles was monitored once daily. Isothermic growth curves were calculated from six replicates, except for the curves for 30°C and 77°C that had three and 12 replicates, respectively. All shifted growth curves consisted of six replicate cultures, except for 40°C and 30°C which had four and three replicates, respectively. To determine growth rates (Figure 1), for

each culture the $lnOD_{600}$ was plotted against growth time and the curve was fitted with a linear trend line. The growth rate was defined as the slope at the log phase. To determine the time span of each growth phase (Figure S1), full composite growth curves were constructed using pooled replicate data. For each curve, OD_{600} for all replicates was plotted against growth time and a polynomial regression trend line was fitted.

2.3 Cultivation of *Kosmotoga* sp. DU53 and *K. arenicorallina,* and confirmation of their growth temperature ranges

Kosmotoga sp. DU53 was grown in KTM as described above for *K. olearia*. *K. arenicorallina* was also grown in KTM; however, maltose was used as substrate (2.5 mL and 0.5 mL 10% maltose was added to serum bottles and Hungate tubes, respectively). One mL of culture was used as inoculum for all cultures (bottles and tubes). The temperature range of each strain was assessed by examination of culture turbidity as a proxy for growth. Starting from cultures grown at optimal growth temperature (~ 65°C for *Kosmotoga* sp. DU53 and 60°C for *K. arenicorallina*), new cultures were shifted in ≤10°C increments. If growth was observed after a shift, then that culture was used to initiate a new culture. The shifting procedure was terminated when growth was no longer evident at a given temperature.

2.4 RNA isolation and processing

For each temperature treatment, RNA was extracted in either mid-log phase or late-log phase, using the Zymo Research Fungal/Bacterial RNA MiniPrep Kit (Cedarlane Laboratories, Ltd.; Burlington, Ontario) and following the manufacturer's protocols (Table S1 in Dataset S1). For experiments at 65°C, 77°C, 40°C, and 30°C cultures used for RNA extraction were inoculated from cultures that had been grown under the same temperature conditions for at least three transfers. The time at which a culture was expected to be in a desired growth phase was determined from the composite growth curves (Figure S1), and was used as a cell harvesting time (listed in Table S1 in Dataset S1). This procedure avoided exposure of the cultures to the lower ambient temperatures in the laboratory during subsampling. For the 30°C cultures, OD_{600} was additionally measured 24 h before harvesting to ensure the culture was in mid log phase. In order to stabilize the transcripts and to avoid any transcriptional response to the change in temperature, an equal volume of "stop solution" (10% phenol in ethanol) was added to sealed cultures via syringe immediately upon removal from the incubator.

Following recommendations in (Haas et al., 2012), we aimed to sequence ~3 million non-ribosomal-RNA reads per sample. Ribosomal RNA (rRNA) depletion was performed on all samples using the Ribo-Zero rRNA Removal Kit (Gram-Positive Bacteria) Magnetic Kit (MRZGP126, Epicentre). 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. Quality and quantity of the total RNA, as well as efficiency of rRNA depletion, were assessed on an Agilent 2100 Bioanalyzer RNA Nano chip or RNA Pico chip following the manufacturer's instructions for "Prokaryote Total RNA". mRNA successfully depleted of rRNA was used to construct RNA-Seq libraries following the manufacturer's instructions, and sequenced on either an Ion Torrent PGM (RNA-Seq kit V2; transcriptomes are labeled with an "IT" suffix) or an Illumina MiSeq (TruSeq RNASeq v2 2x100 bp). The platform and RNA extraction technique used for each transcriptome are summarized in Table S1 in Dataset S1. The transcriptomes

are available in the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the accession number SRP075860.

2.5 Architecture of the *K. olearia* transcriptome

Transcription start and stop sites were identified using Rockhopper (McClure et al., 2013) on the Illumina transcriptomes (Table S1 in Dataset S1). The following settings were used: Strand specific, No; Test for differential expression, Yes; Reverse complement reads, No; Orientation of mate-pair reads, forward-reverse; Maximum bases between paired reads, 500 bp; Allowed mismatches, 0.15; Minimum seed length, 0.33; Identify transcript boundaries, Yes; Predict operons, Yes; Minimum expression of UTRs and ncRNAs, 0.5; Minimum reads mapping to a transcript, 20; Minimum transcript length, 50 bp; Minimum count to seed a transcript, 50; and Minimum count to extend a transcript, 5.

2.6 RNA-Seq analysis

For each transcriptome**,** sequenced reads were analyzed using the "RNA-Seq" function in CLC Genomics Workbench version 7.0.4. Briefly, reads were first trimmed using an ambiguous limit of 2 and a quality limit of 0.05. To remove sequences that matched remaining rRNA transcripts, the trimmed reads were subjected to a relaxed mapping protocol: only rRNA genes were used as a "reference genome", reads were mapped only to the reference sequences using similarity fraction of 0.5; length fraction was set to 0.5; all other parameters were set to default (maximum number of hits for a read = 10, map to both strands, mismatch cost = 2, insertion cost = 3, deletion cost = 3, auto-detect paired distances, color space alignment, color error cost = 3). The mapped reads were designated as rRNA and were removed from further analysis.

The remaining reads were subjected to an RNA-Seq protocol with strict mapping parameters (allowing mapping to intergenic regions, similarity fraction = 0.95; length fraction = 0.95; all other default settings as described above) using the *K. olearia* annotated genome as a reference. Unmapped reads were discarded. Expression levels for every gene were estimated using "Reads Per Kilobase of transcript per Million mapped reads" (RPKM) values.

RPKM values for all genes are listed in Table S4 in Dataset S1. Differentially expressed genes were identified by doing pairwise comparisons of Illumina transcriptomes of the isothermically grown cultures at 30°C, 40°C, and 77°C to the cultures grown at the optimal temperature of 65°C. The analyses used the "Empirical Analysis of DGE" function, which employs the "Exact Test" for two-group comparisons (Robinson and Smyth, 2008). A gene was considered differentially expressed in a pairwise comparison if it had (*i*) > 20 reads in at least one of the two transcriptomes, (*ii*) a statistically significant difference in the RPKM values (corrected for multiple testing using False Discovery Rate [FDR] < 0.05), and (*iii*) a difference in RPKM values at least two-fold in magnitude. Principal Component Analysis (PCA) and biplot visualization were performed using R packages *ade4* and *bpca* respectively (Dray et al., 2007; Faria et al., 2013). Each gene was assigned to a Clusters of Orthologous Groups (COG; (Galperin et al., 2015)) functional category using the Integrated Microbial Genomes (IMG) portal (Markowitz et al., 2014). Genes assigned to more than one COG category were counted in all assigned categories.

Pearson correlation of expression values and growth rates were calculated in Microsoft Excel. ANCOVA, linear regression and likelihood ratio tests were carried out in R. The "growth rate only" model was the linear regression (expression = growth rate). In the ANCOVA model growth rates was set as the quantitative variable and temperature as qualitative variable (expression = growth rate + temperature). The growth rates used were 0.006 for 30°C, 0.087 for 40°C, 0.274 for 65°C and 0.107

for 77°C (see Figure 1). When comparing the most significant temperature coefficient to the growth rate coefficient, the latter was scaled by the average growth rate. The growth rate effect was defined as being greater than the temperature rate effect if |temperature/(growth rate * 0.115)| < 1.

2.7 Identification of genes involved in growth on pyruvate

K. olearia genes predicted to be involved in pathways for pyruvate conversion to acetate, CO_2, H_2 and ATP were retrieved from the KEGG (Kanehisa et al., 2014) and BioCyc (Caspi et al., 2014) databases. Genes encoding hydrogenases were taken from (Schut et al., 2012), and genes encoding the F-type ATPase subunits were identified using IMG (Markowitz et al., 2014).

2.8 Fatty acids analysis

Total lipids were extracted from *K. olearia* grown at 40°C to early stationary phase and 65°C to mid-log, early stationary, late stationary and death phase by using methanol-chloroform (1:1 v/v). Fatty acid methyl esters (FAME) were prepared from total lipids extracts using mild alkaline methanolysis (Guckert et al., 1985). Dried FAME samples were re-dissolved in 300 µl chloroform (HPLC grade, Fisher Scientific) and analyzed by gas chromatography with mass spectrometry (GC-MS) on an Agilent 6890N gas chromatograph with a model 5973 inert mass selective detector (Agilent) fitted with an Agilent HP-5MS capillary column (30 m × 0.25 mm ID, 0.25 µm film thickness; J + W Scientific). Helium was used as the carrier gas with a temperature program of 130°C increasing at 3°C min^{-1} to 290°C and held for 2 min. Sample peaks were identified by comparison to Bacterial Acid Methyl Ester Mix standards (Supelco, Sigma Aldrich) or on the basis of mass spectra and expressed as a percentage of the total FAME quantified in each sample.

2.9 Comparative analyses of three *Kosmotoga* spp. genomes

The genome of *K. olearia* was compared to genomes of *Kosmotoga* sp. DU53 (accession number JFHK00000000) and *K. arenicorallina* (accession number JGCK00000000) (Pollo et al., 2016) using the IMG portal (Markowitz et al., 2014) and Geneious v.9. Specifically, genes were declared homologous if they were significantly similar in BLASTP and TBLASTN (Altschul et al., 1997) searches (E-value < 10^{-3}). For phylogenetic analyses, additional homologs of *K. olearia* genes were retrieved from the NCBI non-redundant (*nr*) protein database and the IMG databases via BLASTP searches, retaining the 100 top-scoring matches with E-value < 10^{-3}. Sequences were aligned using MAFFT (Katoh et al., 2002), and phylogenetic trees were reconstructed using RAxML (Stamatakis, 2006), as implemented in Geneious v. 9.1.3. Homologs from the recently released genome of *Kosmotoga pacifica* (NZ_CP011232) (L'Haridon et al., 2014) were included in gene-specific phylogenetic analyses. Pairwise Average Nucleotide Identity (ANI) (Goris et al., 2007) was calculated using the Enveomics Toolbox (Rodriguez-R and Konstantinidis, 2016).

