nature microbiology

ARTICLES

PUBLISHED: 22 AUGUST 2016 | ARTICLE NUMBER: 16143 | DOI: 10.1038/NMICROBIOL.2016.143

Widespread formation of alternative 3′ UTR isoforms via transcription termination in archaea

Daniel Dar[1], Daniela Prasse[2], Ruth A. Schmitz[2] and Rotem Sorek[1]*

Transcription termination sets the 3′ end boundaries of RNAs and plays key roles in gene regulation. Although termination has been well studied in bacteria, the signals that mediate termination in archaea remain poorly understood. Here, we applied term-seq to comprehensively map RNA 3′ termini, with single-base precision, in two phylogenetically distant archaea: *Methanosarcina mazei* and *Sulfolobus acidocaldarius*. Comparison of RNA 3′ ends across hundreds of genes revealed the sequence composition of transcriptional terminators in each organism, highlighting both common and divergent characteristics between the different archaeal phyla. We find that, in contrast to bacteria, a considerable portion of archaeal genes are controlled by multiple consecutive terminators, generating several alternative 3′ untranslated region isoforms for >30% of the genes. These alternative isoforms often present marked length differences, implying that archaea can employ regulation via alternative 3′ untranslated regions, similar to eukaryotes. Although most of the terminators are intergenic, we discover numerous cases in which termination of one gene occurs within the coding region of a downstream gene, implying that leaky termination may tune inter-transcript stoichiometry in multi-gene operons. These results provide the first high-throughput maps of transcriptional terminators in archaea and point to an evolutionary path linking bacterial and eukaryal non-coding regulatory strategies.

Transcription termination is a highly regulated molecular process important for shaping the transcriptomes of all living organisms[1,2]. In the compact genomes of prokaryotes, efficient termination sets the boundaries between transcriptional units and prevents excessive antisense transcript overlap[3]. In addition, condition-specific regulation of termination via riboswitches allows bacteria to fine-tune their gene expression in response to changes in their environments[4,5].

In bacteria, transcription termination usually occurs via factor-independent intrinsic terminators, which, when transcribed, are known to form a stable RNA hairpin structure followed by a uridine-rich tract. It has been shown that transcription of the uridine-rich region causes the RNA polymerase (RNAP) to pause, and promotes hairpin formation, leading to destabilization of the RNAP complex and to the termination of transcription[1,6,7]. The mechanism of transcription termination in archaea, however, is much less understood. Termination studies in the archaeon *Methanothermobacter thermautotrophicus* have shown that bacterial intrinsic terminators can lead to efficient transcription termination of the archaeal RNAP, *in vitro*[8]. However, although the intrinsic terminator hairpin structure could increase termination efficiency, it was not essential for termination. In contrast, the presence of a short uridine tract was sufficient for eliciting termination on its own. In agreement with these results, mapping of the RNA 3′ ends in a number of genes in two additional Euryarchaea revealed that *in vivo* RNA 3′ ends were uridine-rich and, in some cases, had the potential of assuming stable RNA conformations[9]. Similar *in vitro* results were observed in other organisms[10,11] and were validated *in vivo* using a reporter assay that confirmed the essentiality of the uridine-rich tract for termination[12]. In contrast to these data on transcription termination in organisms belonging to the Euryarchaeota phylum, the signals that mediate termination in the second major archaeal phylum, the Crenarchaeota, are largely unknown[8,9,12,13].

We recently developed term-seq, a sequencing method that enables transcriptome-wide RNA 3′ end mapping in prokaryotes[14]. As part of this method, a single-strand sequencing adaptor is ligated directly to the RNA 3′ end, leading to a sequencing library in which the first base of each sequenced read corresponds to the last base of an exposed RNA 3′ end. We have shown that term-seq identifies, *in vivo*, expressed termination sites in bacteria to single-base resolution, enabling the reconstitution of the known terminator code in *Bacillus subtilis*[14]. In the current study we apply term-seq on representative species from Euryarchaea and Crenarchaea, revealing new insights into the archaeal transcription termination code as well as potential roles for termination in gene regulation.

Results

Genome-wide mapping of RNA 3′ ends reveals terminator determinants in *Methanosarcina mazei*. To study the features that define RNA 3′ ends in archaea, we applied term-seq on the model Euryarchaeon *Methanosarcina mazei*, generating over six million reads representing *in vivo* RNA 3′ end sequences from three biological replicates (see Methods; Supplementary Table 1). To minimize the possibility of sites generated by non-specific RNA degradation, we only analysed positions that appeared in all three biological replicates with significant coverage (see Methods)[14]. Using paired-end sequencing of 100–300 bp library inserts we directly associated between 3′ end termini and their respective genes (Fig. 1a, dashed line over the 3′ end), identifying term-seq predicted termination sites for 641 transcriptional units in *M. mazei*, representing 25% of 2,547 estimated transcriptional units in this organism[15] (Supplementary Table 2; see Methods).

Analysis of the dominant termination positions (the most highly covered site for each gene; see Methods) showed that the predicted transcriptional terminators of *M. mazei* protein-coding genes, as well as non-coding RNA (ncRNA)s[16], are characterized by a uridine-rich tract in the ten bases directly upstream of the 3′ end

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]Christian-Albrechts-University Kiel, Institute of General Microbiology, Am Botanischen Garten 1-9, 24118 Kiel, Germany. *e-mail: rotem.sorek@weizmann.ac.il

