functional coupling. *Proc. Natl Acad. Sci. USA* **96,** 2896–2901 (1999).
28. Haselbeck, R. *et al.* Comprehensive essential gene identification as a platform for novel anti-infective drug discovery. *Curr. Pharm. Des.* **8,** 1155–1172 (2002).
29. Krogan, N. J. *et al.* High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* **13,** 225–239 (2004).
30. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements The authors wish to thank C. J. Ingles and M. Shales for comments on the manuscript. This work was supported by funds from the Ontario Research and Development Challenge Fund and Genome Canada to A.E. and J.G. G.B. was a recipient of a Charles H. Best Post-Doctoral Fellowship. J.M.P.-A. acknowledges support from the Hospital for Sick Children (Toronto, Ontario, Canada) Research Training Centre. Computer analyses were undertaken at the Centre for Computational Biology, Hospital for Sick Children.

Authors' contributions Informatics studies were performed and analysed by J.M.P.-A. and J.P. Experimental design and data analysis were coordinated by G.B. Tagging and purification experiments were performed by W.Y., X.Y., J.L. and G.B. V.C., A.S., D.R., B.B., N.J.K. and M.D. performed and assisted with mass spectrometry analysis. The manuscript was jointly drafted by G.B., A.E., J.G., J.M.P.-A. and J.P. The project was conceived and designed by J.G. and was directed by A.E.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.E. (andrew.emili@utoronto.ca) or J.G. (jack.greenblatte@utoronto.ca).

Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5′- and 3′-halves

Lennart Randau[1,2], Richard Münch[2], Michael J. Hohn[1,3], Dieter Jahn[2] & Dieter Söll[1,4]

[1]*Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA*
[2]*Institut für Mikrobiologie, Technical University Braunschweig, P.O. Box 3329, D-38023 Braunschweig, Germany*
[3]*Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany*
[4]*Department of Chemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA*

Analysis of the genome sequence of the small hyperthermophilic archaeal parasite *Nanoarchaeum equitans*[1,2] has not revealed genes encoding the glutamate, histidine, tryptophan and initiator methionine transfer RNA species. Here we develop a computational approach to genome analysis that searches for widely separated genes encoding tRNA halves that, on the basis of structural prediction, could form intact tRNA molecules. A search of the *N. equitans* genome reveals nine genes that encode tRNA halves; together they account for the missing tRNA genes. The tRNA sequences are split after the anticodon-adjacent position 37, the normal location of tRNA introns. The terminal sequences can be accommodated in an intervening sequence that includes a 12–14-nucleotide GC-rich RNA duplex between the end of the 5′ tRNA half and the beginning of the 3′ tRNA half. Reverse transcriptase polymerase chain reaction and aminoacylation experiments of *N. equitans* tRNA demonstrated maturation to full-size tRNA and acceptor activity of the $tRNA^{His}$ and $tRNA^{Glu}$ species predicted *in silico*. As the joining mechanism possibly involves tRNA *trans*-splicing, the presence of an intron might have been required for early tRNA synthesis.

The origin of the tRNA molecule is the subject of continuing discussions and has led to different models postulating that tRNA evolved by duplication or ligation of an RNA hairpin[3,4]. To examine these models further, the investigation of ancient tRNA genes was central. An interesting organism for this task was *Nanoarchaeum equitans*, currently the only characterized member of the kingdom Nanoarchaeota, which roots early in the archaeal lineage, before the emergence of Euryarchaeota and Crenarchaeota[5]. A significant fraction of the small number of *N. equitans* open reading frames consists of 'split genes' that are encoded as fused versions in other archaeal genomes. Our attention was caught by the 'absence' of four tRNA genes encoding the glutamate, histidine, tryptophan and initiator methionine acceptors[5].

We therefore developed a computational approach to search for tRNA signature sequences in the *N. equitans* genome. Our program, trained by an alignment of 4,000 tRNA gene sequences (taken from ref. 6), identifies sequences comprising the highly conserved T-loop region and defines the adjacent 3′-acceptor stem sequence. The reverse complementary sequence (defining the 5′-acceptor stem sequence) plus a D-stem position weight matrix identifies the corresponding 5′ half. The length of the position weight matrices can be adjusted and mismatches in the acceptor stem can be included. Finally, putative tRNA-halves are ligated *in silico* and analysed by COVE[7]. In addition to identifying the set of tRNAs predicted by the tRNAScan-SE program[8], our algorithm found nine tRNA halves spread throughout the chromosome. Surprisingly, these tRNA halves could be joined *in silico* to form the missing $tRNA^{His}$, $tRNA_i^{Met}$, $tRNA^{Trp}$ and two $tRNA^{Glu}$ species (Fig. 1). Further analysis of the tRNA half genes revealed several striking features. First, the location of the sequence separation that generated all nine tRNA half genes is after position 37, one nucleotide downstream of the anticodon and the common location of tRNA introns[9]. Second, a consensus sequence matching the highly conserved archaeal Box A promoter element[10] was found upstream of all 5′ tRNA halves. Third, this same consensus sequence (5′-TTTT/ATAAA-3′) was located 17–25 base pairs (bp) further upstream of the 3′ tRNA halves, resulting in a transcript with a 12–14-bp-long GC-rich leader sequence. Last, it is remarkable that this leading sequence is in all cases the exact reverse complement to a sequence following the corresponding 5′ tRNA half.