2.10 Detection of laterally transferred genes

A customized version of HGTector (Zhu et al., 2014) (available through https://github.com/ecg-lab/hgtector) was used to identify putatively transferred protein-coding genes in the *K. olearia* genome. Homologs of each annotated protein-coding open reading frame (ORF) in the NC_012785 GenBank record were retrieved from a local copy of the NCBI *nr* database (downloaded November 21, 2014) using the BLASTP program from BLAST 2.2.28+ (Altschul et al., 1997). Sequences were first filtered for low complexity regions using the *seg* algorithm. Then, only matches with E-value

$<10^{-5}$ and sequence coverage ≥ 70% were retained. Database matches were expanded according to the *MultispeciesAutonomousProtein2taxname* file from RefSeq release 68. This was necessary as some genes across various taxonomic ranks were combined into a single entry in RefSeq, which artificially decreased the representation of these genes in Close and Distal groups (see below), and would confound downstream analysis. Taxonomic affiliation of each match was assigned using the NCBI Taxonomy database (downloaded on November 21, 2014). Only 500 top-scoring matches (after filtering for sequence coverage) were used as input for HGTector. The "Self" group was defined as TaxID 651456 (genus *Kosmotoga*), and the "Close" group was defined as either TaxID 1184396 (genus *Mesotoga*, a sister group) or TaxID 2419 (order Thermotogales, comprising *Thermotoga*, *Mesotoga*, and *Kosmotoga*). In either case, the “Distal” group comprised the remaining taxonomic groups. The conservative cutoff (the median between the zero peak and the first local minimum) was used for both the “Close” and “Distal” groups. A gene was designated as putatively transferred if its "Close" score was below the cutoff and its "Distal" score was above the cutoff. Putatively transferred genes with no top-scoring match in Thermotogae were designated as recent transfer events into *K. olearia* (labelled "K" in Table S8 in Dataset S1). Putatively transferred genes for which the difference between Close(Thermotoga) and Close(Mesotoga) scores was <1 were designated as gene transfer events into Kosmotogales (*i.e., Kosmotoga* and *Mesotoga*; labelled "K+M" in Table S8 in Dataset S1).

3 Results and Discussion

3.1 Temperature shifts and isothermic conditions elicit different growth patterns in *K. olearia*.

Under laboratory conditions in liquid anaerobic medium we observed growth of *K. olearia* at temperatures as low as 25°C and as high as 79°C, with optimal growth at 65°C, defined as the temperature affording the fastest growth rate (Figure 1 and Figure S1). Using a non-linear regression model (Ratkowsky et al., 1983) we estimate a growth-permissive temperature range of 20.2 – 79.3°C, consistent with the previously reported wide growth range of this isolate (DiPippo et al., 2009). Interestingly, we were not able to cultivate *K. olearia* at temperatures near its range boundaries (30°C and 77°C) by direct transfer from 65°C cultures. Instead, the growth temperature had to be changed sequentially in ≤10°C increments. Particularly at the extremes, even small temperature shifts caused both a longer lag phase and a slower growth rate compared to isothermal cultures (Figure 1 and Figure S1). This phenomenon has also been noted for mesophilic bacteria, especially for transitions from high to low temperature (Swinnen et al., 2004). Our observations suggest that cells shifted to a new temperature need to undergo large physiological changes that require time (i.e. an ‘acclimation’ period (Barria et al., 2013)) and that these physiological challenges are too great to overcome when temperature changes are large. To illuminate *K. olearia*'s transcriptional responses to changes in temperature we sequenced 19 transcriptomes of replicate mid- to late-log cultures grown isothermally at 30°C, 40°C, 65°C, and 77°C, and of two 30°C cultures shifted to 25°C and 4°C (see Table S1 in Dataset S1 and Text S1).

3.2 Architecture of the *K. olearia* transcriptome

Analysis of transcription start and stop sites predicted a minimum of 916 transcriptional units (TU) in *K. olearia* (Text S1 and Table S2 in Dataset S1), 52% of which consist of a single gene. This fraction of single-gene TUs lies between the 65% reported for *E. coli* (Cho et al., 2009) and the 43% recorded

for *T. maritima*, which has also been shown to have a streamlined genome and a low-complexity transcriptome (i.e., few sub-operonic transcripts and few genes with multiple start sites) (Latif et al., 2013). The average TU length of ~2.39 genes in *K. olearia* is less than the 3.3 genes per transcript of *T. maritima* (Latif et al., 2013) but closer to 2.2 genes per transcript in the mesophilic firmicute *Geobacter sulfurreducens* (Qiu et al., 2010) and 1-2 genes per transcript in bacteria in general (e.g. (Cho et al., 2009)). Given that the *K. olearia* genome has more intergenic DNA than *T. maritima*'s genome (the ratio of the nucleotides located in non-coding vs. coding regions is 0.13 in *K. olearia* and 0.06 in *T. maritima*), the shorter TU lengths in *K. olearia* may point to more flexible transcriptional regulation.

3.3 Consistent energy generation across different temperature conditions

K. olearia produces ATP from pyruvate using a biochemically well-understood fermentation pathway that generates hydrogen, carbon dioxide and acetate ((DiPippo et al., 2009); Figure 2 and data not shown). Given that pyruvate was the carbon and energy source provided in all experiments, we surveyed 51 genes predicted to be involved in core energy metabolism during growth on pyruvate. The model in Figure 2 accounts for all of the major end products during growth at 65°C and contains 15 genes with consistently high expression across all temperature treatments (Table S3 in Dataset S1). In addition to indirectly validating the previously known functional annotations of these genes, we furthermore propose that the most highly expressed ABC-transporter gene cluster, Kole_1509 – 1513, encodes a pyruvate transporter (Figure 2). Its current annotation as a peptide ABC transporter may be erroneous since most of the peptide ABC transporters predicted in *T. maritima* using bioinformatics have been shown instead to bind and transport sugars (Nanavati et al., 2006). However, functional studies of the transporter (e.g. binding assays, expression with different substrates) are needed to confirm this hypothesis.

3.4 Identification of temperature-related transcriptional responses in *K. olearia*

Based on hierarchical clustering, transcriptome replicates at the same temperature group together (Figure S2 and Text S1), suggesting that the observed changes in transcription are due to the culture growth temperature. Principal Component Analysis (PCA) clearly separated the transcriptomes into quadrants corresponding to optimal (65°C), intermediate (40°C), low (30°C, 25°C and 4°C) and high (77°C) growth temperatures (Figure 3). Several genes with a high correlation between their expression level and a specific growth temperature (vectors in Figure 3, Table S4 in Dataset S1) are known to be involved in temperature response (Pollo et al., 2015). For example, expression of the heat shock serine protease Kole_1599 positively correlated with the 77°C transcriptomes, where high expression of this protease was expected based on its involvement in heat shock response in *T. maritima* (Pysz et al., 2004). Similarly, expression of the cold shock protein genes Kole_0109 and Kole_2064 positively correlated with low temperature growth. Lastly, some observed changes presumably were due to the expected decreased metabolic activity of the culture at sub- and supra-optimal temperatures. This can be exemplified by the high expression and strong correlation of the central carbon metabolism gene glyceraldehyde-3-phosphate dehydrogenase (Kole_2020) with the 65°C transcriptomes.

Putative temperature-responsive genes were identified by pairwise comparisons of isothermic temperature treatments to the optimal condition of 65°C (i.e., 30°C vs 65°C, 40°C vs 65°C, and 77°C vs 65°C). To remove any confounding effects of differential sample handling, only the 8 Illumina transcriptomes from cultures processed in identical fashion were included in this analysis (see Table S1 in Dataset S1 and Text S1). Across all comparisons 573 genes fulfilled our criteria for temperature responsiveness (i.e., ≥ 2-fold difference in expression, > 20 reads per transcript, False Discovery Rate < 0.05) with 430, 115, and 169 genes detected in the 30°C vs 65°C, 40°C vs 65°C, and 77°C vs 65°C comparisons respectively (Table S5 in Dataset S1). It should be noted that although we label these genes as "temperature responsive", they may additionally respond to other similar environmental stressors (e.g., DNA damage, desiccation, starvation).

Batch culture cannot selectively discern gene expression that is exclusively influenced by temperature from expression that is solely growth rate-dependent. In theory, continuous culture conducted at a single growth rate could afford such discrimination. However, given the extremely slow growth of *K. olearia* near its temperature maxima and minima (Figure 1), it was not feasible to use anaerobic bioreactors to cultivate cells at this marginal growth rate across the temperature range. Hence, in order to assess how differences in growth rates influence the expression patterns, we examined correlations of the expression pattern of the putative temperature responsive genes with growth rates calculated at each temperature. Indeed, expression of 306 of the 573 putative temperature-responsive genes strongly correlated with growth rate (r> |0.7|; Table S5 in Dataset S1). To detect which of the 306 genes are significantly influenced by growth rate, we performed ANCOVA analyses with growth rate as the quantitative variable and temperatures as qualitative treatments. For 60 genes the growth rate was a significant factor ($p < 0.05$; Table S5 in Dataset 1). However, for 26 of the 60 genes ANCOVA suggested that temperature was also a significant factor and the likelihood ratio test (LRT) rejected the simpler growth-rate-only model for all but 13 of the 60 genes (FDR-corrected p-value < 0.05; Table S5 in Dataset S1). Comparison of the most significant temperature coefficient to the growth rate coefficient revealed that growth rate had larger influence on expression of 51 genes (Table S5 in Dataset S1). LRT rejected the growth-rate-only model for 10 of these 51 genes. Among these genes were the glyceraldehyde-3-phosphate dehydrogenase (Kole_2020) mentioned above, the iron-containing alcohol dehydrogenase (Kole_0742) discussed in Supplemental Material. Therefore, although we retained the designation of "putatively temperature-responsive" for all 573 genes, in the discussion below we focus on the genes not primarily affected by growth rate.

3.5 Detailed analysis of changes in gene expression in response to prolonged growth at different temperatures

Most of the temperature responsive genes were up-regulated compared to the expression at 65°C (Figure S3). However, many genes involved in carbohydrate and energy metabolism were down regulated at 30°C (Clusters of Orthologous Groups [COG] categories C and G), while genes from COG C and G categories were over-represented and up-regulated at 77°C (Figure 4 and Figure S3; discussed in detail below).

In all transcriptomes the list of putative temperature-responsive genes was diminished in genes involved in translation (COG category J) and nucleotide metabolism (COG category F) (Figure 4 and Figure S3) and conversely enriched in genes involved in replication, recombination and repair (COG category L, particularly at 30°C), and signal transduction (COG category T). Most of the

identified COG category L genes are either mobile elements or CRISPR-associated proteins. Movement of mobile genetic elements is a common feature of stress responses (Foster, 2007). Differential expression of the signal transduction genes suggests the importance of these systems for regulating cellular responses at all tested temperatures. Additionally, at both 30°C and 77°C many genes encoding transcription regulators (COG category K, transcription) are up-regulated, suggesting that prolonged growth at sub- and supra-optimal temperatures results in detectable changes in transcriptional gene regulation in *K. olearia*.

Below we discuss the identified temperature-responsive gene expression patterns in more detail, focusing on genes that did not show an effect from growth rate.