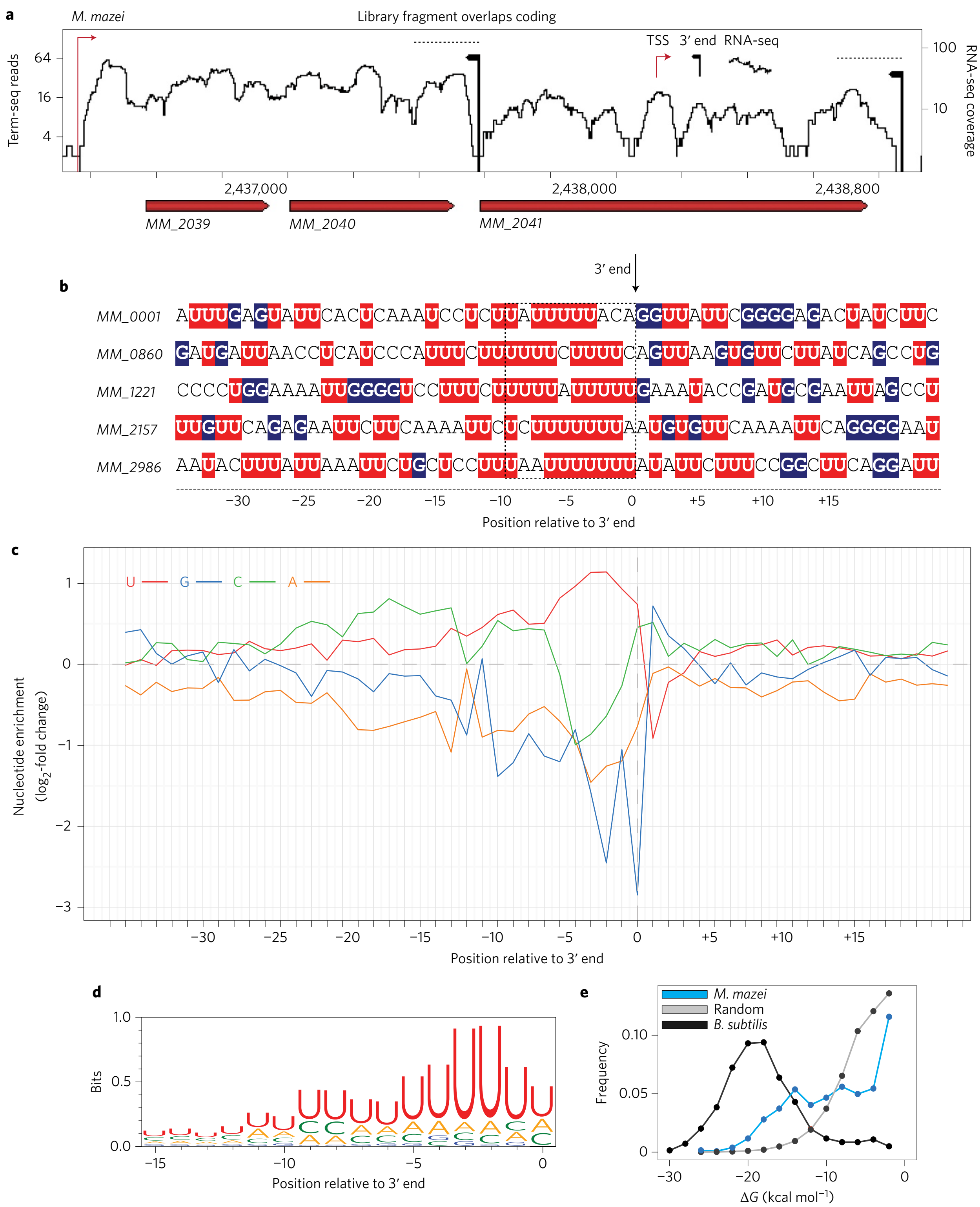

Figure 1 | Sequence and structural determinants of transcription terminators in *M. mazei*. **a**, High-resolution terminator mapping via deep-sequencing of RNA 3′ ends (term-seq). Transcription start sites (TSSs)[16], termination sites and RNA-seq coverage are depicted as red arrows, black arrows and black coverage curves, respectively. The height of the termination signal (black arrow) reflects the average number of supporting term-seq reads in three biological replicates. The average library insert lengths associated with each 3′ position, inferred from paired-end sequencing, were used to associate between RNA 3′ ends and their respective genes (dashed lines over terminator indicate average insert length). **b**, Sequences around the termination points of five representative genes in *M. mazei*. Numbers on the *x* axis depict the position relative to the measured primary termination sites (0 being the exact site of termination). **c**, Nucleotide enrichment meta-analysis across all 641 terminators identified in *M. mazei*. Position-specific nucleotide enrichment was assessed by comparing the terminator set to randomly selected intergenic positions ($n = 10{,}000$). **d**, Logo representation of the sequence terminator signature[18]. **e**, Distributions of predicted RNA structure stabilities (in kcal mol^{-1}) for *M. mazei* terminators (blue), random intergenic positions from the *M. mazei* genome (grey) and term-seq mapped *B. subtilis* terminators (black)[14].