The existence of three $tRNA^{Glu}$ half genes was most exciting. Two 5′ tRNA halves were identified that differed solely by one anticodon base (isoacceptors with UUC and CUC anticodon), whereas only one 3′ $tRNA^{Glu}$ half gene was found. Both 5′ $tRNA^{Glu}$ half genes were followed by the identical 14-bp sequence that was the exact reverse complement of the single 3′ $tRNA^{Glu}$ half upstream sequence. All identified split tRNA genes contained the consensus bases of all archaeal elongator tRNAs[6], namely U8, A14, G15, G18, G19, C32, U33 and the T-loop GTTCA/GAATC (53–61), with the exception of the putative $tRNA^{Trp}$ harbouring an unusual GG sequence preceding the anticodon. The identified $tRNA_i^{Met}$ displays the consensus sequences of archaeal initiator tRNAs such as the anticodon stem/loop nucleotides (nt) 29–41 (GGGCU-CAUAACCC) and the R11:Y24 base pair (G11:C24), which is the reverse of the Y11:R24 base pair found in elongator tRNAs including the annotated *N. equitans* $tRNA^{Met}$. Therefore we define the split $tRNA_i^{Met}$ as the missing initiator tRNA. The sequences also reveal characteristic nucleotides in the respective tRNA species needed for recognition by the cognate aminoacyl-tRNA synthetase. For example, the $tRNA^{His}$ half genes encode the unique G-1:C73 base pair required for aminoacylation of $tRNA^{His}$ by histidyl-tRNA synthetase[11], and the $tRNA^{Glu}$ isoacceptors contain the characteristic D-loop nucleotides 20a and 20b and the deletion of base 47 essential for making the 'augmented D-helix'[12].

We performed reverse transcriptase polymerase chain reaction (RT–PCR) analysis of *N. equitans* total tRNA to verify the computationally predicted sequence of the newly discovered joined tRNAs. Our sequencing results confirmed the sequences for $tRNA^{Glu}$

(UUC), $tRNA^{Glu}$ (CUC) and $tRNA_i^{Met}$ (Fig. 2a). Despite extensive efforts we could not amplify the full-length $tRNA^{Trp}$ and $tRNA^{His}$ (even though its existence was shown by aminoacylation; see below); this might have been due to the extreme thermostability of the GC-rich *N. equitans* tRNAs containing modified nucleosides. Nevertheless, we confirmed the presence of six tRNA half transcripts by RT–PCR and sequence analysis (Fig. 2d). The primary transcripts of these tRNA half genes include the intervening complementary sequences at the position of separation. In addition, RT–PCR of anchor-ligated tRNA (Fig. 2c) revealed that the primary transcript of the 5′ $tRNA^{His}$ half terminates at the AT-rich region following the complementary downstream sequence found in all tRNA half genes.

For a tRNA to participate in protein biosynthesis it must carry a 3′-terminal CCA sequence to which the amino acid will be esterified. In *N. equitans* and most Archaea, this CCA sequence is not encoded in the tRNA genes (including the split tRNA genes) but is added post-transcriptionally by the ATP(CTP):tRNA nucleotidyltransferase[13,14], an enzyme probably encoded by the still uncharacterized NEQ152 gene. By using a RT–PCR approach involving circularization of the tRNA[15] we were able to identify the 5′ and 3′ ends of the mature tRNA. Our sequencing results show size-maturation of the joined $tRNA^{Glu}$, as a CCA sequence is indeed added to the 3′ end of both $tRNA^{Glu}$ isoacceptors after transcription (Fig. 2b). A final requirement for tRNA functionality *in vivo* is the ability to serve as a substrate for amino acid attachment by