3.5.1 At 40°C there are pronounced differences in membrane fatty acid composition but no signs of cold stress

Although the growth rate of *K. olearia* at 40°C is only one-third of that at the optimum 65°C (Figure 1 and Figure S1), clustering analysis suggested that the 40°C transcriptome was most similar to that at 65°C (Figure 3 and Figure S2). The slower growth rate was reflected by the four most highly expressed temperature responsive genes at 40°C showing significantly lower expression than at 65°C, including the growth-rate dependent alcohol dehydrogenase (Kole_742, Table S5 in Dataset S1). Yet, 94 of 115 putative temperature responsive genes were up-regulated (Table S5 in Dataset S1), suggesting that slower metabolism is not the only explanation for the observed transcriptional response to growth at 40°C.

Lipid metabolism (COG category I) appears to be particularly important at 40°C. For instance, all of the predicted fatty acid synthesis genes showed the highest expression at 40°C (Table S5 in Dataset S1 and Figure S4), with two genes involved in synthesis of unsaturated fatty acids (Kole_0968) and initiation of fatty acid synthesis (Kole_0969) having significantly higher expression. Biochemical analyses of total fatty acids at 40°C and 65°C showed a much greater diversity of fatty acids at 40°C (Table S6 in Dataset S1), which may explain the higher demand for these genes at lower temperatures. Interestingly, at 40°C in particular, there was increased expression of a phosphate ABC transporter (Kole_0707 – Kole_0711, Table S5 in Dataset S1), which may be linked to increased production of polar membrane lipids at moderately low temperatures. Maintenance of a functional cell membrane is crucial for survival, and bacteria respond to changes in temperature by altering the membrane's fatty acid composition (de Mendoza, 2014). The observation that lipid metabolism genes were among the highly expressed genes at low temperature, despite the lower growth rate, suggests that changes to the cell membrane composition are one of the most important adaptations for survival of *K. olearia* at lower temperatures.

Proper protein folding at a lower temperature is another physiological challenge that may require enzymatic assistance. For example, proline isomerization happens spontaneously at high temperatures, but at lower temperatures (e.g., 37°C) the reaction needs a catalyzing enzyme - peptidylprolyl isomerase (PPIase) (Godin-Roulling et al., 2015). Not surprisingly, *K. olearia* has three temperature-responsive PPIase genes: two PpiC-type genes (Kole_1682 and Kole_0383) that are both highly expressed at 40°C, and one FKBP-type gene (Kole_1745) that shows high expression at all temperatures except 77°C (Table S5 in Dataset S1). These expression patterns suggest PPIase is particularly important at moderately low temperatures where the cells are still relatively active. However, the enzymes known to assist protein folding in cellular stress responses, chaperones (e.g., GroEL and Hsp) and protease Do, were significantly down-regulated at 40°C (Table S5 in Dataset S1), contributing to the enrichment of differentially expressed genes from COG category O (Post-translational modification, protein turnover, chaperone function) at 40°C in Figure 3. Among other

typical cold stress-related proteins, only one of *K. olearia*'s three cold shock proteins (Kole_0109) showed significantly higher expression at 40°C and its up-regulation was merely moderate when compared to its expression levels at 30°C (Table S5 in Dataset S1). This overall lack of induction of typical stress-related genes, especially when compared to 30°C and 77°C (see below), suggests that 40°C is still within the "Goldilocks" temperature range for *K. olearia.*

3.5.2 *K. olearia* is in cold stress at 30°C

Transcriptomes from 30°C, 25°C, and 4°C cultures were very similar to each other (Figure 3 and Figure S2). However, due to adjustments in culture handling required to obtain enough biomass at lower temperatures (see materials and methods), some gene expression patterns observed at 25°C and 4°C may be due to the cells either responding to fresh medium or displaying an immediate cold shock response. Therefore, we focused further analyses on genes differentially expressed at 30°C and used the 25°C and 4°C transcriptomes only to confirm the patterns observed at 30°C.

Two of three Csp-encoding genes in *K. olearia* (Kole_0109 and Kole_2064, Table S5 in Dataset S1) were among the three most highly expressed up-regulated genes at low temperatures, suggesting that the cells were in a cold-stressed state during growth at ≤30°C. Further support for this hypothesis comes from significant up-regulation of genes linked to bacterial cold response (Barria et al., 2013) including a DEAD/DEAH-box RNA helicase (Kole_0922), *rbfA* (Kole_2103) and *nusA* (Kole_1529). Hence, the thermophile *K. olearia* uses homologs of the cold response genes employed by mesophilic bacteria at physiologically low temperatures.

With decreasing temperature, we observed up-regulation of several ribosomal proteins (Figure 3). Some (L10 (Kole_1840) and L7/L12 (Kole_1839)) have already been linked to both cold shock and prolonged low temperature growth responses in bacteria (e.g., (Alreshidi et al., 2015)). The most dramatic differential expression, however, was observed for a ribosomal protein gene not yet connected to cold response (L34; Kole_0258). L34, a bacteria-specific ribosomal protein hypothesized to be a relatively recent addition to the evolving ribosome (Fox, 2010), is required for proper ribosome formation (Akanuma et al., 2014). A *Bacillus subtilis* mutant lacking the L34 gene showed particularly slow growth at low temperature (Akanuma et al., 2012), suggesting a role for L34 in this condition. Many ribosomal proteins are recruited for extra-ribosomal functions (Bhavsar et al., 2010), hence some of the up-regulated ribosomal proteins may have alternative roles in response to low temperature that are unrelated to the ribosome itself. However, genes encoding ribosomal RNA (rRNA) methyltransferases, *rmlH* (Kole_1718) and *rmlL* (Kole_0897), were also significantly up-regulated, and methylation of rRNAs has been associated with responses to environmental stress, including temperature (Baldridge and Contreras, 2014). Combined with observations that ribosomes need to be fine-tuned to function properly at low temperature (Barria et al., 2013), we hypothesize that *K. olearia* modifies its ribosome by changing stoichiometry of its components and by methylating rRNA. Time required for such ribosomal adjustments could also explain the longer lag phase following temperature shifts (Figure S1).

To detect a decrease in environmental temperature and elicit an appropriate regulatory response, some bacteria have evolved two-component cold sensors (de Mendoza, 2014). These signal transduction systems consist of a sensor, a membrane-integrated protein with a kinase domain that detects changes in the fluidity of the cell membrane, and the cytoplasmic response regulator, a protein that induces expression of cold-responsive genes. In *K. olearia*, a histidine kinase with two predicted transmembrane domains (Kole_1017) and two response regulators (Kole_1015 and Kole_1016) showed a steady increase in expression as temperatures decreased from 65°C, but no

significant change in expression at 77°C (Table S5 in Dataset S1), leading us to hypothesize that these genes encode a cold-sensing two-component system.

3.5.3 Increased amino acid metabolism at sub-optimal temperatures

At lower growth temperatures (and especially ≤30°C) we observed an over-representation of highly expressed genes involved in amino acid metabolism (COG category E). Also, at 30°C, and to a lesser extent at 40°C, several genes in the arginine (Kole_0092 – Kole_0097) and lysine (Kole_0104 – Kole_0107, 30°C only) biosynthesis pathways were up-regulated, suggesting the potential for accumulation of peptides and amino acids (or their intermediates) at lower temperatures. Although the inferred effect of temperature was larger than the effect of growth rate for all lysine biosynthesis genes (Table S5 in Dataset S1), the growth-rate-only model was rejected only for one of the genes, leaving the possibility that the expression of the lysine biosynthesis genes may be growth rate dependent. At 30°C there was also significant up-regulation of a citrate synthase gene (Kole_1230). Intriguingly, in *Staphylococcus aureus* citrate was shown to accumulate during prolonged cold stress (Alreshidi et al., 2015), which could also be the case for *K. olearia*. Alternatively, citrate synthase, together with isocitrate dehydrogenase (Kole_1227), may be involved in converting pyruvate or acetyl-CoA to 2-oxoglutarate, a precursor for several amino acids including arginine. Accumulation of both arginine and lysine was observed during low temperature growth of *Clostridium botulinum*, where these amino acids were suggested to act as compatible solutes (Dahlsten et al., 2014). Interestingly, while the cells may accumulate peptides at 30°C, at 40°C there was increased expression of an oligo-peptidase (Kole_1190) and genes involved in lysine degradation (Kole_0958, Kole_0963 – Kole_0966). Such distinguishably different metabolic responses to moderately low (40°C) and low (≤30°C) temperatures suggest a fine-tuned temperature-dependent peptide turnover. Two paralogs of ornithine carbamoyl-transferase genes (*argF*; Kole_1433 and Kole_2071) showed significantly lower expression at both 40°C and 30°C. The amino acid ornithine is an intermediate of arginine synthesis, and therefore lower expression of *argF* could result in ornithine, rather than arginine, accumulation. However, since the growth-rate-only model was not rejected for Kole_2071 (FDR-corrected p-value = 0.062) (Table S5 in Dataset S1), its high expression at 65°C could also be linked to the higher growth rate at this temperature.

Re-modelling of amino acid metabolism at low temperatures has also been observed in other bacteria (e.g., (Dahlsten et al., 2014; Ghobakhlou et al., 2015)). Interestingly, the genome of strictly mesophilic *M. prima* encodes more genes involved in amino acid metabolism than the genomes of thermophilic *K. olearia* and other Thermotogae (Zhaxybayeva et al., 2012). Amino acid metabolism genes are also among the most numerous bacterial genes laterally acquired by mesophilic archaea, which is hypothesized to reflect archaeal adaptation to low temperature growth (López-García et al., 2015).

3.5.4 *K. olearia* is in heat stress at 77°C

Both the multivariate (Figure 3) and clustering analyses (Figure S2) showed that the 65°C and 77°C transcriptomes are distinct. Since 77°C is near the upper limit for *K. olearia* growth under our laboratory conditions, we hypothesize that the observed differences in expression profiles at 77°C reflect a cell-wide heat stress response. Of the 169 differentially expressed genes, 119 showed increased expression at 77°C (Table S5 in Dataset S1). Hypothetical proteins comprise a sizeable fraction (41 genes; 34%) of the 119 genes, indicating that adaptation to growth at sustained high

temperature remains largely uncharacterized. These genes are scattered across the genome (Table S2 in Dataset S1) in 34 transcriptional units, each containing at most three of these genes. The majority of 41 genes have homologs in genomes of other *Kosmotoga* spp. (N=38), *Mesotoga* spp. (N=23), and other Thermotogae (N=26). Kole_0654 has two paralogs (61 and 88% amino acid identity, respectively) in the genome (Kole_0653 and Kole_0630), while Kole_0801 has an almost identical paralog (Kole_0788, 99% amino acid identity). Putative functions could be inferred for some of the encoded proteins. Seventeen of the 41 proteins have predicted signal peptides, suggesting they are membrane-associated, and five (Kole_0130, Kole_0445, Kole_0994, Kole_1991, Kole_2135) have similarity to domains that are either membrane- or cell-wall-associated. Moreover, Kole_1430 and Kole_1431 are co-transcribed with genes from a two-component system (Kole_1428 and Kole_1429), suggesting they may be involved in environmental sensing or signaling. Kole_0652 carries a PrcB-domain, which interacts with and stabilizes PrtP protease (Godovikova et al., 2010). Kole_1314 (and its paralog Kole_1297) contains an AbiEii-toxin domain, and may be part of a toxin-antitoxin system. Three of the co-transcribed hypothetical genes (Kole_1266, Kole_1267, Kole_1270) are found in a cluster containing CRISPR-genes and two of them (Kole_1266 and Kole_1270) contain RAMP-domains suggesting CRISPR-related function (Makarova et al., 2011).