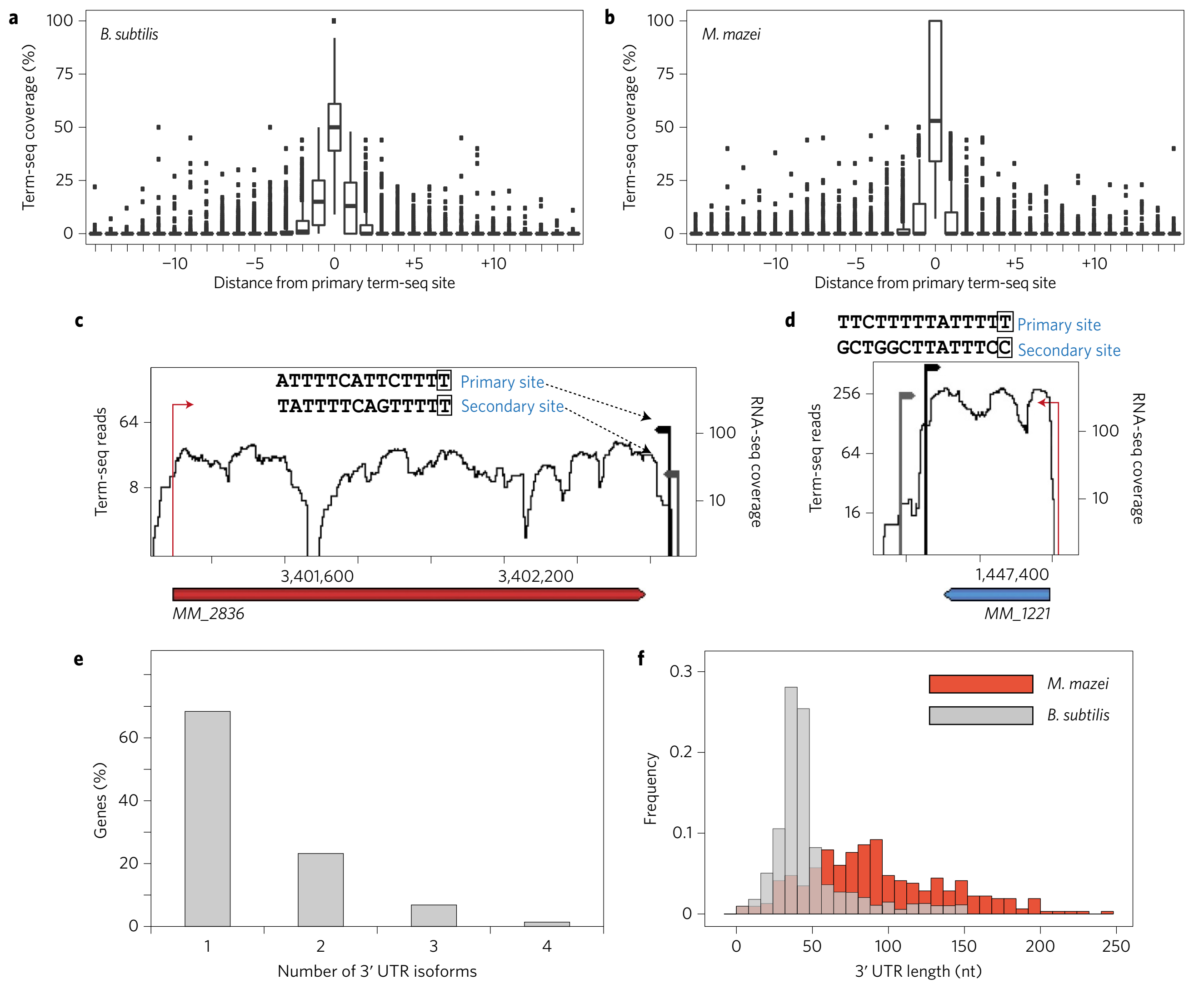

Figure 2 | Multiple terminators control gene boundaries in *M. mazei*. a,**b**, Termination sites determined by term-seq are highly precise in *B. subtilis* (**a**) and *M. mazei* (**b**). The term-seq signal contribution was calculated for the 30 nucleotide positions adjacent to the primary terminator site (15 nt on each side), only using term-seq sites that appeared in all three repeats. Outliers are shown as black dots. **c**,**d**, Examples for genes controlled by multiple terminators in *M. mazei*, along with the sequence directly upstream of the terminator. The boxed nucleotide represents the measured 3′ end. Primary and secondary terminators are shown in black and grey, respectively. The height of the termination signal (black arrow) reflects the average number of supporting term-seq reads in three biological replicates. **e**, Distribution of the number of terminators detected per transcriptional unit in *M. mazei*. **f**, 3′ UTR length (nt) distributions in *M. mazei* (red) and *B. subtilis* (grey).

(Fig. 1b–d). This tract was significantly and specifically depleted of guanosine nucleotides, so that guanosines are almost completely absent at the termination site (Fig. 1c,d; $P < 0.005$). A weaker, cytidine-rich stretch was observed in the ten bases preceding the uridine-rich tract. An approximately twofold enrichment in guanosine and twofold depletion in uridine was observed in the first base directly downstream of the termination site. Together, our results point to an information-rich signal that defines the *M. mazei* sequence code for termination. We found that this signal is 7.4-fold enriched in intergenic regions and largely depleted from gene-coding sequences in *M. mazei* ($P < 1 \times 10^{-16}$, binomial test). Nevertheless, we could detect 117 instances where this consensus sequence was present in expressed protein-coding genes, and in 25% of these cases (29/117) we detected term-seq sequences supporting termination within the gene. The remainder were not supported by term-seq reads, despite obvious expression (see Methods), suggesting that rare termination signals within coding regions may be suppressed by as yet undetermined mechanisms.

As RNA structure is known to mediate termination in bacteria, we used RNAfold[17] to predict the folding of the sequences immediately upstream of our terminator set. We found that terminators tend to produce more stable RNA secondary structures than non-terminating sequences (Fig. 1e). However, although terminators were twice as stable as randomly selected sites on average ($P < 1 \times 10^{-35}$, Wilcoxon rank-sum), many terminators did not contain any significant predicted structure and, as a whole, terminator ΔG values presented a relatively uniform distribution, in marked contrast to bacterial terminators (Fig. 1e and Supplementary Fig. 1b). Thus, although stable secondary structures might play some role in controlling *M. mazei* termination, it is unlikely to be an essential one.

Multi-terminator arrangements commonly control gene termination in *M. mazei*. *In vitro* studies, focused on a small set of model bacterial and archaeal terminators, found that termination occurs in a precise manner, such that transcription usually aborts at a specific position with a flexibility of one to two nucleotides around the termination point[8,12,19]. It has so far remained unknown whether *in vivo* termination behaves similarly. To quantify terminator precision *in vivo*, we examined the term-seq read distributions around all primary sites and compared the relative contributions of each neighbouring nucleotide position to the total signal (see Methods). In agreement with the previous *in vitro* reports, we found that *in vivo* termination, in both *B. subtilis* and *M. mazei*, is highly locally precise, with the vast majority of RNA 3′ ends occurring within a range of one nucleotide from the determined primary site (Fig. 2a,b).

Interestingly, we found that a significant portion of *M. mazei* transcripts, comprising 32% (203/641) of all analysed protein-coding genes and 26% (6/23) of previously validated ncRNAs (ref. 16), are associated with multiple predicted termination sites in addition to the primary (most highly covered) site (Fig. 2c–e and Supplementary Table 3; see Methods). This is in sharp contrast to *B. subtilis*, where only 5% (87/1,708) of the genes presented this pattern (Supplementary Table 4). These secondary sites contained a sequence signal highly similar to that observed for the primary terminators (Supplementary Fig. 2), strongly suggesting that these positions represent actual sites of transcription termination. In contrast, the vast majority of *B. subtilis* secondary sites did not conform to the known terminator code and are thus probably generated by processes other than termination (Supplementary Table 5).

The 3′ untranslated regions (3′ UTRs) of *M. mazei* genes, measured as the distance between the stop codon and the primary sites, were significantly longer and more widely distributed than those found in *B. subtilis* ($P < 1 \times 10^{-122}$, Wilcoxon rank-sum), with a median 3′ UTR of 88 ± 42 nt in *M. mazei* compared with 40 ± 27 nt in *B. subtilis* (Fig. 2f). Taken together, these results suggest that 3′ transcript boundaries in *M. mazei* are frequently controlled by multiple consecutive terminators, reflecting a potentially fundamental—and previously unknown—difference in terminator encoding strategies between bacteria and archaea.

Unique terminator features in Crenarchaeota. To acquire a more comprehensive view of archaeal transcription termination we sequenced the RNA 3′ termini of the model Crenarchaeon *Sulfolobus acidocaldarius* grown to stationary phase, identifying primary term-seq sites for 707 transcriptional units out of 1,336 transcriptional units predicted in this organism[15] (Fig. 3a and Supplementary Tables 1 and 6; see Methods). We found that the consensus sequence of termination positions in *S. acidocaldarius* was composed of two distinct uridine-rich blocks: a distal block at positions −20 to −12 and a proximal block spanning the nucleotides immediately upstream of the termination site (Fig. 3b–d). Significant enrichment for cytosine was observed in position −1 (Fig. 3b,c). Additionally, guanosine depletion was detected approximately 4 nt upstream of the terminator positions (Fig. 3c). Similarly to *M. mazei*, we found 6.5-fold enrichment for these features in intergenic regions ($P < 1 \times 10^{-16}$, binomial test; see Methods), although a significant number of such signals ($n = 390$) were also found in protein-coding genes, only 25% (98/390) of which were supported by term-seq reads. Finally, we found no evidence for a specific RNA structure near the termination sites in *S. acidocaldarius* (Fig. 3e).

