ARTICLE

The essential genome of the crenarchaeal model *Sulfolobus islandicus*

Changyi Zhang1,2, Alex P.R. Phillips1,2, Rebecca L. Wipfler1, Gary J. Olsen1,2 & Rachel J. Whitaker1,2

*Sulfolobus islandicus* is a model microorganism in the TACK superphylum of the Archaea, a key lineage in the evolutionary history of cells. Here we report a genome-wide identification of the repertoire of genes essential to *S. islandicus* growth in culture. We confirm previous targeted gene knockouts, uncover the non-essentiality of functions assumed to be essential to the *Sulfolobus* cell, including the proteinaceous S-layer, and highlight essential genes whose functions are yet to be determined. Phyletic distributions illustrate the potential transitions that may have occurred during the evolution of this archaeal microorganism, and highlight sets of genes that may have been associated with each transition. We use this comparative context as a lens to focus future research on archaea-specific uncharacterized essential genes that may provide valuable insights into the evolutionary history of cells.
Forty-one years ago, Woese and Fox identified the Archaea as a novel microbial lineage distinct from Bacteria1. The same year, Woese and Fox proposed a model of cellular evolution in which early cellular life diverged in two directions, one to the Bacteria and the other to LECA, the Last Eukaryotic and Archaeal Common Ancestor, which subsequently split forming the Archaea and Eukaryota domains2–5. Increases in genome and metagenome sequence data continue to refine this picture, providing reinforcement for many of its key aspects, improving phylogenetic sampling, and providing additional details6–12. With the addition of new lineages, the tree of life itself has evolved with great controversy13–17. Some of these new analyses suggest that the Thaumarcheota, Aigarchaeota, Crenarchaeota, and Korarchaeota (TACK) lineages of Archaea have the closest relationship to the Eukaryota with cultured representatives18–20.

Today, the tree of life provides a framework for studying the evolution of cellular complexity. Genomics and metagenomics provide data on the distribution of genes across this tree and in doing so provide an understanding of the origins and evolutionary dynamics of gene sequences. However, phylogenetic distributions fall short of establishing the functional evolutionary history of the cell since gene presence does not link directly to function. Mapping evolution of today’s complex cells involves a comparative approach in which functional cellular systems and the interactions of their constituent components are examined at a molecular level in organisms representing key evolutionary lineages across the tree of life.

As a step in that direction, we take here a genome-wide functional approach to define 441 genes essential to the growth of *Sulfolobus islandicus*. *Sulfolobus*, a thermooacidophilic genus from geothermal hot springs, is one of the few organisms within the TACK archaea that can be cultured and is genetically tractable20,21, and it is the most developed model for studying the biology of cells in this lineage. We find that the S-layer protein, which is present in most cells in the archaeal domain22, is not essential in this model organism. Furthermore, we explore potential stages of evolution of the essential gene repertoire of the archaeal cell.

**Results and Discussion**

**Identifying essential genes in the genome of *S. islandicus***. We established three independent genome-wide disruption libraries in an agmatine-auxotrophic strain of *S. islandicus* M16.4 by using a modified in vitro transposon mutagenesis system derived from Tn5 (Epicentre, USA). The transposable element was comprised of a nutritional marker cassette, SoargD (arгинine decarboxylase derived from *Sulfolobus sulfataricus* P2), flanked by two 19-base pair (bp) inverted repeats (Fig. 1a). After electroporation-mediated transformation of ArgD- cells with the EZ-Tn5 transposome, cells were allowed 10 days of growth on rich media. While valuable information about metabolic and regulatory genes could have been gained by comparing results from different media conditions, we restricted this study to one rich medium to focus on central cellular rather than metabolic functions. Insertion locations were determined via genome tagging and fragmentation (“tagmentation”) on colony pools, followed by amplification and sequencing of the junction sites, which were then mapped onto the genome. In all, 89,758 unique insertion events with at least 3 reads each were identified across all three libraries, corresponding to an average of 1 insertion every 29 bp and an average expected 29 insertions in each annotated protein-coding gene (see Methods; Supplementary Table 1 contains colony, insertion, and read counts for each library while all insertion locations can be found in Supplementary Data 1).

Essential genes were predicted to be significantly under-represented in the insertion locations extracted from the transposon mutagenesis and sequencing data (Tn-seq). It is important to note that this may make them indistinguishable from genes that are not strictly essential for growth but instead cause a severe growth defect, and thus our definition of “essential” extends to these genes too. To determine the statistical separation between essential and non-essential genes, we used a combination of two programs: ESSENTIALS23 and Tn-Seq Explorer24. Both methods report essential gene candidates by separating essential and non-essential genes into a bimodal distribution of scores. ESSENTIALS does so by calculating a log ratio of observed and expected reads in each gene (log(FC)), while Tn-Seq Explorer uses a sliding window approach to examine the absolute number of insertions in and around genes and calculates an Essentiality Index (EI) for each. The former tends to underestimate the number of essential genes, while the latter tends to overestimate24. Four hundred and forty-five genes lie within the suggested range for both methods (log(FC) ≤ −5.1 and EI < 4), leaving 175 genes within only one range or “unassigned” as essential or non-essential. Fourteen protein-encoding genes could not be assigned a log-FC score due to the presence of identical insertions in the genome (score “NA” in Supplementary Data 2). The remaining 2100 protein-coding genes are likely non-essential for growth under these conditions (Fig. 1b and Supplementary Data 2). Three genes identified as essential through automated methods were additionally removed because misplaced multiply mapped reads falsely reduced read count (M164_0862, M164_1012, and M164_1867; see Supplementary Table 2). Assignments of all genes to categories with their scores for each method are listed in Supplementary Data 2.