Only two of the known heat shock response genes (Pysz et al., 2004), the extreme heat stress sigma factor-24 (*rpoE,* Kole_2150) and the heat shock protease (Kole_1599), were up-regulated. Among the most highly expressed genes were those encoding the structural RNAs *ffs* (Kole_R0010), *ssrA* (Kole_R0006), and *rnpB* (Kole_R0049) (Figure 3), suggesting an increased rate of RNA turnover at supra-optimal temperature. As discussed above, carbohydrate and energy metabolism genes (COG categories C and G) were also up-regulated. It is unclear, however, if the underlying cause is the increased turnover of enzymes at elevated temperatures, or a demand for more enzymes due to increased carbohydrate catabolism. Increased carbohydrate metabolism in response to prolonged growth at supra-optimal temperature has been observed previously in *T. maritima* (Wang et al., 2012) and therefore may be a common adaptation to high temperature growth in the Thermotogae. As for *K. olearia,* the prolonged supra-optimal temperature growth of *T. maritima* also did not involve up-regulation of typical heat-shock response proteins (Wang et al., 2012). This highlights the difference between cellular response to an immediate heat-shock and to prolonged growth at supra-optimal temperature (Balleza et al., 2009), and in general justifies classifying the cellular response to temperature into these two distinct categories.

3.6 Global regulators of temperature response

The transcriptional changes seen at the sub- and supra-optimal temperatures are likely to be controlled by one or a few global regulators (Balleza et al., 2009), and the analysis of COG functional categories revealed enrichment of genes involved in transcription regulation at both 30°C and 77°C (Figure 4). The two-component cold sensor up-regulated at low temperatures (Kole_1015 - Kole_1017) may represent one such global regulator, as observed for similar systems in cyanobacteria (de Mendoza, 2014). Another group of global regulators are sigma factors needed for transcription initiation (Buck et al., 2000). Interestingly, one sigma factor was significantly up-regulated at 77°C (Kole_2150), while another was up-regulated at 30°C and 40°C (Kole_1408) suggesting these may represent temperature-specific sigma factors. Kole_2150 belongs to the sigma-24 ECF subfamily, which is activated in response to environmental stress (Balleza et al., 2009). Kole_1408 belongs to the sigma-54 family, which is involved in enhancer-dependent transcription (Buck et al., 2000), introducing the possibility that this sigma factor may be a target of the two

component cold sensor. Three genes annotated as anti-sigma regulatory factor and anti-sigma factor antagonists (Kole_0441 – Kole_0443) were also up-regulated at 30°C, and are probably involved in sigma factor regulation. Interestingly, Kole_0440, encoding a response regulator which is co-transcribed with these genes (Table S2 in Dataset S1), is also significantly up-regulated at 30°C.

In addition, 23 of the 573 putative temperature-responsive genes are annotated as transcriptional regulators. Of them, 15 were significantly up-regulated at 30°C, one 40°C, and eight at 77°C. Interestingly, only one transcription regulator (Kole_0294) was significantly down regulated (at 30°C). Experiments are needed to determine how these putative regulators activate or repress transcription of their target genes.

3.7 Conservation of *K. olearia*'s temperature-responsive genes across Kosmotogales

All genes that are required for adaptation and response of *K. olearia* strain TBF 19.5.1 to a wide range of growth temperatures are expected to be present in other *K. olearia* isolates, whereas some may be absent from *Kosmotoga* species having a narrower spectrum of growth temperature. Therefore, we compared the *K. olearia* genome to the genomes of *Kosmotoga* sp. DU53 and *Kosmotoga arenicorallina* (Pollo et al., 2016). *Kosmotoga* sp. DU53 has a similar growth temperature range (observed range 25°C - 79°C, Table S7 in Dataset S1) and >99% average nucleotide identity (ANI) when compared to *K. olearia*, while *K. arenicorallina* exhibits a narrower growth temperature range (observed range 35°C - 70°C, Table S7 in Dataset S1) and has only 84% ANI when compared to *K. olearia*.

Indeed, the *Kosmotoga* sp. DU53 genome lacks only 10 of the 573 *K. olearia* putative temperature-responsive genes (BLASTP and TBLASTN searches, E-value < 10^{-3}, Table S5 in Dataset S1). All 10 genes were expressed in *K. olearia* at relatively low levels (the highest average expression value of 453 is for Kole_0200 at 77°C), suggesting that they are unlikely to be essential for high or low temperature growth. On the other hand, the *K. arenicorallina* genome does not have detectable homologs of 103 of the 573 putative temperature-responsive genes in *K. olearia* (BLASTP and TBLASTN searches, E-value < 10^{-3}) (Table S5 in Dataset S1). The list of absent genes includes several of the arginine and lysine biosynthesis genes that are up-regulated in *K. olearia* during growth at ≤30°C, and seven of the hypothetical proteins up-regulated at 77°C. Therefore, we hypothesize that a subset of these 103 genes may play a role in extending the growth range of *K. olearia* to ≤35°C and ≥70°C. However, directed experiments are needed to pinpoint the genes involved.

3.8 Role of lateral gene transfer in thermoadaptation of *K. olearia*

Obtaining "pre-adapted" genes from other genomes is one way prokaryotes adjust to new environmental conditions (Boucher et al., 2003). Using HGTector (Zhu et al., 2014) we predicted that 354 of *K. olearia*'s 2,118 protein coding genes have been acquired laterally by *K. olearia* or the Kosmotogales (i.e., *Kosmotoga* and *Mesotoga*), presumably representing LGT events occurring after the divergence of Kosmotogales from other Thermotogae (Table S8 in Dataset S1). Eighty-eight of the 354 genes were temperature responsive (Table S5 in Dataset S1, Figure S5A and S5B), including several previously discussed highly expressed genes (Table 1 and Table S5 in Dataset S1). Notably, 76% of the 37 presumptively transferred temperature responsive genes that are shared with *Mesotoga*

are upregulated at 30°C (Figure S5C), suggesting that acquisition of these genes has been important in adaptation to low temperature growth. Among these are the previously discussed rRNA methyltransferase genes (Kole_1718 and Kole_0897). The fatty acid synthesis genes (Kole_0969 - Kole_0973) that are up-regulated at 40°C, as well as their Kosmotogales homologs, form a distantly related sister clade to other Thermotogae lineages (Figure S6A), suggesting that these genes may have been acquired from an un-sampled lineage. Similarly, the Csp-encoding gene highly expressed at 30°C (Kole_0109) is placed outside of the Thermotogae clade (Figure S6B). Predicted acquisition of the fatty acid synthesis and *csp* genes by (now mesophilic) archaea (López-García et al., 2015) additionally argues for the importance of these genes in adaptation to low temperature growth.

It is notable that some putative lateral gene acquisitions by *K. olearia* do not have homologs in *Mesotoga*. These include genes encoding the predicted cold temperature sensor (Kole_1015 – Kole_1017), one of the PPIase genes (Kole_1745), as well as the canonical cold response enzyme DEAD/DEAH box RNA helicase (Kole_0922). Absence of these genes in *Mesotoga* suggests their potential importance for *K. olearia*'s ability to grow over a wide temperature range.

3.9 Role of gene family expansion and lineage-specific gene evolution in thermoadaptation

Expansion of cold-responsive gene families may represent a common strategy for low temperature adaptation, as has been noted in many bacteria, especially in psychrophiles (e.g. Piette et al., 2010). *K. olearia* exhibits the same trend. For example, when compared to other Thermotogae, all three analyzed *Kosmotoga* genomes harboured more copies of Csp-encoding genes (Table S9 in Dataset S1). Additionally, *K. olearia* has extra homologs (Kole_0111 and Kole_0110) of the putative cold sensor system discussed above. The observed gene family expansions might be important not only for low temperature growth, but also for growth over a wide temperature interval. For example, *Mesotoga* functions with only a single *csp* gene, demonstrating that having more copies of this gene is not required for low temperature growth. Having several versions of these genes could, however, have made differential regulation easier. Similarly, the additional homologs of the two-component cold sensor genes do not show coordinated temperature responses: Kole_0110 is up-regulated at 40°C, while Kole_0111 is up-regulated at 77°C (Table 1). Therefore, these additional homologs may encode sensors tuned to different temperatures.

Gene family expansions can be achieved via within-lineage gene duplication or through LGT. A combination of these mechanisms appears to be at work in *K. olearia*, as demonstrated by the phylogenetic analyses of *csp* genes (Figure S6B). Similarly, even though several Thermotogae genomes contain as many copies of PPIase genes as do *Kosmotoga* genomes (Table S9 in Dataset S1), phylogenetic analysis suggests that in the Kosmotogales this gene family has only recently been expanded by both LGT (the FKBP-type, Table 1) and duplication (the PpiC-type, Figure S6C).

However, the role of within-lineage evolution of specific genes in response to changing environmental conditions should not be neglected. For example, typical cold response genes *rbfA* (Kole_2103) and *nusA* (Kole_1529) were not laterally acquired, but nevertheless show high expression only at 30°C. Deciphering adaptive changes that occurred in such genes compared to thermophilic homologs may elucidate molecular mechanisms of low temperature adaptation.

3.10 Why maintain the capacity for growth over such a wide temperature range?

Most bacteria are under selection to eradicate extraneous DNA (and genes) from their genomes (Graur, 2016), and among free-living bacteria Thermotogae in general have very compact genomes.