Remarkably, as in the case in *M. mazei*, 39% (274/707) of the genes in *S. acidocaldarius* were associated with more than one termination site (Fig. 3a and Supplementary Table 7). Because most of these sites recapitulated the signals that define primary sites (Supplementary Fig. 3 and Supplementary Table 4), they probably represent alternative terminators for the same gene. However, in contrast to primary sites, uridine enrichment, while still significant, was limited to the seven nucleotides immediately upstream of the termination site (Supplementary Fig. 3). In addition, although cytosine enrichment at the −1 position was apparent, the effect of guanosine depletion was significantly reduced at these secondary terminator positions (Supplementary Fig. 3). Therefore, the secondary termination sites possibly represent less efficient terminator sequences.

The 3′ UTR lengths of *S. acidocaldarius*, defined as the distance between the stop codon and the primary termination site, were significantly larger than those of *B. subtilis* ($P < 1 \times 10^{-28}$ Wilcoxon rank-sum), but somewhat shorter than those detected in *M. mazei* ($P < 1 \times 10^{-30}$ Wilcoxon rank-sum) (Fig. 3f). Nonetheless, we find that the short 3′ UTRs in *S. acidocaldarius* were usually (97/114 of 3′ UTRs shorter than 40 nt, 85%) associated with secondary sites downstream, but not upstream, of the primary terminator, therefore extending the effective 3′ UTR lengths.

These analyses highlight a set of conserved terminator features shared between *M. mazei* and *S. acidocaldarius* (primarily the proximal uridine tract and general lack of RNA structure), but also show phylum-specific divergence in the terminator signals. Together, our results strongly support that multiple-terminator control of gene boundaries is a common and conserved phenomenon in archaea.

Termination-induced antisense overlap in *S. acidocaldarius*. Efficient transcription termination allows minimal overlapping of transcriptional units, which in the case of antisense transcription may interfere with gene expression via mRNA destabilization or reduced translation efficiency[20–22]. Remarkably, in 52% (156/301) of the gene pairs presenting convergent orientations in *S. acidocaldarius*, the terminator of one gene appeared within the coding sequence of the other gene and vice versa, resulting in significant potential antisense overlap between the transcripts of the two genes (Fig. 4a–c). This is in stark contrast to the 3% (27/796) and 8% (22/264) of such cases in *B. subtilis* and *M. mazei*, respectively. Interestingly, genomic annotation in this organism showed pervasive overlap between the coding regions of convergent genes, with 34% (102/301) of such antisense oriented gene pairs having no intergenic region. In comparison, the *B. subtilis* gene set contained only 13 such events and in *M. mazei* we could not detect even a single case in our data.

Termination within the coding sequence was also not infrequent for genes encoded on the same strand (Fig. 4d–g). We found that in 16% (62/380) of the consecutive gene pairs in *M. mazei* and 41% (168/408) in *S. acidocaldarius* the terminator of the upstream gene occurred at least 30 nt into the coding sequence of the downstream gene. In some cases, transcriptional termination of the upstream gene appeared inefficient, generating partial readthrough into the downstream gene (Fig. 4d–g). Such a leaky termination strategy can generate two RNA isoforms encoded from an operon governed by a single promoter, with the downstream genes expressed in lower amounts.

Regulatory potential of alternative 3′ UTR isoform abundance. In eukaryotes, regulated switching between alternative 3′ UTR isoforms can control gene expression by including or excluding regulatory regions, for example, micro-RNA (miRNA) binding sites[23]. Our discovery of widespread multi-terminator control in both *M. mazei* and *S. acidocaldarius* demonstrates that multiple 3′ UTR isoforms exist for at least one-third of the genes in archaea, similarly to eukaryotes. These isoforms frequently add or subtract a substantial amount of RNA sequence from the transcript 3′ UTR, with an average difference of 50 ± 34 nt and 52 ± 38 nt between isoforms in *M. mazei* and *S. acidocaldarius*, respectively (Supplementary Fig. 4a,b). To examine whether potential base-pairing interaction may occur between ncRNAs and 3′ UTRs, we compared the sequences of 23 previously validated ncRNA in *M. mazei*[16] to the set of 641 3′ UTR regions we mapped in this organism using term-seq. We detected a potential for ncRNA/3′ UTR base-pairing, defined here as at least 11 consecutive