**Genetic confirmation of essential gene criteria.** To support our informatic essentiality/non-essentiality criteria, 129 genes were compared with gene knockout studies performed in our model *S. islandicus* M16.4 and another two genetically tractable *S. islandicus* strains: RYE15A and LAL 14/1 (Supplementary Table 3). We were unable to acquire knockouts for 42 of the 45 predicted essential genes in this set. Two exceptions, *topR2* (M164_1245) and *apt* (M164_0158), were identified to have significant growth defects on plates once they were knocked out (Supplementary Fig. 1c and 2a and ref.25), likely resulting in their under-representation in our transposon library. The third, *cdvB3* (M164_1510), a paralogue of *cdvB*, may be incorrectly called essential in our Tn-seq analysis. We can readily obtain *cdvB3* disruption mutants (Supplementary Fig. 3b) and the growth of a *cdvB3* mutant strain is indistinguishable from the wild-type strain (data not shown), thus this gene was removed from the essential gene list. An explanation of why this gene is mischaracterized would require further investigation, but it is possible that, because the score distributions for essential and non-essential genes overlap, this gene was simply not hit enough times to achieve significance. This could be true for a small number of other genes as well and is a fundamental limitation of Tn-seq.

To further investigate our automated assignments, we screened eight “unassigned” genes in *S. islandicus* M16.4 that were called essential by one method or the other but not both. We were unable to obtain mutants for six of them. Of these, five genes, i.e., *lig* (M164_1953), *priL* (M164_1568), *priX* (M164_1652), *rnhII* (M164_0197), and *tfs2* (M164_1524), were called essential via EI but not log(FC), while *thrS1* (M164_0290) was called essential based on log(FC) but not EI. In contrast, knockouts of the two “unassigned” genes called essential by EI but not log(FC), *udp4* (M164_0085), encoding uracil-DNA glycosylase family 4, and...
rpo8 (M164_1872), encoding a subunit of RNA polymerase, were obtained after an extended 14 days incubation of transformation plates, again consistent with a severe growth defect (Supplementary Fig. 2b, c and 3b). This suggests the presence of false negatives and a stronger bias to underestimate than overestimate the true number of essential genes. Because not all genes in the unassigned categories were genetically tested, we conservatively excluded all unassigned genes from the essential gene list. By contrast, knockouts for all 76 non-essential genes tested were successfully obtained and verified by PCR analysis (Supplementary Table 3 and Supplementary Fig. 3a and 3b). These include hjm/hel308a (M164_0269), cdvB1 (M164_1700), topR1 (M164_1732), and three DExD/H-box family helicase genes (M164_0809, M164_2103, and M164_2020), the homologs of which were previously thought to be essential in a related strain S. islandicus Rey15A26 (Supplementary Table 3 and Supplementary Fig. 3b). Taken together, these experimental results supported the overall validity of our computational approaches for conservatively classifying putative gene essentiality.

**Essential gene repertoire.** The functional repertoire of the predicted essential, unassigned, and non-essential genes of *S. islandicus* is shown in Fig. 2. With the above adjustments, the size of this essential genome (441 genes) is close in size to that observed for other bacteria and archaean 30. For example, ~526 genes are required for growth in *Methanococcus maripaludis* S231 and 473 genes within the engineered *Mycoplasma mycoides* JCVI Syn3.0 minimal bacterial cell 32. The proportion of different functional categories represented in this set (as defined by archaeal clusters of orthologous genes 12 (arCOGs)) are also similar to that observed in other studies 31,32 (Fig. 2), with the largest fraction of genes (178, ~40%) representing information processing (translation, transcription, and DNA replication/recombination/repair) and 76 (~17%) either classified as “function unknown” or “general functional prediction only.” The latter two categories are hereby collectively referred to as “poorly characterized.” Descriptions of the specific essential components found in central information processing and the cell cycle, as well as central carbon metabolism, are detailed in Supplementary
S-layer is non-essential in *S. islandicus*. Our essential gene predictions include several surprising findings. First, SlaA (M164_1763) and SlaB (M164_1762), the two known components of the surface layer (*S* of *Sulfolobus* cells) were shown to be non-essential. SlaA is the dominant component of the S-layer that forms a quasi-crystalline matrix outside the cell membrane. Current models suggest a “stalk-and-cap” structure in which the C-terminal-transmembrane-helix-domain-containing SlaB projects from the cell membrane and anchors SlaA to the cell membrane. The cellular function of the *Sulfolobus* S-layer is unknown but is believed to provide resistance to osmotic stress and contribute to cell morphology. S-layer-deficient mutants have never been successfully cultivated before in any archaeal species, therefore it was assumed to be essential.