Kosmotogales, however, have notably larger genomes than other thermophilic Thermotogae (Pollo et al., 2015; Zhaxybayeva et al., 2012), raising the possibility that expanded genomes are advantageous in *K. olearia*'s habitat. As discussed above, many of the genes in *K. olearia*, such as the cold-sensor system, were expressed only at specific sub- or supra-optimal temperatures, but do not seem to be important for growth at other temperatures (Table 1 and Table S5 in Dataset S1). The regulated response to low temperatures and the preservation of the laterally acquired genes specifically expressed at 40°C and 30°C suggest that *K. olearia* encounters environments with very different temperatures frequently enough to maintain these genes in its genome. Such environments may include oil reservoirs located at different depths, as well as marine sediments influenced by the mixing of cold deep-sea water and hydrothermal fluids (Sievert and Vetriani, 2012). Perhaps, *K. olearia* migrates between such locations via subsurface fluids and, as a result, may have been selected to become a temperature generalist. This conjecture is supported by the environmental conditions of the subsurface environments and marine hydrothermal vents from which *Kosmotoga* spp. have been isolated (DiPippo et al., 2009; Nunoura et al., 2010; L'Haridon et al., 2014). *K. olearia* originated from a deep subsurface oil reservoir with *in situ* temperature of 68°C (DiPippo et al., 2009), but its 16S rRNA sequences have also been detected in many oil fields having *in situ* temperatures of 20°C–50°C (Nesbø et al., 2010). *Kosmotoga* sp. DU53, which is most similar to *K. olearia*, was isolated from an oil reservoir with an *in situ* temperature of ~50°C, while *K. arenicorallina* was obtained from hydrothermal sediments with a temperature of ~40°C (Nunoura et al., 2010). Notably, *K. olearia* was also identified as a major constituent of a metagenome from a deep subsurface oil reservoir with *in situ* temperature of 85°C and pressure of 25MPa (Kotlar et al., 2011). While the reservoir temperature is higher than the maximum *K. olearia* growth temperature reported here, elevated pressure could extend *K. olearia*'s temperature maximum, as has been demonstrated for some archaea (e.g. (Takai et al., 2008)). Therefore, *K. olearia*'s growth temperature range under natural conditions may be even broader than 20-79°C.

4 Conclusions.

The present study demonstrates that even bacteria with relatively small genomes can use transcriptional changes to respond effectively to large changes in temperature. We showed that *K. olearia*'s response to sustained exposure to a non-optimal temperature includes up-regulation of hundreds of genes. Several key genes with known temperature-related functions apparently have been acquired laterally, suggesting that LGT is an evolutionarily successful strategy for expansion of temperature tolerance. However, gene duplication and subsequent sub-functionalization of the paralogs likely also play an important adaptive role.

The ability of *K. olearia* to inhabit both high and low temperature environments suggests that members of this lineage encounter environments with large temperature fluctuations and/or migrate across ecological niches within the deep biosphere (e.g., between deep and shallow subsurface oil reservoirs). Therefore, the subsurface environments, as well as their microbial populations, might be viewed as a connected archipelago instead of isolated islands. As a corollary, we speculate that *K. olearia*-like ecological generalists could also facilitate LGT among seemingly isolated deep biosphere microbial communities adapted to a narrower ecological niche. For example, we have previously demonstrated high levels of gene flow among hyperthermophilic *Thermotoga* populations in subsurface oil reservoirs and marine hydrothermal vents (Nesbø et al., 2015), environments that are separated by non-thermophilic surroundings but are hydrologically linked. The mechanism of such gene flow is not yet known, but *K. olearia*-like Thermotogae capable of growing both in subsurface oil reservoirs and adjacent marine sediments could serve as mediators of gene exchange.

Although some of the identified 573 temperature-responsive genes are already known to be expressed in bacteria and archaea grown at high or low temperatures, most of the up-regulated genes have not previously been implicated in temperature response and are in need of better functional and biochemical characterization. For example, the majority of the *K. olearia* genes responsive to elevated temperature encode proteins of unknown functions. Versatile proteins that work across a broad range of temperatures also warrant further biochemical and evolutionary analyses, as understanding of their enzymatic flexibility can aid the design of commercially important thermostable proteins. Finally, other regulatory mechanisms (e.g. DNA methylation, post-transcriptional modifications) not studied here may reveal additional proteins important in *K. olearia*'s temperature responses, and should be targeted in future studies.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

SP, JF, OZ, CN designed research. SP, JF and CN managed the project. SP performed the RNA-Seq experiments and data analyses. AA performed the fatty acid analyses and interpreted the data. TS carried out the HGTector analyses. CN and SP did the comparative genomic analyses. CN did the phylogenetic analyses and ANCOVA analyses of the RNA-Seq data. SP, OZ, JF and CN interpreted all data. SP, OZ, and CN wrote the paper. All authors reviewed, revised and approved the final manuscript.

7 Funding

This work was supported by a Research council of Norway award (project no. 180444/V40) to C.L.N.; a Simons Investigator award from the Simons Foundation, a Dartmouth College Walter and Constance Burke Research Initiation Award, and Dartmouth College start-up funds to O.Z.; and an NSERC Alexander Graham Bell Canada Graduate Scholarship CGS-M to S.M.J.P. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

8 Acknowledgments

We thank Rhianna Charchuk for technical assistance, Dr. Amanda J. Lohan for advice and discussion of RNA-Seq methodology, Dr. Karen Budwill for head-space gas analyses, and Jennifer Franks for initial discussions of lateral gene transfer detection methods.

9 References

Akanuma, G., Nanamiya, H., Natori, Y., Yano, K., Suzuki, S., Omata, S., et al. (2012). Inactivation of ribosomal protein genes in *Bacillus subtilis* reveals importance of each ribosomal protein for cell proliferation and cell differentiation. *J. Bacteriol.* 194, 6282-6291. doi: 10.1128/JB.01544-12.

Akanuma, G., Kobayashi, A., Suzuki, S., Kawamura, F., Shiwa, Y., Watanabe, S., et al. (2014). Defect in the formation of 70S ribosomes caused by lack of ribosomal protein L34 can be suppressed by magnesium. *J. Bacteriol.* 196, 3820-3830. doi: 10.1128/JB.01896-14.

Alreshidi, M. M., Dunstan, R. H., Macdonald, M. M., Smith, N. D., Gottfries, J., and Roberts, T. K. (2015). Metabolomic and proteomic responses of *Staphylococcus aureus* to prolonged cold stress 121, 44-55. doi: 10.1016/j.jprot.2015.03.010.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402. doi: 10.1093/nar/25.17.3389.

Baldridge, K. C., and Contreras, L. M. (2014). Functional implications of ribosomal RNA methylation in response to environmental stress. *Crit. Rev. Biochem. Mol. Biol.* 49, 69-89. doi: 10.3109/10409238.2013.859229.

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., et al. (2009). Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol. Rev.* 33, 133-151. doi: 10.1111/j.1574-6976.2008.00145.x.

Barria, C., Malecki, M., and Arraiano, C. M. (2013). Bacterial adaptation to cold. *Microbiology* 159, 2437-2443. doi: 10.1099/mic.0.052209-0.

Bhavsar, R. B., Makley, L. N., and Tsonis, P. A. (2010). The other lives of ribosomal proteins 4, 327-344. doi: 10.1186/1479-7364-4-5-327.

Boonyaratanakornkit, B. B., Simpson, A. J., Whitehead, T. A., Fraser, C. M., El-Sayed, N. M. A., and Clark, D. S. (2005). Transcriptional profiling of the hyperthermophilic methanarchaeon *Methanococcus jannaschii* in response to lethal heat and non-lethal cold shock. *Environ. Microbiol.* 7, 789-797. doi: 10.1111/j.1462-2920.2005.00751.x.

Boucher, Y., Douady, C. J., Papke, R. T., Walsh, D. A., Boudreau, M. E. R., Nesbø, C. L., et al. (2003). Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* 37, 283-328. doi: 10.1146/annurev.genet.37.050503.084247.

Buck, M., Gallegos, M., Studholme, D. J., Guo, Y., and Gralla, J. D. (2000). The bacterial enhancer-dependent σ^{54} (σ^{N}) transcription factor
. *J. Bacteriol.* 182, 4129-4136. doi: 10.1128/JB.182.15.4129-4136.2000.Updated.

772 Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., et al. (2014). The
773 MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of
774 pathway/genome databases. *Nucleic Acids Res.* 42, D459-D471. doi: 10.1093/nar/gkt1103.

775 Cho, B., Zengler, K., Qiu, Y., Park, Y. S., Knight, E. M., Barrett, C. L., et al. (2009). The
776 transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.* 27, 1043-1049. doi:
777 10.1038/nbt.1582.

778 Dahlsten, E., Isokallio, M., Somervuo, P., Lindström, M., and Korkeala, H. (2014). Transcriptomic
779 analysis of (group I) *Clostridium botulinum* ATCC 3502 cold shock response 9, e89958. doi:
780 10.1371/journal.pone.0089958.

781 de Mendoza, D. (2014). Temperature sensing by membranes. *Annu. Rev. Microbiol.* 68, 101-116.
782 doi: 10.1146/annurev-micro-091313-103612.

783 Dinsdale, A. E., Halket, G., Coorevits, A., van Landschoot, A., Busse, H. J., de Vos, P., et al. (2011).
784 Emended descriptions of *Geobacillus thermoleovorans* and *Geobacillus thermocatenulatus. Int. J.*
785 *Syst. Evol. Microbiol.* 61, 1802-1810. doi: 10.1099/ijs.0.025445-0.

786 DiPippo, J. L., Nesbø, C. L., Dahle, H., Doolittle, W. F., Birkland, N., and Noll, K. M. (2009).
787 *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil
788 production fluid. *Int. J. Syst. Evol. Microbiol.* 59, 2991-3000. doi: 10.1099/ijs.0.008045-0.

789 Dray, S., Dufour, A. B., and Chessel, D. (2007). The ade4 Package - II : Two-table and K-table
790 methods 7, 47-52. doi: 10.1159/000323281.

791 Faria, J. C., Demétrio, C. G. B., and Allaman, I. B. (2013). *bpca: Biplot of multivariate data based*
792 *on principal components analysis*. Ilheus, Bahia, Brasil and Piracicaba, Sao Paulo, Brasil: UESC and
793 ESALQ.

794 Foster, P. L. (2007). Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol.* 42, 373-
795 397. doi: 10.1080/10409230701648494.Stress-Induced.

796 Fox, G. E. (2010). "Origin and evolution of the ribosome," in *The Origins of Life*, eds. D. Deamer,
797 and J. W. Szostak: Cold Spring Harb Perspect Biol), 2:a003483.

798 Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial
799 genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.*
800 43, D261-D269. doi: 10.1093/nar/gku1223.

801 Ghobakhlou, A., Johnston, A., Harris, L., Antoun, H., and Laberge, S. (2015). Microarray
802 transcriptional profiling of Arctic *Mesorhizobium* strain N33 at low temperature provides insights
803 into cold adaption strategies. *BMC Genomics* 16, 383-396. doi: 10.1186/s12864-015-1611-4.