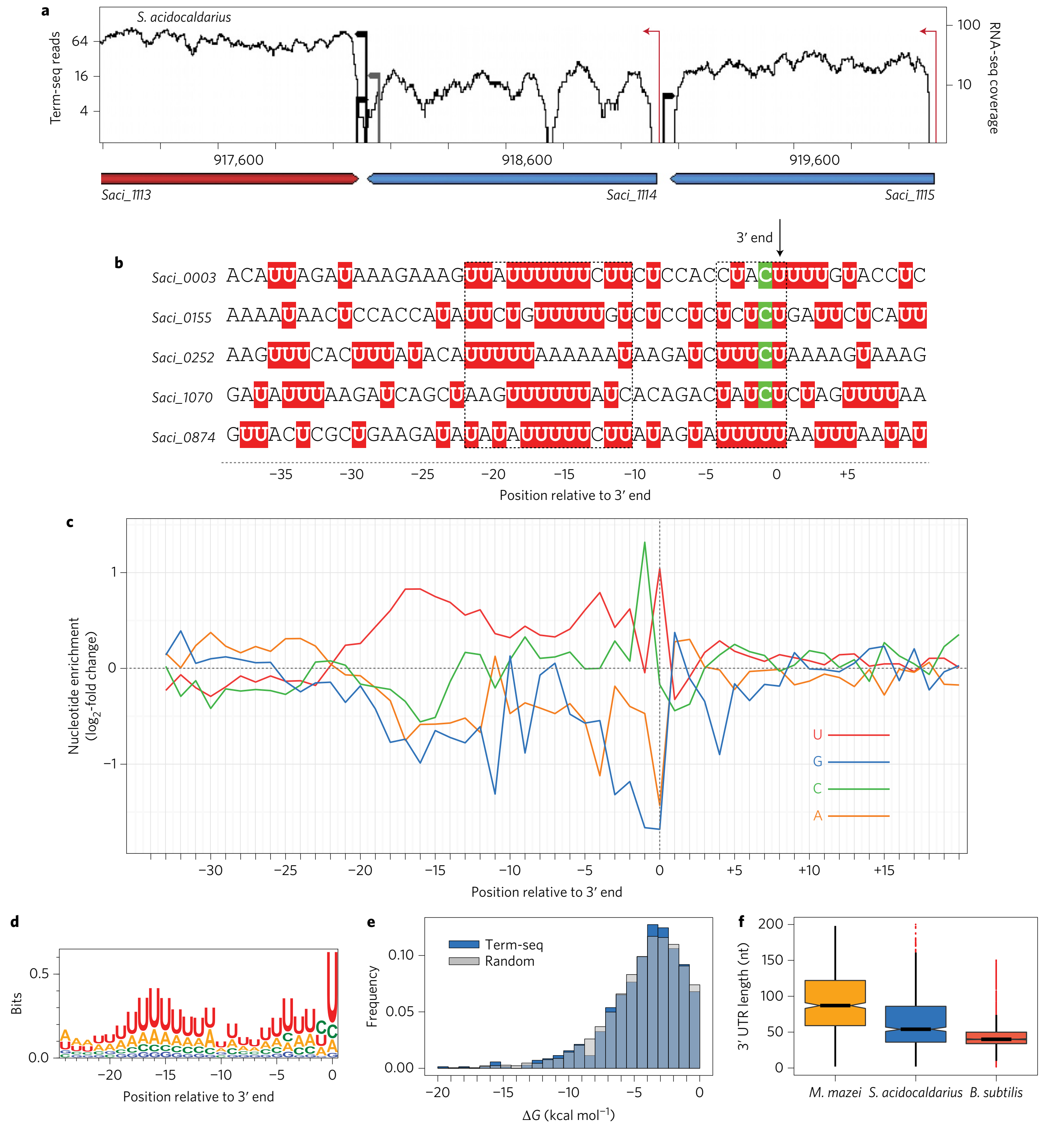

Figure 3 | Determinants of transcription terminators in *S. acidocaldarius*. **a**, Terminator mapping in *S. acidocaldarius*. TSSs, termination sites and RNA-seq coverage are depicted as in Fig. 1. **b**, Sequences around the termination points of representative genes in *S. acidocaldarius*. **c**, Nucleotide enrichment meta-analysis across all 707 terminators identified in *S. acidocaldarius*. Position-specific nucleotide enrichment was assessed by comparing the terminator set to randomly selected intergenic positions ($n = 10{,}000$). **d**, Logo representation of the terminator signature. **e**, Distributions of predicted RNA structure stabilities (kcal mol^{-1}) for *S. acidocaldarius* terminators (blue) and random intergenic positions (grey). **f**, Distributions of the 3′ UTR lengths of *M. mazei* (yellow), *S. acidocaldarius* (blue) and *B. subtilis* (red) depicted as a notched box plot. The median value is shown as a black horizontal line in each box and outliers are marked as red dots.

matching bases, for 87% (20/23) of the ncRNAs, which together were mapped to 14% (88/641) of the *M. mazei* 3′ UTRs (see Methods; Supplementary Tables 8 and 9). Therefore, control of transcript 3′ UTR isoform abundance could, in principle, mediate 3′ UTR-based gene regulation in archaea.

Discussion

In this study we have comprehensively mapped the *in vivo* RNA 3′ termini of archaea belonging to the Euryarchaea and Crenarchaea phyla, represented here by *M. mazei* and *S. acidocaldarius*, respectively. The single-base resolution of the derived RNA 3′ ends enabled