To confirm the non-essentiality of the S-layer genes, we constructed in-frame deletion mutants of slaA, slaB, and slaAB via a MID (marker insertion and unmarked target gene deletion) recombination strategy. PCR amplification with two primer sets, which bind the flanking and internal region of S-layer genes, respectively (Fig. 3a), confirmed the successful deletion of slaA, slaB, and slaAB from the chromosome of the genetic host RJW004 (wild type) (Fig. 3b). We next tested for absence of the S-layer proteins in growing cells. Isolation of a white precipitant, described as the S-layer previously, was possible only in the wild type and to a much lesser extent in the ΔslaA mutant strain (Supplementary Fig. 4a and 4b). Transmission electron microscopy (TEM) analysis confirmed this extracted protein precipitate from both wild type and ΔslaB mutant strains formed crystalline lattice structures (Supplementary Fig. 4c). Finally, we tested the mutant phenotypes by comparing their growth profiles with wild type in a standard laboratory condition (pH 3.3, 76 °C). As shown in Fig. 3c, cells lacking the S-layer protein lattice SlaA (including slaA and slaAB mutants) are viable but have a measurable growth defect. This confirms the non-essentiality of the S-layer lattice in *S. islandicus*. The deletion of slaB alone had no significant impact on the growth rate in comparison with that of wild type (Fig. 3c). For a complete knockout of all potential S-layer components, we successfully created a viable triple knockout of slaA, slaB, and a paralog of SlaB encoded by M164_1049 (42% coverage, 53% amino acid identity via BLAST), demonstrating non-essentiality of all S-layer components together in *S. islandicus* (Supplementary Fig. 5).

We performed thin-section TEM analyses of the RJW004 (wild type) and S-layer gene knockout strains. The thin-section micrographs of wild-type cells clearly revealed that the S-layer was separated from the cytoplasmic membrane by a quasi-periplasmic space (Fig. 4a, e), in agreement with previous studies in *Sulfolobus acidocaldarius* and *Sulfobolus shibatae*. The S-layer in the wild type was observed as a distinct dark band on the outermost edge of the cell, and the quasi-periplasmic space was seen as a light gray band between the outermost band and the cell membrane (Fig. 4a, e). However, the dark, outermost layer surrounding the cell was not observed in the ΔslaA or ΔslaAB mutant cells (Fig. 4b, d, f, h), confirming that SlaA contributes to the formation of the outermost layer. Additionally, the cell surface appeared diffuse in the *slaA* mutant cell, which was attributed to the periodic extensions of membrane proteins, likely including the SlaB protein, and/or their extensive N-glycosylation.

Incomplete complementarity of reverse gyrase. As an additional surprise from our genome-wide essential gene identification, we found incomplete complementarity between two copies of the reverse gyrase in *S. islandicus* M16.4. Unlike Euryarchaeota and most extremely thermophilic bacteria, Crenarchaeota possess two copies of reverse gyrase, both believed to be essential for growth. Tn-seq analysis indicated that the topR1 (M164_1732) was non-essential, which was confirmed by a successful disruption (Supplementary Fig. 1a and 1b). Interestingly, as mentioned above, topR2 (M164_1245) was called essential but we could obtain topR2 disruption mutants (Supplementary Fig. 1a and 1c) if we prolonged the incubation time (up to 14–20 days) of transformation plates in gene knockout experiments. These observations suggest that topR2 plays a more important role than topR1 in *Sulfolobus* cell survival at optimal temperature.

Lethal deletion mutants. Tn-seq also uncovered genes that may not be essential to growth but instead are toxic when disrupted. Among them, arCOG analysis predicts that M164_0131, M164_0217, M164_0268, M164_2076, M164_1728, and M164_1060 are antitoxin-encoding genes. We reason that inactivation of these antitoxin genes might cause overproduction of toxins and then trigger cell death. This finding suggests that associated toxins are constitutively expressed in our laboratory conditions. Interestingly, unlike most of the family II (VapBC) and family HEPN-NT toxin/antitoxin gene pairs in *S. islandicus*
M.16.4 (Supplementary Data 3), partners (toxin genes) adjacent to these predicted antitoxin genes (with the exception of M164_1060; see Supplementary Fig. 6) were not observed. This indicates that VapB-VapC or HEPN-NT do not always correspond to their neighbors and some gene pairs might have exchanged counterparts. The Tn-seq-based analyses also classified cas5 (M164_0911), a part of the Cascade (CRISPR-associated complex for antiviral defense) complex43, as essential. Consistent with this assignment, disruption of cas5 by replacing it with the StaurD marker cassette via homologous recombination failed after repeated attempts. However, the entire Type-IA module of CRISPR-Cas system, consisting of eight genes with cas5 included, could be deleted from the S. islandicus M.16.4 chromosome with this assignment, disruption of slaAB with this assignment, disruption of slaAB mutants with two primer sets, which bind the flanking and internal regions of S-layer genes, respectively. Expected sizes of amplicons can be found in Supplementary Table 8.