804 Godin-Roulling, A., Schmidpeter, P. A. M., Schmid, F. X., and Feller, G. (2015). Functional
805 adaptations of the bacterial chaperone trigger factor to extreme environmental temperatures. *Environ.*
806 *Microbiol.* 17, 2407-2420. doi: 10.1111/1462-2920.12707.

807 Godovikova, V., Wang, H. T., Goetting-Minesky, M. P., Ning, Y., Capone, R. F., Slater, C. K., et al.
808 (2010). *Treponema denticola* PrcB is required for expression and activity of the PrcA-PrtP
809 (dentilisin) complex. *J. Bacteriol.* 192, 3337-3344. doi: 10.1128/JB.00274-10.

810 Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M.
811 (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence
812 similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81-91. doi: 10.1099/ijs.0.64483-0.

813 Graur, D. (2016). "Prokaryotic Genome Evolution," in *Molecular and Genome Evolution*Anonymous
814 (Sunderland, MA: Sinauer Associates), 451.

815 Guckert, J. B., Antworth, C. P., Nichols, P. D., and White, D. C. (1985). Phospholipid, ester-linked
816 fatty acid profiles as reproducible assays for changes in prokaryotic community structure of estuarine
817 sediments. *FEMS Microbiol. Lett.* 31, 147-158. doi: 10.1016/0378-1097(85)90016-3.

818 Haas, B. J., Chin, M., Nusbaum, C., Birren, B. W., and Livny, J. (2012). How deep is deep enough
819 for RNA-Seq profiling of bacterial transcriptomes?. *BMC Genomics* 13, 734-745. doi: 10.1186/1471-
820 2164-13-734.

821 Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data,
822 information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199-
823 D205. doi: 10.1093/nar/gkt1076.

824 Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple
825 sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059-3066. doi:
826 10.1093/nar/gkf436.

827 Kotlar, H. K., Lewin, A., Johansen, J., Throne-Holst, M., Haverkamp, T., Markussen, S., et al.
828 (2011). High coverage sequencing of DNA from microorganisms living in an oil reservoir 2.5
829 kilometres subsurface 3, 674-681. doi: 10.1111/j.1758-2229.2011.00279.x.

830 Latif, H., Lerman, J. A., Portnoy, V. A., Tarasova, Y., Nagarajan, H., Schrimpe-Rutledge, A. C., et
831 al. (2013). The genome organization of *Thermotoga maritima* reflects its lifestyle 9, e1003485. doi:
832 10.1371/journal.pgen.1003485.

833 L'Haridon, S., Jiang, L., Alain, K., Chalopin, M., Rouxel, O., Beauverger, M., et al. (2014).
834 *Kosmotoga pacifica* sp. nov., a thermophilic chemoorganoheterotrophic bacterium isolated from an
835 East Pacific hydrothermal sediment. *Extremophiles* 18, 81-88. doi: 10.1007/s00792-013-0596-7.

836 López-García, P., Zivanovic, Y., Deschamps, P., and Moreira, D. (2015). Bacterial gene import and
837 mesophilic adaptation in archaea 13, 447-456. doi: 10.1038/nrmicro3485.

838 Makarova, K. S., Aravind, L., Wolf, Y. I., and Koonin, E. V. (2011). Unification of Cas protein
839 families and a simple scenario for the origin and evolution of CRISPR-Cas systems. 6, 38-65. doi:
840 10.1186/1745-6150-6-38.

841 Markowitz, V. M., Chen, I. A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., et al. (2014). IMG 4
842 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* 42,
843 D560-7. doi: 10.1093/nar/gkt963.

844 McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. a., et al. (2013).
845 Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res.* 41, e140. doi:
846 10.1093/nar/gkt444.

847 Mega, R., Manzoku, M., Shinkai, A., Nakagawa, N., Kuramitsu, S., and Masui, R. (2010). Very rapid
848 induction of a cold shock protein by temperature downshift in *Thermus thermophilus. Biochem.*
849 *Biophys. Res. Commun.* 399, 336-340. doi: 10.1016/j.bbrc.2010.07.065.

850 Mykytczuk, N. C. S., Foote, S. J., Omelon, C. R., Southam, G., Greer, C. W., and Whyte, L. G.
851 (2013). Bacterial growth at -15°C; molecular insights from the permafrost bacterium *Planococcus*
852 *halocryophilus* Or1 7, 1211-1226. doi: 10.1038/ismej.2013.8.

853 Nanavati, D. M., Thirangoon, K., and Noll, K. M. (2006). Several archaeal homologs of putative
854 oligopeptide-binding proteins encoded by *Thermotoga maritima* bind sugars. *Appl. Environ.*
855 *Microbiol.* 72, 1336-1345. doi: 10.1128/AEM.72.2.1336-1345.2006.

856 Nesbø, C. L., Kumaraswamy, R., Dlutek, M., Doolittle, W. F., and Foght, J. (2010). Searching for
857 mesophilic *Thermotogales* bacteria: "Mesotogas" in the wild. *Appl. Environ. Microbiol.* 76, 4896-
858 4900. doi: 10.1128/AEM.02846-09.

859 Nesbø, C. L., Swithers, K. S., Dahle, H., Haverkamp, T. H., Birkeland, N., Sokolova, T., et al.
860 (2015). Evidence for extensive gene flow and *Thermotoga* subpopulations in subsurface and marine
861 environments 9, 1532-1542. doi: 10.1038/ismej.2014.238.

862 Nunoura, T., Hirai, M., Imachi, H., Miyazaki, M., Makita, H., Hirayama, H., et al. (2010).
863 *Kosmotoga arenicorallina* sp. nov. a thermophilic and obligately anaerobic heterotroph isolated from
864 a shallow hydrothermal system occurring within a coral reef, southern part of the Yaeyama
865 Archipelago, Japan, reclassification of *Thermococcoides shengliensis* as *Kosmotoga*
866 *shengliensis* comb. nov., and emended description of the genus *Kosmotoga. Arch. Microbiol.* 192,
867 811-819. doi: 10.1007/s00203-010-0611-7.

868 Phadtare, S. (2004). Recent developments in bacterial cold-shock response. *Curr. Issues Mol. Biol.* 6,
869 125-136.

870 Phadtare, S., Hwang, J., Severinov, K., and Inouye, M. (2003). CspB and CspL, thermostable cold-
871 shock proteins from *Thermotoga maritima* 8, 801-810. doi: 10.1046/j.1365-2443.2003.00675.x.

872 Piette, F., D'Amico, S., Struvay, C., Mazzucchelli, G., Renaut, J., Tutino, M. L., et al. (2010).
873 Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic
874 bacterium *Pseudoalteromonas haloplanktis* TAC125. *Mol. Microbiol.* 76, 120-132. doi:
875 10.1111/j.1365-2958.2010.07084.x.

Pollo, S. M. J., Charchuk, R., and Nesbø, C. L. (2016). Draft genome sequences of *Kosmotoga* sp. DU53 and *Kosmotoga arenicorallina* S304. *GenomeA*, in press.

Pollo, S. M. J., Zhaxybayeva, O., and Nesbø, C. L. (2015). Insights into thermoadaptation and the evolution of mesophily from the bacterial phylum *Thermotogae. Can. J. Microbiol.* 61, 655-670.

Pysz, M. A., Ward, D. E., Shockley, K. R., Montero, C. I., Conners, S. B., Johnson, M. R., et al. (2004). Transcriptional analysis of dynamic heat-shock response by the hyperthermophilic bacterium *Thermotoga maritima*. *Extremophiles* 8, 209-217. doi: 10.1007/s00792-004-0379-2.

Qiu, Y., Cho, B. K., Park, Y. S., Lovley, D., Palsson, B., and Zengler, K. (2010). Structural and operational complexity of the *Geobacter sulfurreducens* genome. *Genome Res.* 20, 1304-1311. doi: 10.1101/gr.107540.110.

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, 803-805.

Ratkowsky, D. A., Lowry, R. K., McMeekin, T. A., Stokes, A. N., and Chandler, R. E. (1983). Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *J. Bacteriol.* 154, 1222-1226.

Robinson, M. D., and Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data 9, 321-332. doi: 10.1093/biostatistics/kxm030.

Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes 4, e19001.

Schumann, W. (2009). Temperature sensors of eubacteria. *Advances in applied microbiology* 67, 213-256. doi: 10.1016/S0065-2164(08)01007-1.

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, 182-203. doi: 10.1111/j.1574-6976.2012.00346.x.

Sievert, S., and Vetriani, C. (2012). Chemoautotrophy at Deep-Sea Vents: Past, Present, and Future. *Oceanography* 25, 218-233. doi: 10.5670/oceanog.2012.21.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690. doi: 10.1093/bioinformatics/btl446.

Swinnen, I. A. M., Bernaerts, K., Dens, E. J. J., Geeraerd, A. H., and Van Impe J. F. (2004). Predictive modelling of the microbial lag phase: A review. *Int. J. Food Microbiol.* 94, 137-159. doi: 10.1016/j.ijfoodmicro.2004.01.006.

909 Swithers, K. S., DiPippo, J. L., Bruce, D. C., Detter, C., Tapia, R., Han, S., et al. (2011). Genome
910 sequence of *Kosmotoga olearia* strain TBF 19.5.1, a thermophilic bacterium with a wide growth
911 temperature range, isolated from the Troll B oil platform in the North Sea. *J. Bacteriol.* 193, 5566-
912 5567. doi: 10.1128/JB.05828-11.

913 Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., et al. (2008). Cell
914 proliferation at 122°C and isotopically heavy CH_4 production by a hyperthermophilic methanogen
915 under high-pressure cultivation. *Proc. Natl. Acad. Sci. U. S. A.* 105, 10949-10954. doi:
916 10.1073/pnas.0712334105.

917 Wang, Z., Tong, W., Wang, Q., Bai, X., Chen, Z., Zhao, J., et al. (2012). The temperature dependent
918 proteomic analysis of *Thermotoga maritima* 7, e46463. doi: 10.1371/journal.pone.0046463.

919 Zhaxybayeva, O., Swithers, K. S., Foght, J., Green, A. G., Bruce, D., Detter, C., et al. (2012).
920 Genome sequence of the mesophilic Thermotogales bacterium *Mesotoga prima* MesG1.Ag.4.2
921 reveals the largest Thermotogales genome to date. *Genome Biol. Evol.* 4, 700-708. doi:
922 10.1093/gbe/evs059.

923 Zhaxybayeva, O., Swithers, K. S., Lapierre, P., Fournier, G. P., Bickhart, D. M., DeBoy, R. T., et al.
924 (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the
925 *Thermotogales. Proc. Natl. Acad. Sci. U. S. A.* 106, 5865-5870. doi: 10.1073/pnas.0901260106.