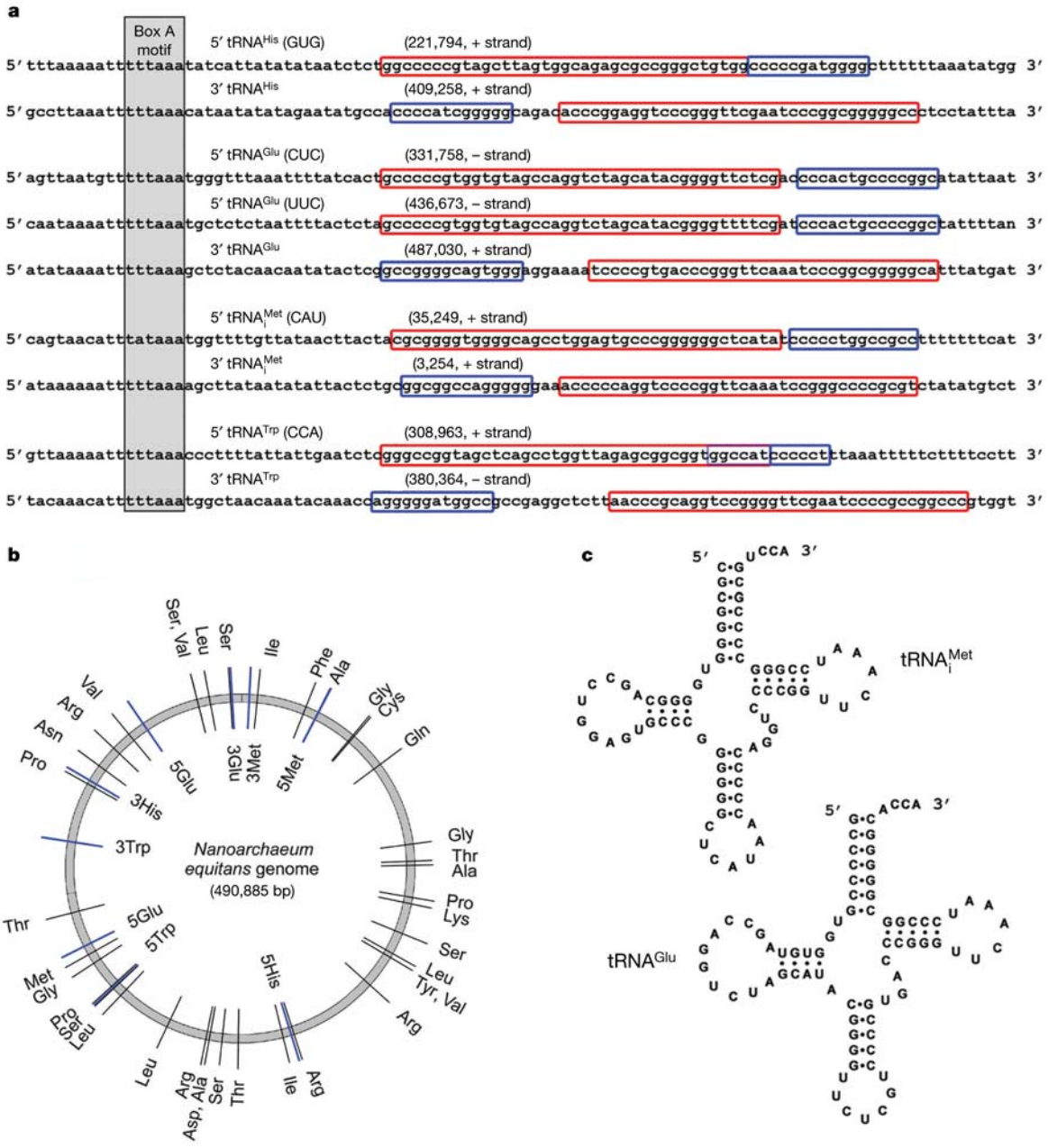


Figure 1 Predicted split *N. equitans* tRNA genes. **a**, tRNA half genes. The archaeal RNA polymerase III promoter consensus box A motif, the tRNA half genes (red) and intervening reverse complementary sequences (blue) are indicated. The positions of the tRNA halves in the *N. equitans* genome and the strand are indicated. **b**, Schematic representation of the genomic distribution of tRNA genes (indicated by the amino-acid three-letter code) and tRNA half genes (5 indicates the 5′ tRNA half gene, 3 indicates the 3′ tRNA half gene) identified by our search algorithm. **c**, The joined sequences of $tRNA^{Glu}$ and $tRNA_i^{Met}$ as verified by RT–PCR and sequencing (see Fig. 2).