**Shared essential genes.** To establish how this essential gene set compares with those found in other organisms, we retrieved sets of essential genes from the database of essential genes30,44 in eight model organisms that span the tree of life31,45–49, including the minimal genes set in the JCVI Syn 3.0 Mycoplasma mycoides genome12 (Fig. 5, Supplementary Data 4). We find that 242 S. islandicus essential genes are essential in at least 1 other organism we surveyed, while 199 essential genes are uniquely essential in S. islandicus. Eighty-nine genes are essential in representatives of all 3 domains, 78 of which are also essential in Syn 3.0 (Supplementary Data 4). As shown in Fig. 5, comparisons of shared essential genes support the shared cellular systems between the archaeal and eukaryotic domains. More total S. islandicus essential gene orthologs are shared with archaea and eukaryotes (gray in Fig. 5), and more of these shared orthologs are essential (colors), than are shared between S. islandicus and the bacteria we use for comparison. The highest number of shared essential genes (187) is between S. islandicus and M. maripaludis S231, an organism from the euryarchaeal lineage of the archaeal domain (Table 1). The large size of the essential gene set shared between Sulfolobus and Methanococcus, in spite of their widely different habitats and life styles, reinforces the fundamental nature of Archaea as a distinct cell type50.

**Phyletic distributions of essential genes.** To investigate the broader phyletic distributions of S. islandicus essential genes, we
used assignments from the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) database\textsuperscript{31} (see Methods) to map the presence and absence of putative essential gene orthologs from 168 complete genomes representing major clades in all 3 domains based on a previously published analysis\textsuperscript{50} (Supplementary Data 5 and Supplementary Data 6). Figure 6 graphically shows the \emph{S. islandicus} essential genes shared in other genomes in a set of hierarchical clusters based on Euclidean distance. From this Fig. 4, primary transitions emerge in the evolution of the contemporary \emph{S. islandicus} essential genome. The number of genes in phyletic groups (Table 2) is significantly different from random sampling among phyletic categories (Supplementary Table 4). Similarly ranked distributions are seen in two additional datasets: (1) all genomes in the eggNOG database subsampled to have equal representation in each domain and (2) all genomes in the eggNOG database for which assignments are available. These data are supported by parsimony analysis with bootstrap support for the grouping of each of the three major domains (Supplementary Figs. 7 and 8). Together these data support four primary stages in the evolution of the contemporary \emph{S. islandicus} cells and allow us to assign specific essential genes to these potential transitions in the evolution of the cell.

The highest number of essential genes are shared broadly across the tree of life (Universal in Table 2), supporting the early evolution of the majority of essential gene functions in the contemporary archaeal cell. Most of these have putative functional assignments in information processing, particularly translation and transcription (Supplementary Data 7). Many previous studies have reported the evolutionary conservation of information processing components going back to the Last Universal Common Ancestor (LUCA) using computational methods\textsuperscript{7,9,11,32}. We find that in all studies the majority of conserved orthologous gene sets that we could interrogate in this system are essential (Supplementary Table 5 and Supplementary Data 8). Of the 200 metabolic COGs identified in the \emph{S. islandicus} genome from a recent estimate of the LUCA gene set\textsuperscript{10}, only 19 were found to be essential (Supplementary Data 8). This is expected, due to our use of rich medium. The first phase of the cell contains the universal set of genes with conserved cellular components that are likely to have evolved early in evolutionary history remain essential components of the contemporary \emph{S. islandicus} genome today.

The next largest category of essential genes is found between \emph{Sulfolobus} and other organisms in the Eukarya/Archaea (EA) domains (Table 2). These genes are largely involved in core information processing functions and support the shared evolutionary ancestry of the Archaea and Eukarya after their divergence from Bacteria. Only one gene in this category is poorly characterized: \emph{M164_0237}, a homolog to eukaryotic \emph{zpr1}. \emph{zpr1} is a gene essential for transcription and cell cycle progression in fungal and mammalian cells\textsuperscript{53-56} and has recently been reported as a regulator of circadian rhythm in plants\textsuperscript{57}. Though it has been noted that this gene is exclusively shared in EA\textsuperscript{56}, it remains uncharacterized in the Archaea outside of our results recognizing its essentiality in \emph{Sulfolobus} (Supplementary Table 6).

Fifty-five essential genes belong to NOGs that are shared by organisms in the archaeal domain (Table 2). Functional assignments of the archaeal-specific genes represent a diversity of functions split between core functions (translation, transcription, and replication) and peripheral functions, such as transport, defense (including all the above-mentioned predicted antitoxin genes), and metabolism. Archaea-specific DNA replication/recombination/repair genes are \emph{mtrA} and \emph{gins15}, while genes in arCOG category “transcription (K)” are largely transcription factors and do not represent core RNA polymerase functionality like the EA genes mentioned above. Fourteen of the archaeal-specific genes are poorly characterized (Table 2), 9 of which are also essential in \emph{M. maripaludis} S2 (Supplementary Data 4).

In an evolutionary context, this set of poorly characterized, but essential, archaea-specific genes are key targets for future molecular characterization since they likely highlight the unique biology of archael cells. We also show that the majority of \emph{S. islandicus} genes are conserved in evolutionary history through the archaeal domain.