926 Zhu, Q., Kosoy, M., and Dittmar, K. (2014). HGTector: An automated method facilitating genome-
927 wide discovery of putative horizontal gene transfers. *BMC Genomics* 15, 717. doi: 10.1186/1471-
928 2164-15-717.

929

TABLE 1. Gene expression in selected laterally acquired temperature-responsive genes. At each temperature, the listed RPKM values represent the average expression levels across replicates. Values that are significantly different from 65°C are shown in bold font. None of the temperature responsive expression patterns show significant effect of growth rate and, where tested (i.e. R > 0.7), the growth-rate-only model was rejected.

		RPKM values				
Locus Tag	Functional annotation	30°C[a]	40°C	65°C	77°C	Identified by
Kole_0109	Cold shock protein	**5602**	**892**	222	119	Phylogenetic analysis[a]
Kole_0110	Histidine kinase	175	**333**	144	312	Phylogenetic analysis
Kole_0111	Response regulator	166	204	173	**446**	HGTector
Kole_0505	Glycerol dehydrogenase	721	**2668**	752	1242	HGTector
Kole_0506	Hypothetical protein	559	**2037**	461	783	Phylogenetic analysis
Kole_0507	Hypothetical protein	555	**2193**	521	809	HGTector
Kole_0508	Poly (3-hydroxybutyrate) depolymerase-like protein	212	423	200	314	HGTector
Kole_0897	Ribosomal RNA methyltransferase, rmlL	**503**	498	228	232	HGTector
Kole_0922	DEAD/DEAH box helicase	**755**	**288**	89	102	HGTector
Kole_0969	3-oxoacyl-ACP synthase III , FabH	**2386**	**3063**	939	1424	HGTector
Kole_0970	enoyl-ACP reductase II , fabK	2226	3243	1486	1641	HGTector[a]
Kole_0971	malonyl CoA-acyl carrier protein transacylase, fabD	2304	4211	2303	2647	HGTector
Kole_0972	acyl carrier protein	6531	12601	4850	4241	HGTector
Kole_0973	3-oxoacyl-ACP synthase II, fabF	4815	9257	4753	4498	HGTector
Kole_1015	Response regulator	**1289**	**515**	95	130	HGTector
Kole_1016	Response regulator	**783**	**280**	54	72	HGTector
Kole_1017	Histidine kinase	**697**	**275**	59	90	Phylogenetic analysis
Kole_1718	Ribosomal RNA methyltransferase, rmlH	**531**	332	211	203	HGTector
Kole_1745	PPIase FKBP-type	2783	2382	1541	**430**	HGTector

a) Phylogenetic tree is shown in Supplemental Figure 6.

936 **Figure Legends:**

937 **FIGURE 1. Growth rate of *K. olearia* as a function of temperature.** Isothermic growth curves
938 were generated at each temperature from an inoculum grown at that temperature for at least three
939 transfers (except for 25°C and 80°C, for which an inoculum from the same temperature could not be
940 generated; see materials and methods). Up-shifted and down-shifted growth curves were generated
941 from an inoculum that was grown at lower and higher temperatures, respectively. Red squares,
942 growth temperature up-shifted from 65°C to 77°C or from 40°C to 65°C; Blue circles, growth
943 temperature down-shifted from 77°C to 65°C, 65°C to 40°C, or 40°C to 30°C. Data points represent
944 the mean of replicate cultures (see materials and methods); error bars represent standard error.

945

946 **FIGURE 2. Model of energy generation pathway in *K. olearia* during growth on pyruvate.** The
947 model includes genes likely involved in conversion of pyruvate to acetate, CO_2, H_2, and ATP. The
948 genes were selected from the list of genes highly expressed across all temperature conditions (Table
949 S3 in Dataset S1). Acetate transport is not shown. The dashed box indicates hydrogenase activity.
950 The two highly expressed hydrogenases are shown, but their potential interactions with each other or
951 with the membrane are not known. Increased expression of citrate synthase at low temperature,
952 which could redirect acetyl-CoA away from acetate production, is shown in grey. The inclusion of
953 the ABC transporter is based on its high expression level but experiments are needed to confirm its
954 involvement in pyruvate transport. The model also explains the observed lower ratio of carbon
955 dioxide to hydrogen produced by growth on maltose vs. pyruvate (not shown), because during
956 growth on maltose reduced electron carriers would be generated from the conversions of maltose to
957 pyruvate.

958

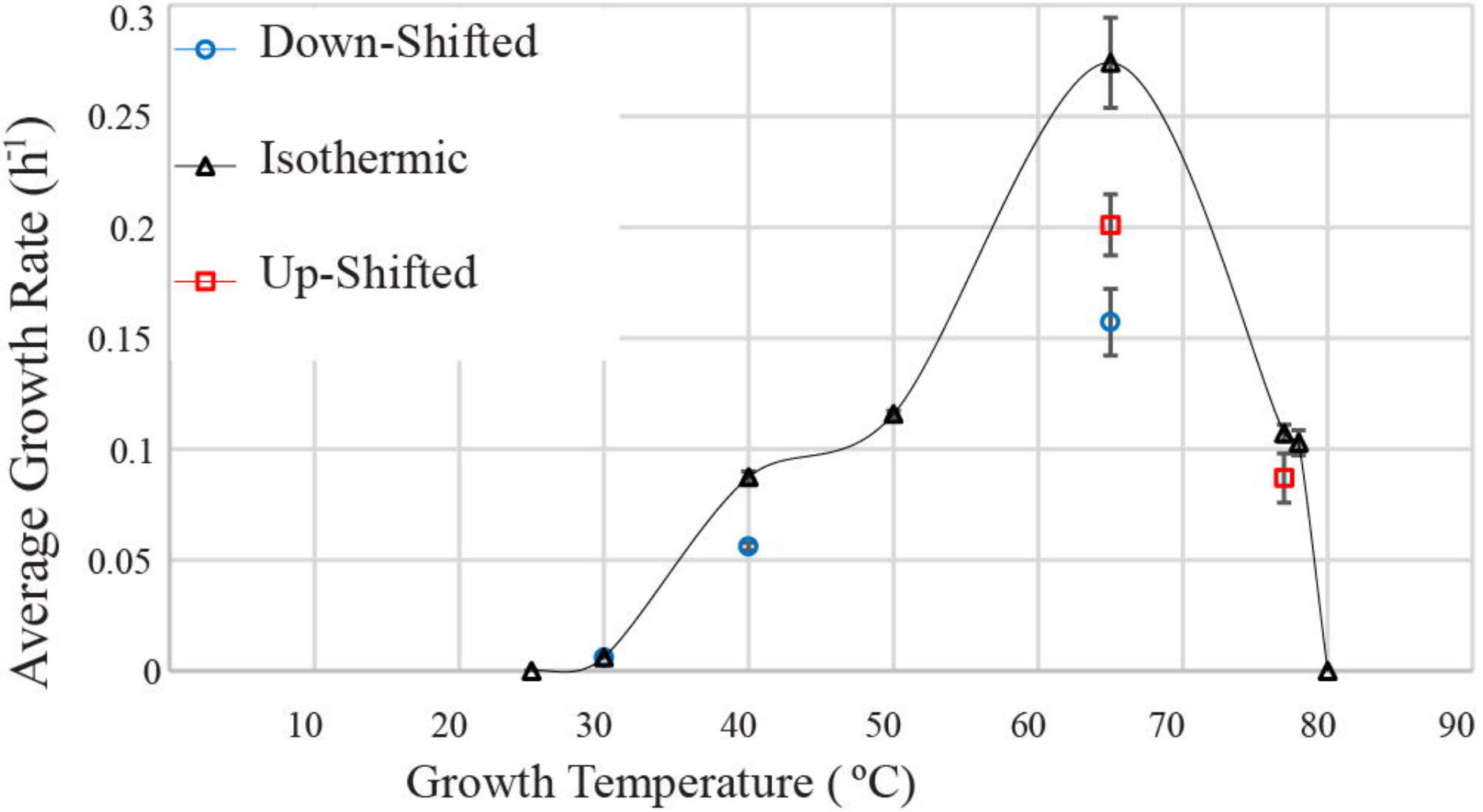
959 **FIGURE 3. Biplot of the principal component analysis of 12 transcriptomes.** Each transcriptome
960 is denoted by a point and genes are represented by vectors. Genes that point into a specific
961 "temperature quadrant" are up-regulated at the growth temperature(s) of that quadrant, and the five
962 longest (i.e., most highly correlated) gene vectors pointing to each quadrant are shown. Co-ordinates
963 and vector length for all genes can be found in Table S4 in Dataset S1. It should be noted that the *ffs*
964 (Kole_R0010) transcript is only 115 nt, and may not have been fully represented in every
965 transcriptome due to our isolation protocol which selects against small RNA (<200 nucleotides).
966 Also, the high expression of the alcohol dehydrogenase (Kole_0742) is probably due to the addition
967 of stop solution before RNA isolation (see Text S1).