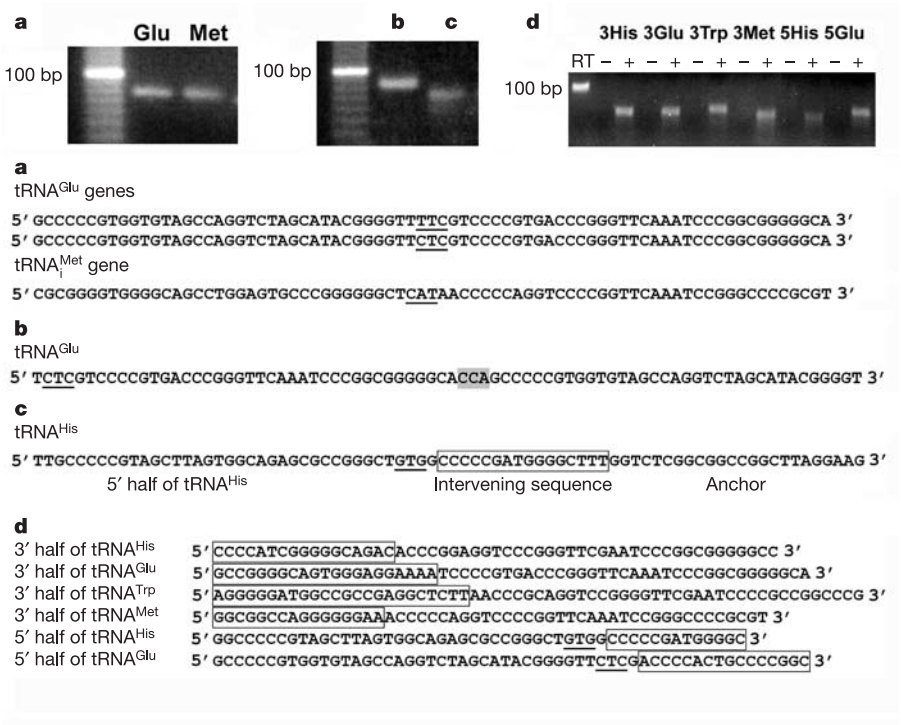


Figure 2 RT–PCR amplification of the newly identified tRNAs in *N. equitans*. The RT–PCR products were separated on a 4% ethidium bromide-stained acrylamide gel and compared with a 10-bp ladder marker (the 100-bp fragment is indicated). For RT–PCR of the tRNA halves, control experiments without reverse transcriptase (RT −) were performed to ensure that solely RNA transcripts were amplified (RT +). The different RT–PCR techniques employed are detailed in Methods. Sequencing results are displayed with the anticodon underlined. The post-transcriptionally edited CCA terminus is shaded. The intervening complementary regions of the primary tRNA half transcripts are boxed. **a**, Detection of full-length tRNA by RT–PCR; **b**, detection of 3′ maturation (CCA addition) by RT–PCR of circularized tRNA; **c**, detection of primary transcripts by RT–PCR of anchor-ligated tRNA; **d**, detection of primary transcripts by RT–PCR.

aminoacyl-tRNA synthetases. Aminoacylation reactions were performed to verify acceptor activity of the joined tRNAs. For this reason, the *N. equitans* genes encoding histidyl-tRNA synthetase (HisRS) and glutamyl-tRNA synthetase were cloned. The two enzymes were produced in *Escherichia coli*, and HisRS was purified by flocculation at 80 °C. Both synthetases were active and were able to acylate total *N. equitans* tRNA with their cognate ^{14}C-labelled amino acids (Fig. 3). Direct proof of the identity of the aminoacylated tRNA was obtained by northern blot analysis of acid/urea gels[16] after the separation of Glu-tRNA and His-tRNA from deacylated tRNA due to a difference in electrophoretic mobility between the two species (Fig. 3). The oligonucleotide probes for hybridization were complementary to a region comprising the anticodon stem/loop of the full-length tRNA^{His} and tRNA^{Glu}. These results strongly indicate an active role of these mature joined tRNAs in protein biosynthesis.

The 'missing' *N. equitans* tRNAs identified here reveal the necessity for assembly of two tRNA half molecules. What is the mechanism of joining these tRNA halves? We propose a model based on the discovery of extended reverse complementary intervening sequences (Fig. 4). Earlier studies showed that a GC-rich intervening RNA duplex increases the efficiency of intermolecular splicing (*trans*-splicing) of mRNA precursors *in vitro*[17,18]. *Trans*-splicing *in vivo* occurs in several plant mitochondrial transcripts encoding subunits of the NADH dehydrogenase complex. In most cases the exon-flanking regions form a group II intron structure. An extended GC-rich duplex in the split intron is thought to facilitate base pairing of the two intron halves[19]. Similarly, during *trans*-splicing of the *N. equitans* tRNA halves a 12–14-bp fully paired RNA duplex in the intervening sequences would be the primary nucleation region in annealing the two RNA sequences. This duplex would

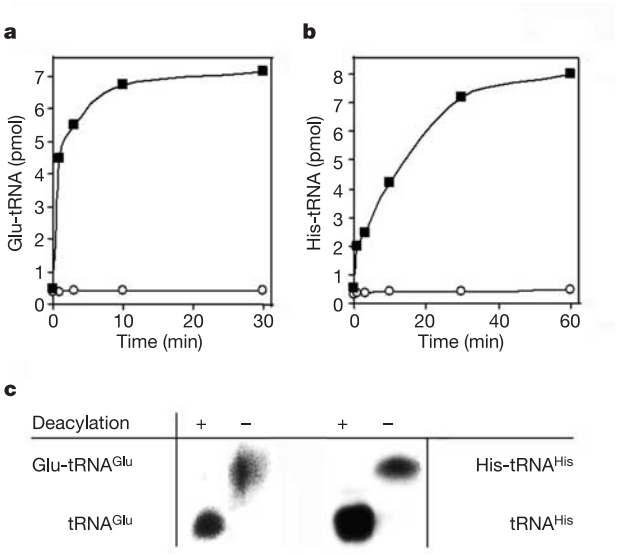


Figure 3 Aminoacylation of unfractioned *N. equitans* tRNA by *N. equitans* glutamyl-tRNA synthetase and histidyl-tRNA synthetase. The purification and aminoacylation procedures are detailed in Methods. **a**, Glutamylation with *N. equitans* glutamyl-tRNA synthetase (filled squares) at 37 °C. Open circles, control reactions without tRNA. **b**, Histidylation with *N. equitans* histidyl-tRNA synthetase (filled squares) at 80 °C. Open circles, control reactions without tRNA. **c**, Northern blot analysis of *N. equitans* tRNA glutamylated by glutamyl-tRNA synthetase or histidylated by HisRS for 1 h reaction time. The resulting aminoacyl-tRNA was loaded onto an acid/urea gel directly (−) or after deacylation (+). The blots were probed with a ^{32}P-labelled oligonucleotide spanning the joined anticodon stem/loop.