The final set of essential genes are specific or largely specific to the Sulfolobales, most of which have uncharacterized functions.

Fig. 4 Thin-section TEM analysis of the wild-type and S-layer gene knockout strains. a–d Representative TEM micrographs of thin-sectioned cells of the wild-type, \(\Delta slaA\), \(\Delta slaB\), and \(\Delta slaAB\) mutant strains, respectively. Images e–h are closeups of images a–d, respectively. Magenta arrows indicate the breaking points of S-layer. CM cytoplasmic membrane, SL surface layer, QPS quasi-periplasmic space, SlaA surface layer protein A. Scale bars, 500 nm (a–d) and 200 nm (e–h).
of essential gene content in a model crenarchaeon. Our
Methods evolutionary steps in the Archaea. characterization are likely to provide further understanding for
provides a list of genes whose future molecular and systems understanding the functional interactions among essential
stages of evolution of the contemporary NOGs to the need for further study even in this model archaeon.

**Fig. 5** Shared essential genes across the three domains of life. Heatmap shows the presence of essential (colored) or non-essential (gray) shared NOGs compared with the *S. islandicus* M16.4 essential genome. Single-letter codes for functional categories are as follows: J, translation, ribosomal structure, and biogenesis; K, transcription; L, DNA replication, recombination, and repair; C, energy production and conversion; E, amino acid transport and metabolism; H, coenzyme transport and metabolism; F, nucleotide transport and metabolism; G, carbohydrate transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolite biosynthesis, transport, and catabolism; O, posttranslational modification, protein turnover, chaperone functions; M, cell wall/membrane/envelope biogenesis; U, intracellular trafficking, secretion, and vesicular transport; D, cell cycle control and mitosis; N, cell motility; T, signal transduction; S, function unknown. CPS, cellular processes and signaling

| Table 1 Number of *S. islandicus* essential genes shared with seven model organisms |
| Phyletic category<sup>a</sup> | Universal | EA | Archaea | TACK | Sulfolobales | Other |
|---|---|---|---|---|---|---|
| **Archaea** | **Methanococcus maripaludis** | 128 | 77 | 42 | 2 | 24 | 42 |
| | Shared | Essential | 93 | 55 | 20 | 2 | 0 | 17 |
| **Eukarya** | **Saccharomyces cerevisiae** | 134 | 77 | 2 | 0 | 1 | 27 |
| | Shared | Essential | 68 | 41 | 1 | 0 | 0 | 4 |
| **Schizosaccharomyces pombe** | Shared | Essential | 134 | 78 | 2 | 0 | 1 | 26 |
| **Bacteria** | **Bacteroides fragilis** | 82 | 43 | 1 | 0 | 1 | 10 |
| | Shared | Essential | 12 | 10 | 0 | 0 | 2 | 40 |
| **Bacillus subtilis** | Shared | Essential | 70 | 1 | 0 | 0 | 0 | 12 |
| **Escherichia coli** | Shared | Essential | 131 | 7 | 5 | 1 | 4 | 49 |
| | Shared | Essential | 72 | 0 | 0 | 0 | 0 | 11 |
| **JCVI Syn 3.0** | Shared | Essential | 136 | 7 | 6 | 0 | 0 | 64 |
| | Shared | Essential | 73 | 0 | 1 | 0 | 0 | 12 |

<sup>a</sup>Genees are put into a category if they are present in ≥50% of the organisms in each group, i.e., universal is in ≥50% of all the Bacteria, Archaea, and Eukarya groups. “Other” refers to genes that do not meet these criteria

(Table 2). The essentiality of these genes and whether they fit into central cellular functions as non-orthologous gene replacements or peripheral ones are important subjects of future work. This set of genes, unique to this lineage, may represent environmental adaptations. The fact that they are poorly characterized attests to the need for further study even in this model archaeon.

The key next steps toward comparative cell biology will be understanding the functional interactions among essential genes so that new gene inventions, non-orthologous gene transfers, and/or loss of specific functions can be identified. From the unique perspective of the TACK archaea, this work provides a list of genes whose future molecular and systems characterization are likely to provide further understanding for evolutionary steps in the Archaea.

In conclusion, this is the first comprehensive genome-wide study of essential gene content in a model crenarchaeon. Our profile of *S. islandicus* essential genes uncovers several surprising findings, most notably the non-essentiality of the Sulfolobus S-layer. Comparative phylogenetic patterns provide a perspective on the stages of evolution of the contemporary *S. islandicus*, its shared ancestry with the eukaryotes, and the key components that define its uniqueness as an archaeal cell.