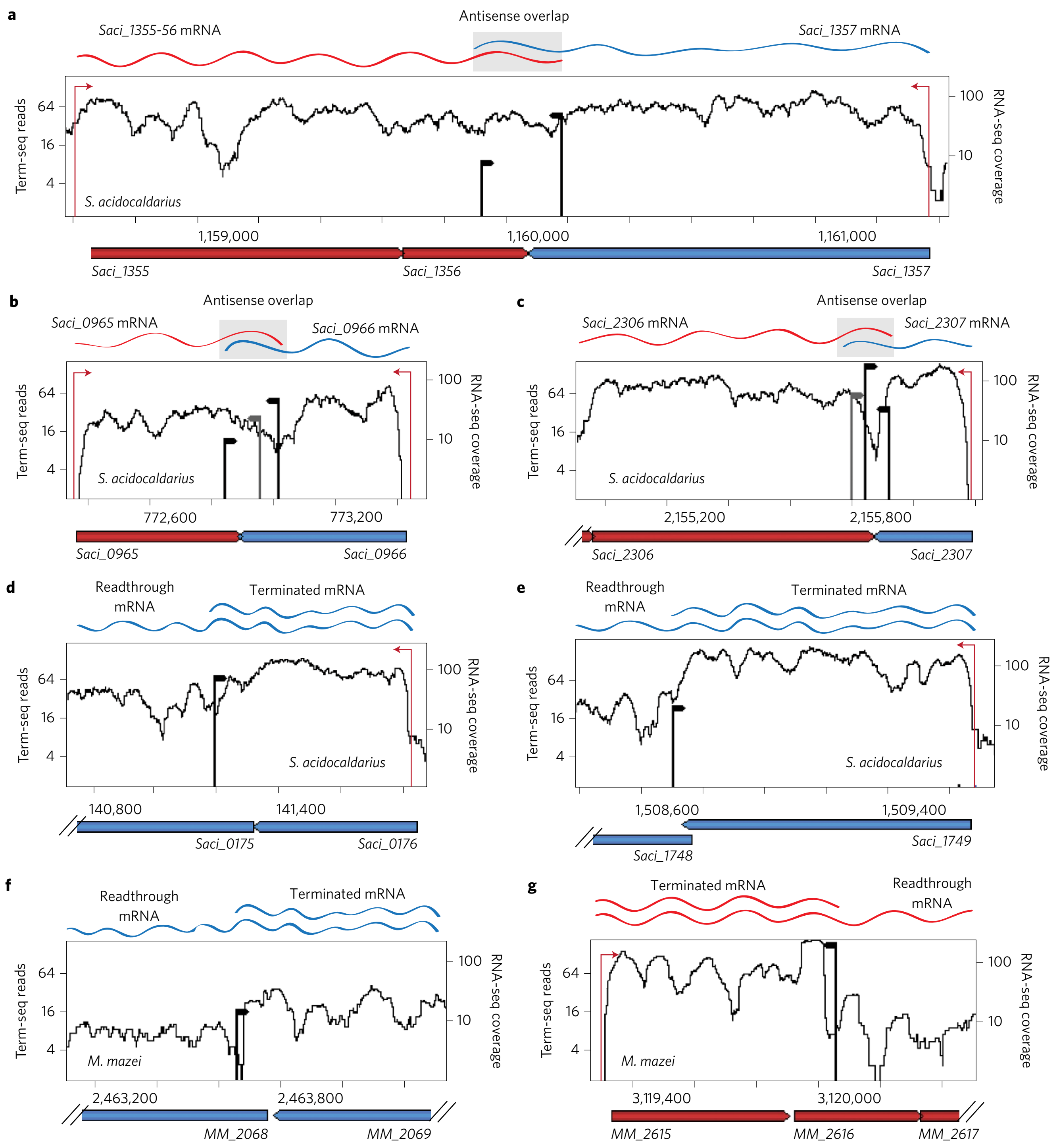

Figure 4 | Abundant overlapping termination suggests possible role in gene regulation. a–c, Examples of antisense oriented genes displaying transcription termination within the coding region of the opposite gene. Primary and secondary terminators are shown as black and grey arrows, respectively. Red and blue lines above represent the mRNAs and the predicted transcript overlaps are shown as shaded boxes. **d–g**, Examples where termination occurs within the coding region of genes co-transcribed together as a polycistron. Red and blue lines represent the transcript mRNAs generated either by termination (short mRNA) or readthrough. Black curve represents RNA-seq data.

the detection of a unique set of characteristics that together define a typical terminator model for each organism and collectively point to a divergence in the mechanisms of termination between the two archaeal phyla. Our data not only reveal the terminator code for archaea, but also provide a reference terminator set that will facilitate studies of gene regulation for individual genes in these two organisms. For this, a transcriptome browser integrating the term-seq, RNA-seq and transcription start site (TSS) data was established and can be accessed at http://www.weizmann.ac.il/molgen/Sorek/archaea_termseq/.

It has previously been shown that *in vivo* and *in vitro* termination in the Euryarchaeon *T. kodakarensis* can be mediated by a stretch of uridine residues[12]. Our results in *M. mazei*, another Euryarchaeon, validate these observations across hundreds of different loci and demonstrate that termination *in vivo* occurs immediately downstream of the uridine tract (Fig. 1b–d). Moreover, comparison of termination positions for 641 genes enabled the detection of previously unknown signals, including guanosine depletion upstream of termination sites and uridine depletion at the +1 position following the terminator (Fig. 1c). Although we find some enrichment for predicted RNA structure upstream of termination sites in *M. mazei*, the overall fold stability and the lack of structure in nearly half of the terminators suggest that it is not essential for termination (Fig. 1e). Possibly, such RNA structures contribute to termination efficiency in *M. mazei*, as has been shown *in vitro*[8].

Previous RNA 3′ mapping of several genes in the *Sulfolobus* virus-like particle SSV1 predicted that termination occurs immediately downstream of TTTTTYT sequences, usually in the context of a 16–19 nt pyrimidine-rich stretch[13]. Our results show that the uridine-rich region is in fact composed of two regions, a proximal one immediately upstream of the terminator and an additional distal region found −20 to −12 nt upstream (Fig. 3b–d). Nevertheless, a large fraction of the terminators and specifically the secondary terminators, were associated with a proximal uridine tract only, suggesting that multiple termination signals may be at play in archaea of this phylum. In addition, we could not detect any significant RNA secondary structure associated with *S. acidocaldarius* terminators. This may be attributed to the high temperatures in which this organism thrives, which can impede the formation of short RNA secondary structures (Fig. 3e).

Similarly to bacteria and archaea, the eukaryotic RNAPIII requires a short stretch of uridines for the termination of tRNAs and other small RNA genes. However, as the archaeal RNAP is more similar to RNAPII (refs 24,25), a polymerase not known to terminate at poly-U stretches, it is unclear how termination signals evolved. Nevertheless, it is now clear that poly-U sequences are important for terminating transcription in all three domains of life.

An important result of this study is the finding that many archaeal genes employ multiple consecutive terminators. Previous tiling-array analysis in *Halobacterium salinarum* reported decay in the hybridization signal downstream of genes, consistent with our high-resolution data[26]. Interestingly, this observation suggests that the RNAP does not always recognize a terminator sequence and in some cases can even skip multiple consecutive termination signals before finally dissociating, suggesting termination is either inefficient or potentially regulated. A previous *in vitro* study found that upstream template sequences, as far as 200 nt, can influence termination efficiency through an unknown mechanism in archaea[8]. It is thus possible that the observed multiple terminators occur downstream of genes where upstream sequences inhibit termination. Alternatively, it may be that termination efficiency is regulated by transcription elongation or termination factors that associate with the RNAP to enhance or bypass specific termination sequences and so select the site of RNA dissociation.

An interesting consequence of multi-terminator control is that the effective 3′ UTR lengths in archaea are significantly larger than those of bacteria (Figs 2f and 3f), offering a possible mode of regulation that depends on alternative isoforms differing in their non-coding 3′ UTRs. Long 3′ UTRs represent a common mode of mRNA regulation in eukaryotes[23], and such 3′ UTRs typically contain binding sites for non-coding RNAs (miRNAs). In bacteria, however, regulatory 3′ UTRs are rare, and mRNA/non-coding RNA interactions typically occur in the 5′ UTR (refs 27,28). The extension of the 3′ UTRs in archaea is therefore intriguing, particularly in light of the fact that ~70% of genes in Crenarchaea are devoid of 5′ UTR (ref. 29). Archaea possibly represent an evolutionary transition point between regulation by 5′ UTRs, common in bacteria, and the 3′ UTR regulation in eukaryotes[23].

Although we were able to predict potential base-pairing interactions between ncRNAs and 3′ UTRs in *M. mazei*, sequence-based target prediction for ncRNA/mRNA interactions is notoriously non-specific[30], and such predictions should be considered with caution. Therefore, the nature of the putative 3′ UTR-mediated regulation in archaea remains to be explored further. Specifically, it would be interesting to examine whether 3′ UTR isoform frequencies for particular genes are dynamically altered in different physiological or growth conditions, which would imply active regulation.

The central role of transcription termination is to set the 3′ boundaries between adjacent genes. Nevertheless, in *S. acidocaldarius* we found many cases in which termination of one gene occurs well within the coding region of its downstream neighbouring gene. In antisense-oriented gene pairs, such termination often generates the potential for extensive antisense overlap, which may have regulatory consequences, as has been shown in bacteria[20,21] (Fig. 4a–c). In addition, our finding of within-operon leaky termination offers a possible mechanism to shape gene stoichiometry within transcripts belonging to the same operon, allowing differential expression of genes transcribed from a single promoter (Fig. 4d–g).