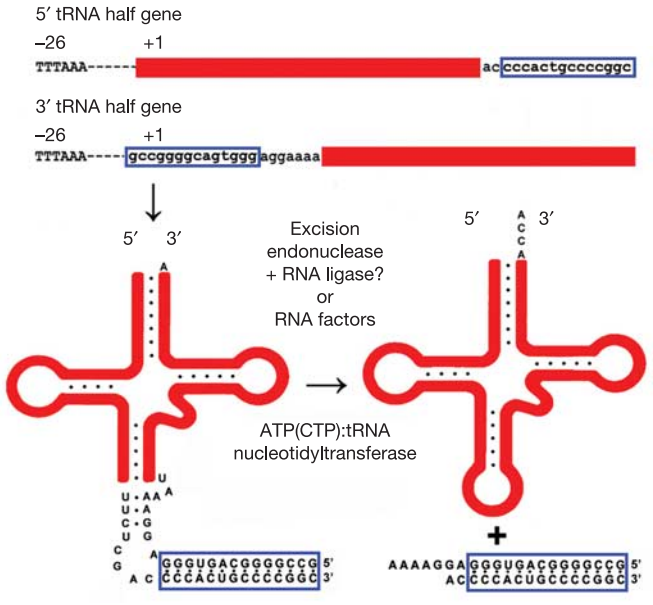


Figure 4 Schematic representation of a 5′ tRNA half gene ($tRNA^{Glu}$) and the corresponding 3′ tRNA half gene found in *N. equitans*. The archaeal RNA polymerase III promoter consensus sequence (TTTAAA), the tRNA half genes (red) and the intervening reverse complementary sequences that are supposed to facilitate joining of the halves (blue) are indicated.

facilitate folding of the whole tRNA body and stabilize the cloverleaf structure of the tRNA. It should be noted that for the joined $tRNA^{His}$ and $tRNA^{Glu}$ the region between the folded tRNA and the intervening RNA duplex resembles the consensus bulge–helix–bulge motif structure described for archaeal and eukaryotic tRNA introns[20,21]. Because this structure is located at the position where most archaeal tRNA introns occur, a similar mechanism of tRNA maturation is possible. In this case one of the two putative tRNA splicing endonucleases (NEQ205 and NEQ261)[22,23] might be responsible for intermolecular tRNA splicing, and an RNA ligase would join the 5′ and 3′ tRNA half molecules[24]. A second possibility is an RNA-mediated *trans*-splicing mechanism. Both possibilities will be investigated.

Why does *N. equitans* employ this strategy? The advantage of tRNA *trans*-splicing is not apparent, given the small size of a tRNA gene. What is remarkable is the finding that one single exon (3′ $tRNA^{Glu}$ half) is *trans*-spliced to two exons (5′ $tRNA^{Glu}_{UUC}$ half and 5′ $tRNA^{Glu}_{CUC}$ half). It has been suggested[4] that in the pre-biotic world two RNA hairpins had the simplest RNA structure and folded by given similarity into a cloverleaf-like structure. After the birth of the cloverleaf shape some template RNAs would evolve into ancient tRNAs. The intervening complementary sequences of the *N. equitans* split tRNAs might indicate an intermediate state in which the hairpins are still separated and have to be joined and stabilized by a GC-rich duplex. It is possible that the 3′ tRNA half comprised of a highly conserved T-loop minigene could be fused to various anticodon-containing 5′ tRNA halves to satisfy the growing complexity of protein translation during evolution[25,26]. Thus, introns in modern tRNAs might be remnants of this duplex from an earlier world, which still performs its function in *N. equitans*.

Is this the only example of split genes for stable RNAs? Although we do not know of any other occurrences, it should be noted that the *N. equitans* genome does not possess an orthologue of the RNA component of RNase P[27] nor any recognizable fragment genes. It remains to be determined whether the *N. equitans* tRNA transcripts contain leader sequences upstream of the 5′ end of tRNA. In their absence, there would not be a need for the presence of RNase P in the organism.