**Methods**

**Strains and culture conditions.** The complete list of strains and plasmids used in this study is shown in Supplementary Table 7. All *S. islandicus* strains were routinely grown aerobically at 76–78 °C and pH 3.3 without shaking in basal salt medium containing 0.2% [wt/vol] dextrin (Sigma-Aldrich, USA) and 0.1% [wt/vol] tryptone (BD Biosciences, USA) (the medium is hereafter named as DY). When required, agmatine, uracil, and 5-FOA were added to a final concentration of 50, 20, and 50 μg/ml respectively. For solid plates, 2 × DY medium was supplemented with 25 mM MgSO<sub>4</sub> and 7 mM CaCl<sub>2</sub>·2H<sub>2</sub>O and mixed with 1.4% geltrol (Sigma-Aldrich, USA) with a ratio of 1:1 [vol/vol]. Plates were put into sealed bags and generally incubated for 10–14 days at 76–78 °C. Cell culture growth was monitored by optical density measurements at 600 nm using a portable cell density meter (C08000, WPA, Cambridge, UK).

**Construction of *S. islandicus* transposon mutant library.** The 755-bp argD gene cassette (SoargD) was PCR-amplified from the genomic DNA of *S. solfataricus* P2 using the primer set Soarg-D-F1/R1, introducing the Sall and Xmal sites, respectively. The resultant PCR products were digested with Sall/Xmal and then cloned into the EZ-Tn5<sup>TM</sup> pMOD<sup>TM</sup> 2 < MCS > Transposon Construction Vector (Episure, USA) in the corresponding sites, generating pT-SoargD. The Tn5 < SoargD > transposon DNA was prepared by PCR amplification from linearized pT-SoargD with 5′phosphorylated primers PCRFP/PCRPR. The PCR products, consisting of a nutritional marker flanked by a 19-bp inverted repeat (Mosaic Ends, ME), were purified and highly concentrated using the DNA Clean & Concentrator<sup>TM</sup>-5 Kit (Zymo Research, USA). Preparation of transposomes was made in a 10 μl reaction system as follows: 2.2 μg of transposon, 1 μl of EZ-Tn5 transposase (Episure, USA), and 2.5 μl of 100% glycerol. The reaction was incubated at room temperature for 30 min and then switched to 4 °C for another 72 h. In all, 1–2 μl of transposomes were transformed into *S. islandicus* RJW008 (AargD) via electroporation as described in ref. 37. Cell transformation assays were repeated dozens of times in order to collect a sufficient number of transfectants to achieve saturation mutagenesis. The theoretical number of transposon insertion colonies was calculated using a derivative of Poisson’s law: N = ln(1 − Pf)(ln(1 − B), with average gene size (900.64 bp)/genome size (2,586,647 bp). To make sure the transposon insertions cover approximately 99.99% (P = 0.0099) of the genome, around 26,448 colonies per library are required. The transformed cells were plated...
on DY plates either by glass beads or over-lay. After 10 days of incubation, the ArgD revertants were harvested from plates either by manually picking or with sterile spreaders and then pooled into three independent transposon mutant libraries (CYZ-TL1, CYZ-TL2, and CYZ-TL3), with approximately 100,000 colonies/plate. Sterile spreaders and then pooled into three independent transposon mutant libraries. beads were converted into demultiplexed compressed fastq files using the bcl2fastq v1.8.4 Conversion Software (Illumina, CA) at the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign.

DNA library preparation and high-throughput DNA sequencing. Genomic DNA from each mutant pool was extracted as described in ref. and then quantified with Qubit® 2.0 Fluorometer (Invitrogen, USA). DNA libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina, USA) with proper modifications. Briefly, 2 µg of input genomic DNA in total was simultaneously fragmented and tagged with sequencing adapters in a single enzymatic reaction tube. Afterwards, a primer mixture of Tn-seq-F (Supplementary Data 9) and N705 primer set into the XmalI site of a shutter vector. The pSed-SsoargD marker cassette, amplified from S. solfataricus P2 genomic DNA with a primer set SsoargD-F2/R2, into the XmalI site of a Sulfolobus-E.coli shuttle vector pSed. Thus, the estimated frequency of transposition is $10^{-2}$ per cell.

Table 2 Number of S. islandicus essential genes shared with 168 full genome sequences spanning the tree of life

| Phyletic category | Universal | EA | Archea | TACK | Sulfolobales | Other |
|-------------------|-----------|----|--------|------|-------------|-------|
| Shared genes      | 141       | 80 | 55     | 18   | 73          | 74    |
| Poorly characterized | 5       | 1  | 14     | 7    | 46          | 3     |

*Genes are put into a category if they are present in ≥50% of the organisms in each group, i.e., universal is in ≥50% of each of the Bacteria, Archea, and Eukarya groups. “Other” refers to genes that do not meet these criteria.

**NOG categories “Function unknown” or “General functional prediction only”. Full list shown in Supplementary Table 6.**
as "random." The program uses log_{10} FC as its measure of essentiality, which is proportional to log_{10} (reads observed/reads expected) for each gene and sets a cutoff automatically as the local minimum between essential and non-essential distributions in a density plot of the scores. The program suggested a putative maximum log_{10} FC of −5.1 for essential genes.

For the Trn-Seq Explorer software, insertion sites of all three libraries were combined and insertion sites with <4 reads were excluded for analysis due to their vast over-representation in the insertion sites and the uncertainty of their source (Supplementary Table 1). The program uses a sliding window approach and returns an Ei based on the number, location, and spatial concentration of insertion sites within each individual gene. It also allows for the adjustment of the stated start and end of the gene. As it is default, adjustments in the first 5% and last 20% of genes were excluded to compensate for misannotated start codons and proteins for which C-terminal deletions are tolerated, respectively. The program suggested an Ei maximum of 3 (Fig. 1b).