Transcriptome-wide studies on transcript boundaries in archaea have so far focused on RNA 5′ ends, and such studies provided numerous mechanistic insights into transcription initiation and gene regulation[16,29,31]. However, largely due to technical limitations, the 3′ ends of archaeal transcripts have so far not been mapped at genome scales. This study, in which we provide the first high-throughput terminator analysis of model archaeal organisms, points to an unexpected regulatory complexity in archaeal mRNA 3′ ends and sets the framework for future studies on 3′ UTR-mediated non-coding RNA regulation in archaea.

Methods

Strains, growth conditions and RNA extraction. *M. mazei* cultures were grown under anaerobic conditions in 50 or 70 ml closed serum bottles with a gas phase of 80% N_2 and 20% CO_2 (refs 32,33). The minimal medium was supplemented with 150 mM methanol and 40 mM acetate as sole energy and carbon source. Growth was monitored by determining culture turbidity at 600 nm. For cultivation under nitrogen sufficiency (NS) 10 mM ammonium chloride was added to the medium. For RNA isolation, *M. mazei* cultures were grown to a turbidity of 0.15–0.25 (NF) or 0.4–0.5 (NS) at 600 nm, which corresponds to the exponential growth phase under the respective growth condition, and harvested at 4°. The obtained cell pellet was resuspended in Isol-RNA Lysis Reagent (5 Prime), followed by RNA isolation using the Direct-zol RNA MiniPrep Kit (Zymo Research), according to the manufacturer's instructions. RNA integrity was evaluated using a Tapestation 2200 Instrument (Agilent).

S. acidocaldarius cells were grown in Brock medium supplemented with 0.1% trypton (Difco)[34]. Frozen bacterial pellets were lysed using the Fastprep homogenizer (MP Biomedicals) and RNA was extracted with the FastRNA PRO[T] blue kit (MP Biomedicals, 116025050) according to the manufacturer's instructions. RNA levels and integrity were determined by Qubit RNA BR Assay Kit (Life Technologies, Q10210) and Tapestation, respectively. All RNA samples were treated with TURBO DNase (Life Technologies, AM2238).

Library construction, deep-sequencing and read mapping. Term-seq and RNA-seq libraries were constructed as in ref. 14 and were sequenced on the Illumina Nextseq 500 platform either as single or paired ends (Supplementary Table 1). The sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession no. PRJEB14292. Reads were then mapped to the reference genomes using NovoAlign (Novocraft) V3.02.02 with default parameters, discarding reads that mapped to more than one genomic position. In the case of paired ends, we required that both reads map uniquely, discarding pairs in which the mapped insert length was greater than 500 nt. Gene annotation and sequences were downloaded from Genbank: NC_003901.1 and NC_007181.1 for *M. mazei* Goe1 and *S. acidocaldarius* DSM 639, respectively.

TSSs for *B. subtilis*, *M. mazei* and *S. acidocaldarius* were taken from refs 14, 16 and 35 respectively. RNA-seq data for *S. acidocaldarius* were taken from ref. 35 and analysed as described above.

Terminator identification. The number of 3′ ends mapped to each genomic position was recorded, and sites appearing in all three replicates with a minimum number of reads and average coverage greater or equal to a set threshold were collected for downstream analysis. The minimum number of reads and average coverage thresholds were set as follows: two reads per sample and an average of four for *M. mazei*, four and four for *B. subtilis*, and two and six for *S. acidocaldarius*. For each 3′ site the average library insert length was calculated using the paired-end read mapping positions. Sites were then associated with their respective genes, requiring that the average insert length allow at least 1 nt overlap with the gene coding region. Annotated tRNAs, rRNAs and ncRNAs were excluded as they are generally present in a stable processed form. In the case of *B. subtilis*, as the term-seq data was previously sequenced with single ends, we allowed a maximal 3′ UTR length of 150 nt and up to 10 nt invasion into the downstream gene if this gene was on the same strand. In the case that the downstream gene was on the opposite strand, the length of the invasion was not restricted.

Primary term-seq predicted terminators were assigned as the most highly covered gene-associated position. In cases where equal coverage was shared between multiple sites, the position further downstream was selected as the primary terminator. Secondary term-seq predicted terminators were assigned as all additional, non-primary sites that had the maximal coverage in a region spanning 12 nt upstream and downstream of their respective position and that were covered by at least 10% of the average term-seq coverage over all dominant (primary and secondary) positions associated with the gene.

Terminator sequence and structure analysis. The upstream and downstream sequences relative to the exact termination sites were collected from the latest version of the reference genomes (as described above). For the nucleotide usage enrichment analyses, the frequency of each base in a given position relative to the terminators (−1, +1 and so on) was calculated using the sequences above and compared to an identical calculation resulting from sampling 10,000 randomly chosen intergenic positions. The $\log_2$-fold enrichment was calculated and plotted as in Figs 1–3.

To assess the significance of the enrichment/depletion we performed the following simulation. For each organism and its respective terminator collection we generated 1,000 randomly selected sets of intergenic positions, with each such set fixed to the same size as the terminator collection ($n = 641$ and $n = 707$ for *M. mazei* and *S. acidocaldarius*, respectively). For each such simulated set we calculated the enrichment for each nucleotide in each position as described for the terminator set, generating a distribution of possible enrichments for each nucleotide type and position but from non-terminating positions. This distribution was used to calculate an empirical P value for enrichment. Sites in which the calculated P value was smaller or equal to 0.005 were considered significant.

Predicted RNA structural stability analysis was performed by folding the 45 nt long DNA sequence found upstream to the terminators using the RNAfold software[17]. Distribution comparison was performed with the Wilcoxon rank-sum rank-sum test R package.

Terminator sequence enrichment in intergenic versus coding regions was performed by searching for terminator motifs in the genomic sequence in the entire genome, examining both plus and minus strands in the case of intergenic sites. In *M. mazei*, the signal was defined as at least 80% T content in the upstream 10 nt and the last nucleotide being an A, C or G, but not T. In *S. acidocaldarius*, we required a five consecutive T tract to be positioned at a distance of 8–15 nt from the selected genomic positon, as well as a pyrimidine frequency of at least 70% in the first upstream 10 nt and the first two nucleotides being either C or T. Enrichment was calculated by comparing the number of motifs found in intergenic versus coding regions and normalizing by the number of searches in coding and intergenic regions, respectively (correcting for double strand counting in intergenic regions).