An extensive search of the available bacterial and archaeal genome sequences did not reveal split tRNA genes in other organisms. Future sequences of other very small or compact genomes should reveal whether split tRNA genes are signs of a very early genome[2] or whether they are created in a later process of genome size reduction. The sequencing of other Nanoarchaeota genomes is therefore eagerly awaited. □

Methods

Computational method for tRNA identification

tRNA genes were predicted by use of a new bioinformatics approach and the program Virtual Footprint (http://www.prodoric.de/sts/). Position weight matrices were generated from both a conserved, continuous 3′ region of tRNA genes (nt 54–76) and a 5′ region of tRNA genes (nt 1–16) in an alignment of more than 4,000 tRNA genes (taken from ref. 6). For this purpose the information content was used as scoring function[28] with slight modifications. tRNA gene searches were performed with these position weight matrices on a genome scale with the highest sensitivity (the threshold score was taken from the lowest scoring sequence of the training set). The information that the 3′ region contains a pairing stretch of 7 nt to a reverse complementary part in the 5′ region (the tRNA's acceptor stem) was used to identify matching pairs of tRNA gene halves. Using this approach, all the previously annotated tRNAs were identified, including nine additional tRNA halves that fell into the threshold range of the annotated tRNAs.

Cell culture and tRNA isolation

N. equitans cells were grown in a 300-l fermenter in a simultaneous culture with *Ignicoccus* sp. and purified by gradient centrifugation as described[1]. The cell pellet was lysed by chemical digestion with 2% SDS, followed by the isolation and purification of total RNA as described[29]. The tRNA was further purified by MonoQ HR 5/5 anion-exchange chromatography to eliminate residual genomic DNA contamination. The tRNA was eluted with a linear 60-ml gradient of 0–1 M NaCl in 20 mM MOPS pH 6.2.

Reverse transcription and sequencing

Total tRNA from *N. equitans* was reverse transcribed with Thermoscript reverse transcriptase and PCR amplified with Platinum Taq DNA polymerase (Invitrogen) according to the manufacturer's directions. The tRNA was denatured at 100 °C for 5 min and snap-cooled on ice for 5 min to facilitate transcription through the highly stable secondary structure of the tRNA. PCR products were cloned with the pCR-2.1-TOPO cloning kit (Invitrogen). Plasmids were sequenced at the W. M. Keck Facility. The following oligonucleotides were used for PCR amplification of the indicated full-length tRNAs: 5′-TGCCCCCGCCGGATTTGAACC-3′ and 5′-GCCCCCGTGGTGTAGCCAGG TCTAGC-3′ ($tRNA^{Glu}$ (UUC), $tRNA^{Glu}$ (CUC)); 5′-ACGCGGGGCCCGGATTTGAACC-3′ and 5′-CGCGGGGTGGGGCAGCCTGGAGTGC-3′ ($tRNA_i^{Met}$). The reverse primers for RT–PCR of the 3′ tRNA halves (Fig. 2d) comprised 22 nt, whereas the forward primers were 17 nt ($tRNA^{His}$, $tRNA^{Glu}$) or 18 nt ($tRNA^{Trp}$, $tRNA^{Met}$) long. The reverse primers for RT–PCR of the 5′ tRNA halves comprised 26 nt and the forward primers were 17 nt ($tRNA^{His}$) or 19 nt ($tRNA^{Glu}$) long.

A different RT–PCR approach uses total *N. equitans* tRNA circularized by RNA-ligase-mediated joining of the 5′ and 3′ ends of a tRNA as described previously[15]. Circularized $tRNA^{Glu}$ was PCR amplified, cloned and sequenced as described above by using the oligonucleotides 5′-CGAGAACCCCGTATGCTAGACCTGGCTACAC-3′ and 5′-TCTCGTCCCCGTGACCCGGGTTCAAATCCC-3′. In a third RT–PCR approach an anchor oligonucleotide (5′-pGGTCTCGGCGGCCGGCTTAGGddC-3′) was ligated to total *N. equitans* tRNA as described[30] and cDNA clones were produced and sequenced as described above. Two oligonucleotides were used for PCR amplification of the anchor-ligated 5′ $tRNA^{His}$ half: 5′-GCCCCCGTAGCTTAGTGGCAGAG-3′ and 5′-CCTAAGCC GGCCGCCGAGACC-3′.

Preparation of proteins and aminoacylation assay

N. equitans hisS (NEQ102) and *gltX* (NEQ302) genes were amplified by PCR from genomic DNA. The genes were cloned into the *NdeI*/*EcoR*I site of pET12b(+) (Invitrogen) for *hisS* and into the *EcoR*V site of petBlue (Invitrogen) for *gltX*, to facilitate expression of the proteins in the *E. coli* BL21-codon plus (DE3)-RIL strain (Stratagene). Cultures were grown at 37 °C in Luria–Bertani medium supplemented with 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol. Expression of the recombinant proteins was induced for 4 h at 37 °C by the addition of 1 mM isopropyl α-D-thiogalactopyranoside before cell harvesting. Cells were resuspended in buffer containing 50 mM Tris-HCl pH 7.5 and 300 mM NaCl, and broken by sonication. The fractions were extensively flocculated at 80 °C for 30 min, then centrifuged for 30 min at 20,000 *g*. Aminoacylation was performed in a 0.1 ml reaction at 80 °C in 50 mM Hepes pH 7.0, 50 mM KCl, 4 mM ATP, 15 mM $MgCl_2$, 3 mM dithiothreitol, 5 μg total *N. equitans* tRNA, 50 μM [^{14}C]glutamate (256 mCi $mmol^{-1}$; 9.47 GBq $mmol^{-1}$) or 50 μM [^{14}C]histidine (314 mCi $mmol^{-1}$; 11.6 GBq $mmol^{-1}$), and the aminoacyl-tRNA synthetase. Glutamyl-tRNA synthetase activity assays revealed advanced activity at 37 °C reaction temperature. Aliquots of 20 μl were removed at the intervals indicated in Fig. 3, and radioactivity was measured as described[29]. Separation of tRNA by acid/urea gel electrophoresis (9.6% gel run for 40 h ($tRNA^{Glu}$) and 6.4% gel run for 22 h ($tRNA^{His}$)) and electroblotting onto Hybond N^+