**Construction of S. islandicus mutants.** Except where otherwise stated, disruption of the chromosomal genes was achieved by replacing their coding regions (57–100% of the length of the gene was deleted) with the argD expression cassette (StoargD) derived from Sulfolobus tokodaii via a microhomology-mediated gene inactivation approach we recently developed69. Briefly, a functional argD gene was PCR-amplified from a linearized Sulfolobus-E.coli shuttle vector pSesD-StoargD with 35–40 bp homology of the targeted gene introduced, yielding the gene disruption cassettes. The resultant PCR products were purified and electroporated into the argD auxotrophic strain S. islandicus RJW008, selecting Arg^+ transformatants for the appropriate antibiotic resistance. S. islandicus strains slaA, slaB, and slaAB were deleted from the chromosome of the genetic host S. islandicus RJW004 via an improved MID strategy70,71 with knockout plasmids pMID-slaA, pMID-slaB, and pMID-slaAB, respectively. The resulting ΔslaA and ΔslaB mutants harbored an in-frame deletion of nucleotides +52 to +3687 relative to the start codon of slaA (3609 bp in length), and +13 to +1185 relative to the start codon of slaB (1194 bp in length), respectively. The ΔslaAB mutant was constructed similarly leaving 51 bp of the slaB (nt 1–51 relative to the start codon of slaB), 6 bp of restriction enzyme (MulI) site, and 9 bp of slaB (nt 1186–1194 relative to the start codon of slaB) in the chromosome of S. islandicus RJW004.

Verification of each gene replacement or deletion mutants was determined through PCR diagnosis with both flanking primers (bind outside of the targeted region) and internal primers (bind inside of targeted region), which examined the genotype and purity of mutants, respectively. The primers used to generate and confirm gene disruptions or deletions were described in Supplementary Data 9, and the expected sizes of amplicons generated from the genetic host (wt) and mutant strains were provided in Supplementary Table 8.

**Transmission electron microscopy.** Proteinaceous S-layer was extracted from S. islandicus cell cultures as described in ref. 38. To prepare the samples that were used for TEM, glow-discharged, carbon-stabilized Formvar-coated 200-mesh copper grids (Carbon Type B, cat. no. 01811, Ted Pella, Inc., USA) were placed on 8–20 μl droplets of each sample for 3 min, rinsed with degassed water, and negatively stained with 2% uranyl acetate for 15–60 s. Thin-sectioned S. islandicus cells were prepared essentially as described in ref. 67, with minor modifications as follows: after microwave fixation with the primary fixative, cells were washed in Sorenson’s Phosphate buffer with no further additives. All samples were observed using a Philips CM200 transmission electron microscope at 120 kV. Images were taken at various magnifications using a TVIPS (Tietz Video and Image Processing Systems GmbH; Germany) 2 k × 2 k Pelletier-cooled CCD camera. Scale bars were added with the ImageJ software.

**Homology search.** Homologs for the 441 essential genes found in Supplementary Data 2 were found across the 168 genomes listed in Supplementary Data 6 via the European Molecular Biology Laboratory eggNOG database72. Genomes were downloaded from the National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov). The genome to be searched was based on the sequence available on eggNOG. With the following additions: M. maripaludis S288C; additional Sulfolobales genomes were added for intra-order comparison of essential gene content (listed in Supplementary Data 6). While not included in the phylogenetic distribution analysis, the sequences for Lskarhacaea sp. GC14_7513 (1977).

**Data availability.** The raw Trn-Seq data of three independent transposon insertion libraries CYZ-TL1, CYZ-TL2, and CYZ-TL3 have been deposited at NCBI under BioSample accession numbers SAMN08628694, SAMN08628695, and SAMN08628696, respectively; Bioproject accession number PRJNA436600; and Sequence Read Archive (SRA) accession number SRP133799. Analyzed data showing the insertion locations across three independent transposon libraries can be found in Supplementary Data 1. All other data that support the findings of this work are available from the corresponding author upon request.

Received: 31 August 2018 Accepted: 24 October 2018 Published online: 21 November 2018

**References**

1. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl Acad. Sci. 74, 5088–5090 (1977).

2. Woese, C. R. & Fox, G. E. The concept of cellular evolution. J. Mol. Evol. 10, 1–6 (1977).

NATURE COMMUNICATIONS | DOI: 10.1038/s41467-018-07379-4 | www.nature.com/naturecommunications
3. Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S. & Miyata, T. Evolutionary relationship of archaeabacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. 86, 9355–9359 (1989).

4. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. 87, 4576–4579 (1990).

5. Guy, L. & Eitner, T. J. G. The archael protein ACF superfamily and the origin of eukaryotes. Trends Microbiol. 19, 580–587 (2011).

6. Raykumar, K., Brochier-Armanet, C. & Gribaldo, S. The two-domain tree of life is linked to a new root for the Archaea. Proc. Natl. Acad. Sci. 112, 6670–6675 (2015).