To evaluate whether terminator-like sequences that are located within protein-coding regions are supported by term-seq reads, we collected all such positions associated with significantly expressed genes, covered by a minimum of 50 and 100 reads per kilobase (r.p.k.) for *M. mazei* and *S. acidocaldarius*, respectively. The total term-seq coverage for each position was calculated as the sum of reproducible term-seq reads in the five adjacent nucleotides. In cases where terminator-like sequences were adjacent to one another, the most highly covered position was selected. Terminator-like sequences, detected as described above and that were supported by at least two reproducible reads, were considered as terminating sequences in this analysis.

ncRNA target analysis. Small RNA target prediction was performed by comparing the sequences of the 23 validated ncRNAs (Supplementary Table 8) against the *M. mazei* genome using blastall version 2.2.26 with word size set to 7 and a non-limiting E-value. The blast results were compared to the set of term-seq annotated 3′ UTRs, using the longest isoform for the comparison (Supplementary Table 9).

Accession codes. The sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession no. PRJEB14292.

Received 21 April 2016; accepted 14 July 2016;
published 22 August 2016

References

1. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: the RNA 3′-end chronicles. *J. Mol. Biol.* **412,** 793–813 (2011).
2. Porrua, O. & Libri, D. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16,** 190–202 (2015).
3. Peters, J. M. *et al.* Rho and NusG suppress pervasive antisense transcription in *Escherichia coli*. *Genes Dev.* **26,** 2621–2633 (2012).
4. Mellin, J. R. & Cossart, P. Unexpected versatility in bacterial riboswitches. *Trends Genet.* **31,** 150–156 (2015).
5. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8,** R239 (2007).
6. Santangelo, T. J. & Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* **9,** 319–329 (2011).
7. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol. Cell* **3,** 495–504 (1999).
8. Santangelo, T. J. & Reeve, J. N. Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J. Mol. Biol.* **355,** 196–210 (2006).
9. Brenneis, M., Hering, O., Lange, C. & Soppa, J. Experimental characterization of *cis*-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet.* **3,** e229 (2007).
10. Spitalny, P. & Thomm, M. A polymerase III-like reinitiation mechanism is operating in regulation of histone expression in archaea. *Mol. Microbiol.* **67,** 958–970 (2008).
11. Hirtreiter, A., Grohmann, D. & Werner, F. Molecular mechanisms of RNA polymerase-the F/E (RPB4/7) complex is required for high processivity *in vitro*. *Nucleic Acids Res.* **38,** 585–596 (2009).
12. Santangelo, T. J., Cubonová, L., Skinner, K. M. & Reeve, J. N. Archaeal intrinsic transcription termination *in vivo*. *J. Bacteriol.* **191,** 7102–7108 (2009).
13. Reiter, W. D., Palm, P. & Zillig, W. Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. *Nucleic Acids Res.* **16,** 2445–2459 (1988).
14. Dar, D. *et al.* Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352,** aad9822 (2016).
15. Taboada, B., Ciria, R., Martinez-Guerrero, C. E. & Merino, E. ProOpDB: Prokaryotic Operon DataBase. *Nucleic Acids Res.* **40,** D627–D631 (2012).
16. Jäger, D. *et al.* Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. *Proc. Natl Acad. Sci. USA* **106,** 21878–21882 (2009).
17. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6,** 26 (2011).
18. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14,** 1188–1190 (2004).
19. Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of bacterial transcription termination: all good things must end. *Annu. Rev. Biochem.* **85,** 319–347 (2016).
20. Wurtzel, O. *et al.* Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. *Mol. Syst. Biol.* **8,** 583 (2012).
21. Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R. & Cossart, P. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat. Rev. Microbiol.* **11,** 75–82 (2012).
22. Georg, J. & Hess, W. R. *cis*-Antisense RNA, another level of gene regulation in bacteria. *Microbiol. Mol. Biol. Rev.* **75,** 286–300 (2011).
23. Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43,** 853–866 (2011).
24. Jun, S. *et al.* The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat. Commun.* **5,** 1–11 (2014).
25. Werner, F. & Grohmann, D. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* **9,** 85–98 (2011).
26. Koide, T. *et al.* Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol. Syst. Biol.* **5,** 285 (2009).
27. Ruiz de los Mozos, I. *et al.* Base pairing interaction between 5′- and 3′- UTRs controls icaR mRNA translation in *Staphylococcus aureus*. *PLoS Genet.* **9,** e1004001 (2013).
28. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by small RNAs in bacteria: expanding frontiers. *Mol. Cell* **43,** 880–891 (2011).
29. Wurtzel, O. *et al.* A single-base resolution map of an archaeal transcriptome. *Genome Res.* **20,** 133–141 (2010).
30. Pain, A. *et al.* An assessment of bacterial small RNA target prediction programs. *RNA Biol.* **12,** 509–513 (2015).
31. Jäger, D., Förstner, K. U., Sharma, C. M., Santangelo, T. J. & Reeve, J. N. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics* **15,** 684 (2014).
32. Deppenmeier, U., Blaut, M., Mahlmann, A. & Gottschalk, G. Reduced coenzyme F420: heterodisulfide oxidoreductase, a proton-translocating redox system in methanogenic bacteria. *Proc. Natl Acad. Sci. USA* **87,** 9449–9453 (1990).
33. Weidenbach, K., Ehlers, C. & Schmitz, R. A. The transcriptional activator NrpA is crucial for inducing nitrogen fixation in *Methanosarcina mazei* Gö1 under nitrogen-limited conditions. *FEBS J.* **281,** 3507–3522 (2014).
34. Brock, T. D., Brock, K. M., Belly, R. T. & Weiss, R. L. Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. für Mikrobiol.* **84,** 54–68 (1972).
35. Cohen, O. *et al.* Comparative transcriptomics across the prokaryotic tree of life. *Nucleic Acids Res.* **44,** W46–W53 (2016).

Acknowledgements

The authors thank M. Shamir, S. Doron, A. Millman and A. Lopatina for discussions. R.S. was supported, in part, by the Israel Science Foundation (personal grant no. 1303/12 and I-CORE grant no. 1796/12), the ERC-StG programme (grant no. 260432), the Abisch–Frenkel Foundation, the Pasteur–Weizmann council grant, the Minerva Foundation, the Leona M.

and Harry B. Helmsley Charitable Trust and by a DIP grant from the Deutsche Forschungs Gemeinschaft (DFG). D.P. was funded by the DFG (Schm1052/9-2).

Author contributions

D.D., R.A.S. and R.S conceived and designed the research studies. D.D. and D.P. performed the experiments. D.D. and R.S. analysed the data. D.D., RA.S., D.P. and R.S. wrote the manuscript.

Additional information

Supplementary information is available for this paper. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.S.

Competing interests

The authors declare no competing financial interests.