membrane (Amersham Biosciences) were performed as described[16]. Northern analysis was performed with ^{32}P-labelled oligonucleotides complementary to bases 12–50 of *N. equitans* tRNA^{Glu} and bases 11–50 of *N. equitans* tRNA^{His}.

Received 30 August; accepted 2 December 2004; doi:10.1038/nature03233.

1. Huber, H. *et al.* A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417,** 63–67 (2002).
2. Waters, E. *et al.* The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl Acad. Sci. USA* **100,** 12984–12988 (2003).
3. Di Giulio, M. The non-monophyletic origin of the tRNA molecule. *J. Theor. Biol.* **197,** 403–414 (1999).
4. Tanaka, T. & Kikuchi, Y. Origin of the cloverleaf shape of transfer RNA—the double-hairpin model: Implication for the role of tRNA intron and the long extra loop. *Viva Origino* **29,** 134–142 (2001).
5. Boucher, Y. & Doolittle, W. F. Something new under the sea. *Nature* **417,** 27–28 (2002).
6. Marck, C. & Grosjean, H. tRNomics: analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8,** 1189–1232 (2002).
7. Eddy, S. R. & Durbin, R. RNA sequence analysis using covariance models. *Nucleic Acids Res.* **22,** 2079–2088 (1994).
8. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).
9. Marck, C. & Grosjean, H. Identification of BHB splicing motifs in intron-containing tRNAs from 18 archaea: evolutionary implications. *RNA* **9,** 1516–1531 (2003).
10. Hain, J., Reiter, W. D., Hudepohl, U. & Zillig, W. Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res.* **20,** 5423–5428 (1992).
11. Connolly, S. A., Rosen, A. E., Musier-Forsyth, K. & Francklyn, C. S. G-1:C73 recognition by an arginine cluster in the active site of *Escherichia coli* histidyl-tRNA synthetase. *Biochemistry* **43,** 962–969 (2004).
12. Sekine, S. *et al.* Major identity determinants in the 'augmented D helix' of tRNA(Glu) from *Escherichia coli*. *J. Mol. Biol.* **256,** 685–700 (1996).
13. Schurer, H., Schiffer, S., Marchfelder, A. & Mörl, M. This is the end: processing, editing and repair at the tRNA 3′-terminus. *Biol. Chem.* **382,** 1147–1156 (2001).
14. Xiong, Y., Li, F., Wang, J., Weiner, A. M. & Steitz, T. A. Crystal structures of an archaeal class I CCA-adding enzyme and its nucleotide complexes. *Mol. Cell* **12,** 1165–1172 (2003).
15. Lohan, A. J. & Gray, M. W. Methods for analysis of mitochondrial tRNA editing in *Acanthamoeba castellanii*. *Methods Mol. Biol.* **265,** 315–332 (2004).
16. Varshney, U., Lee, C. P. & RajBhandary, U. L. Direct analysis of aminoacylation levels of tRNAs *in vivo*. Application to studying recognition of *Escherichia coli* initiator tRNA mutants by glutaminyl-tRNA synthetase. *J. Biol. Chem.* **266,** 24712–24718 (1991).
17. Konarska, M. M., Padgett, R. A. & Sharp, P. A. *Trans* splicing of mRNA precursors *in vitro*. *Cell* **42,** 165–171 (1985).
18. Solnick, D. Does *trans* splicing *in vitro* require base pairing between RNAs? *Cell* **44,** 211 (1986).
19. Wissinger, B., Schuster, W. & Brennicke, A. *Trans* splicing in *Oenothera* mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group II intron sequences. *Cell* **65,** 473–482 (1991).
20. Abelson, J., Trotta, C. R. & Li, H. tRNA splicing. *J. Biol. Chem.* **273,** 12685–12688 (1998).
21. Fabbri, S. *et al.* Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. *Science* **280,** 284–286 (1998).
22. Li, H. & Abelson, J. Crystal structure of a dimeric archaeal splicing endonuclease. *J. Mol. Biol.* **302,** 639–648 (2000).
23. Kleman-Leyer, K., Armbruster, D. W. & Daniels, C. J. Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell* **89,** 839–847 (1997).
24. Salgia, S. R., Singh, S. K., Gurha, P. & Gupta, R. Two reactions of *Haloferax volcanii* RNA splicing enzymes: joining of exons and circularization of introns. *RNA* **9,** 319–330 (2003).
25. Maizels, N. & Weiner, A. M. Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl Acad. Sci. USA* **91,** 6729–6734 (1994).
26. Nagaswamy, U. & Fox, G. E. RNA ligation and the origin of tRNA. *Orig. Life Evol. Biosph.* **33,** 199–209 (2003).
27. Gopalan, V., Vioque, A. & Altman, S. RNase P: variations and uses. *J. Biol. Chem.* **277,** 6759–6762 (2002).
28. Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **188,** 415–431 (1986).
29. Curnow, A. W., Tumbula, D. L., Pelaschier, J. T., Min, B. & Söll, D. Glutamyl-tRNA^{Gln} amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc. Natl Acad. Sci. USA* **95,** 12838–12843 (1998).
30. Williams, M. A., Johzuka, Y. & Mulligan, R. M. Addition of non-genomically encoded nucleotides to the 3′-terminus of maize mitochondrial mRNAs: truncated rps12 mRNAs frequently terminate with CCA. *Nucleic Acids Res.* **28,** 4444–4451 (2000).