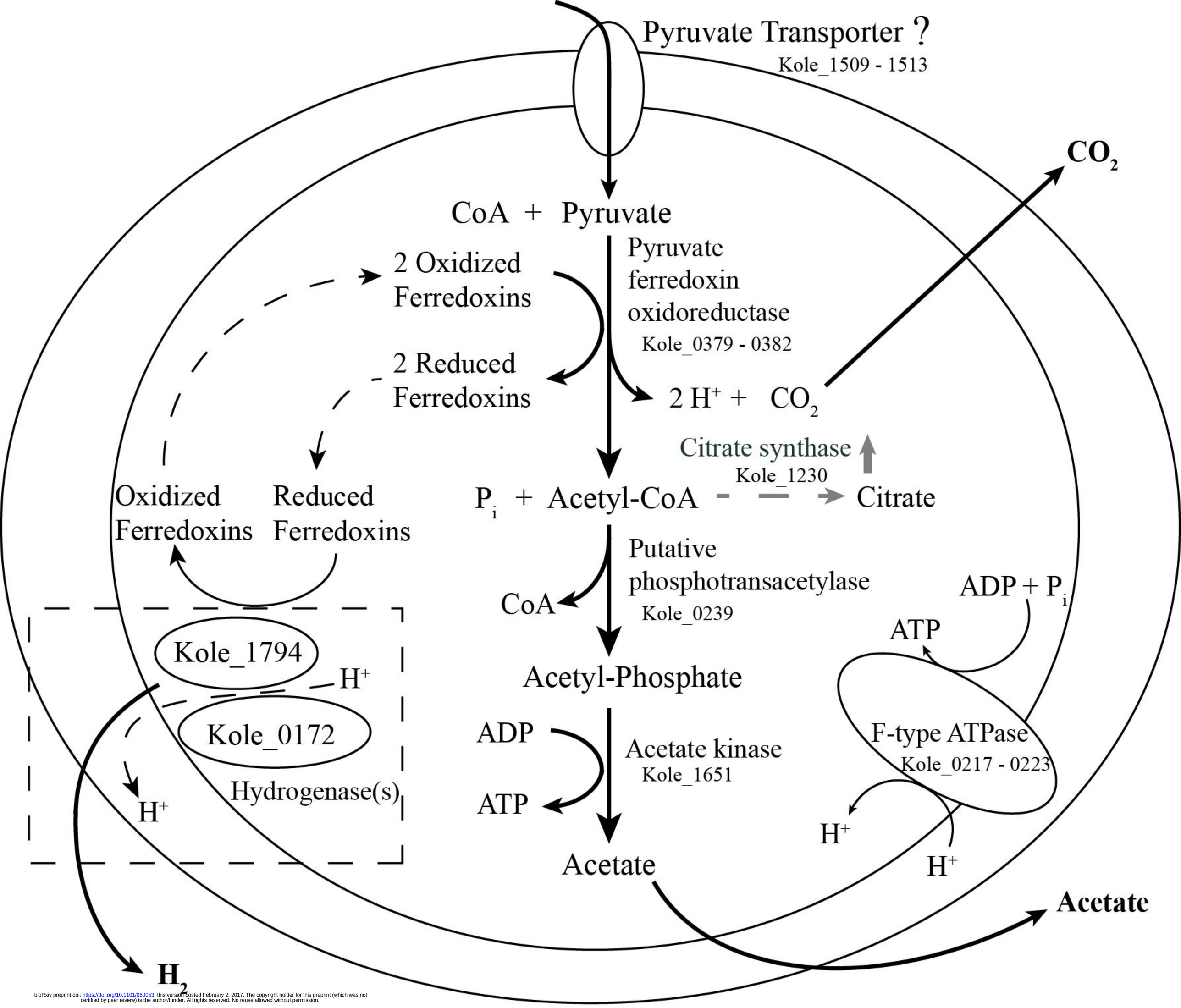
968

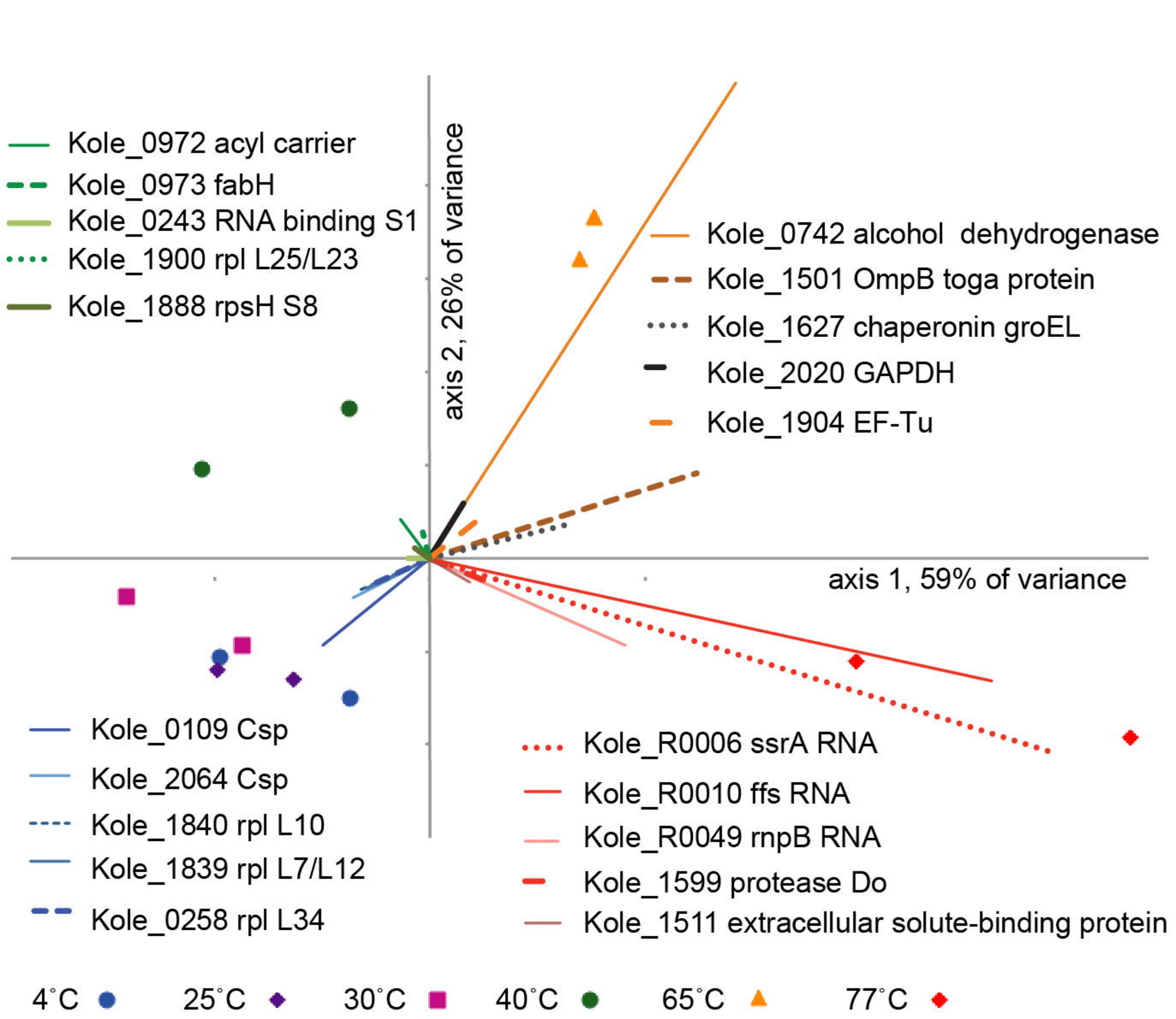
969 **FIGURE 4. Difference between observed and expected number of temperature responsive**
970 **genes across functional categories.** Functional categories were assigned using the Clusters of
971 Orthologous Groups (COG) database as implemented in IMG (Markowitz et al., 2014) and are
972 denoted by one-letter abbreviations along the X-axis (see Figure S3 legend for notations). NC, for
973 "no category", denotes genes not assigned to a functional category. For each temperature treatment
974 (30°C, 40°C and 77°C) only the temperature-responsive fraction of the *K. olearia* genome was
975 considered. If the temperature-responsive genes were randomly distributed across functional
976 categories we would expect the same fraction of temperature-responsive genes in each COG

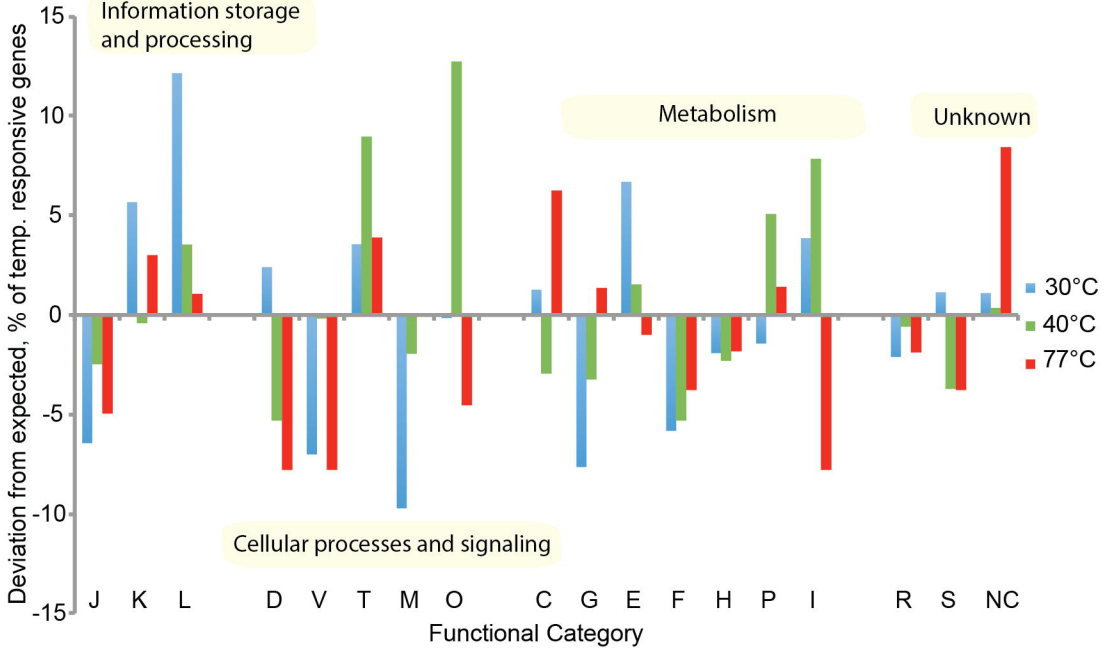
977 category. The difference (in percent) between the observed and expected number of temperature
978 responsive genes is plotted on the Y-axis with positive and negative values referring to over- and
979 under-representation of the temperature-responsive genes, respectively. For actual numbers of genes
980 in each COG category see Figure S3.

981


Down-Shifted
Isothermic
Up-Shifted
Average Growth Rate (h-1)
0.3
0.25
0.2
0.15
0.1
0.05
0
10
20
30
40
50
60
70
80
90
Growth Temperature (°C)




Kole_0972 acyl carrier
Kole_0973 fabH
Kole_0243 RNA binding S1
Kole_1900 rpl L25/L23
Kole_1888 rpsH S8
axis 2, 26% of variance
Kole_0742 alcohol dehydrogenase
Kole_1501 OmpB toga protein
Kole_1627 chaperonin groEL
Kole_2020 GAPDH
Kole_1904 EF-Tu
axis 1, 59% of variance
Kole_0109 Csp
Kole_2064 Csp
Kole_1840 rpl L10
Kole_1839 rpl L7/L12
Kole_0258 rpl L34
Kole_R0006 ssrA RNA
Kole_R0010 ffs RNA
Kole_R0049 rnpB RNA
Kole_1599 protease Do
Kole_1511 extracellular solute-binding protein
4˚C
25˚C
30˚C
40˚C
65˚C
77˚C


Information storage and processing
Metabolism
Unknown
Cellular processes and signaling
Deviation from expected, % of temp. responsive genes
15
10
5
0
-5
-10
-15
J
K
L
D
V
T
M
O
C
G
E
F
H
P
I
R
S
NC
Functional Category
30°C
40°C
77°C