Acknowledgements We thank H. Huber and K. O. Stetter for advice and spirited discussions, M. Thomm for the use of laboratory facilities, and J. Yuan and J. Sabina for critically reading the manuscript. This work was supported by grants from the National Institute of General Medical Sciences and the Department of Energy (to D.S.) and by the German Federal Ministry of Education and Research (BMBF) for the Bioinformatics Competence Center 'Intergenomics' (to D.J.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.S. (soll@trna.chem.yale.edu).

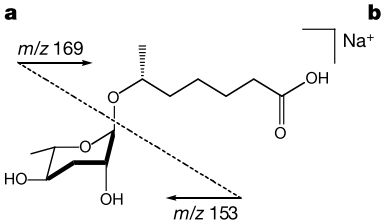
Chemical structure and biological activity of the *Caenorhabditis elegans* dauer-inducing pheromone

Pan-Young Jeong[1], Mankil Jung[2], Yong-Hyeon Yim[4], Heekyeong Kim[2], Moonsoo Park[2], Eunmi Hong[1], Weontae Lee[1], Young Hwan Kim[5], Kun Kim[3] & Young-Ki Paik[1]

[1]*Department of Biochemistry and Yonsei Proteome Research Center,* [2]*Department of Chemistry,* [3]*Bioproducts Research Center, Yonsei University, Seoul 120-749, Korea*
[4]*Korea Research Institute of Standards and Science, Taejon 305-600, Korea*
[5]*Korea Basic Science Institute, Taejeon 305-333, Korea*

Pheromones are cell type-specific signals used for communication between individuals of the same species. When faced with overcrowding or starvation, *Caenorhabditis elegans* secrete the pheromone daumone, which facilitates communication between individuals for adaptation to adverse environmental stimuli[1–4]. Daumone signals *C. elegans* to enter the dauer stage, an enduring and non-ageing stage of the nematode life cycle with distinctive adaptive features and extended life. Because daumone is a key regulator of chemosensory processes in development and ageing[5,6], the chemical identification of daumone is important for elucidating features of the daumone-mediated signalling pathway. Here we report the isolation of natural daumone from *C. elegans* by large-scale purification, as well as the total chemical synthesis of daumone. We present the stereospecific chemical structure of purified daumone, a fatty acid derivative. We demonstrate that both natural and chemically synthesized daumones equally induce dauer larva formation in *C. elegans* (N2 strain) and certain dauer mutants, and also result in competition between food and daumone. These results should help to elucidate the daumone-mediated signalling pathway, which might in turn influence ageing and obesity research and the development of antinematodal drugs.

Although general properties of partially enriched daumone have been known for more than two decades[2], its biochemical identity and other chemical characteristics have not yet been determined. Because there is no chemical method for confirming the presence of very low concentrations of daumone in crude cell extracts[2], we used a large-scale culture (with a 300-litre fermenter), and performed a dauer formation assay with every fraction at each purification step. Following two consecutive organic solvent extractions, the ethyl acetate extracts were loaded onto a silica gel column, which was then washed once with a solution of hexane/ethyl acetate/methanol



b

c
$C_{13}H_{24}O_6Na^+$ 299.1471	Used as the lock mass
$C_7H_{14}O_3Na^+$ 169.0841	169.0838 (1.6 p.p.m.)
$C_6H_{10}O_3Na^+$ 153.0528	153.0518 (6.3 p.p.m.)

Figure 1 Molecular structure of daumone: (−)-(6*R*)-(3,5-dihydroxy-6-methyltetrahydropyran-2-yloxy)heptanoic acid. **a**, Structure of daumone and the proposed fragmentation pattern for sodiated daumone, using *m*/*z* 299.1471 as the lock mass for mass calibration. **b**, Elemental composition and accurate mass calculated. **c**, Accurate mass of each fragment measured.