7. Puigbò, P., Wolf, Y. I. & Koonin, E. V. Search for a universal ancestor. Proc. Natl Acad. Sci. 89, 215–227 (1989).

8. Voigt, A. et al. Accession, Sulfolobus and Metallosphaera surface layers: structure, composition and gene expression. Mol. Microbiol. 73, 58–72 (2009).

9. Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B. & Kyaw, C. M. Archaeal S-layers: overview and current state of the art. Front. Microbiol. 8, 2597 (2017).

10. Zhang, C., Cooper, T. E., Krause, D. J. & Whitaker, R. J. Augmenting the genetic toolbox for Sulfolobus islandicus with a stringent positive selectable marker for agmatine prototrophy. Appl. Environ. Microbiol. 79, 5539–5549 (2013).

11. Feyoom, E. et al. The S-layer glycoprotein of the Crenarchaeote Sulfolobus acidocaldarius is glycosylated at multiple sites with chitobiose-linked N-glycans. Arch. 2010, 75101 (2010).

12. Reitz, T. et al. Spectroscopic study on uranyl carboxylic complexes formed at the surface layer of Sulfolobus acidocaldarius. Dalton Trans. 44, 2684–2692 (2015).

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

14. Veeranagouda, Y., Husain, F., Tenorio, E. L. & Wexler, H. M. Identification of genes required for the survival of B. fragilis using massive parallel sequencing of a saturated transposon mutant library. BMC Genomics 15, 429 (2014).

15. Kobayashi, K. et al. Essential Bacillus subtilis genes. Proc. Natl Acad. Sci. 100, 4678–4683 (2003).

16. Commichau, M., Pietack, N. & Stülke, J. Essential genes in Bacillus subtilis: a re-evaluation after ten years. Mol. Biosyst. 9, 1068–1075 (2013).

17. Woese, C. R. On the evolution of cells. Proc. Natl Acad. Sci. 99, 8742–8747 (2002).

18. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44, D286–D293 (2016).

19. Ciccarelli, F. D. et al. Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283 (2006).

20. Galcheva-Gargova, Z., Konstantinov, K. N., Wu, I.-H. & Klier, F. G. Binding of zinc finger protein ZPR1 to the epidermal growth factor receptor. Science 272, 1799 (1996).

21. Gangwani, L., Mikrut, M., Galcheva-Gargova, Z. & Davis, R. J. Interaction of ZPR1 with translation elongation factor-1α in proliferating cells. J. Cell Biol. 143, 1471–1484 (1998).

22. Gangwani, L. Deficiency of the zinc finger protein ZPR1 causes defects in transcription and cell cycle progression. J. Biol. Chem. 281, 40330–40340 (2006).

23. Schuur, A. K., Gangwani, L., Davis, R. J. & Lambright, D. G. Structural insights into the interaction of the evolutionarily conserved ZPR1 domain with eukaryotic EF1α, receptors, and SMN complexes. Proc. Natl Acad. Sci. 104, 13930–13935 (2007).

24. Kielbowicz-Matuk, A., Czarnacka, J., Banachowicz, E., Rey, P. & Rorat, T. Solanum tuberosum ZPR1 encodes a light-regulated nuclear DNA-binding protein adjusting the circadian expression of SIBBX4 to light cycle. Plant Cell Environ. 40, 424–440 (2017).

25. Deng, L., Zhu, H., Chen, Z., Liang, Y. X. & She, Q. Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon Sulfolobus acidocaldarius. Extremophiles 13, 735 (2009).

26. Giaever, G. et al. A synthetic arabinose-inducible promoter confers high levels of recombinant protein expression in hyperthermophilic archaean Sulfolobus islandicus. Appl. Environ. Microbiol. 78, 5630–5637 (2012).
Acknowledgements

We thank Chris L. Wright and Alvaro G. Hernandez from W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign (UIUC) for advice with primer design, DNA library construction, and sequencing. We also thank Whitney E. England, Angelo Blancaf, and Ted Kim for assistance with primer design, DNA library construction, and sequencing. We thank Scott J. Robinson from Beckman Institute for technical assistance with TEM imaging and sample preparation. We thank Lou A. Miller from Frederick Seitz Materials Research Laboratory Central Research Facilities, UIUC for preparing the thin-sectioned *S. islandicus* cells. Funding for this work was mainly provided by the National Aeronautics and Space Administration (NASA) through the NASA Astrobiology Institute under cooperative agreement no. NNA13AA91A, issued through the Science Mission Directorate. This work was also partially supported by Division of Environmental Biology (DEB: 1355171 to R.J.W.), US National Science Foundation, the Department of Microbiology Alice Helm Graduate Research Excellence Fellowship, UIUC (to A.P.R.P.), the Carl R. Woese Institute for Genomic Biology Undergraduate Research Scholar program, and the Office of Undergraduate Research, UIUC (to R.I.W.).

Author contributions

C.Z., A.P.R.P. and R.J.W. conceived and designed the research; C.Z. and R.I.W. carried out experimental work; A.P.R.P., C.Z., G.J.O., R.L.W. and R.J.W. analyzed the data; R.I.W. and G.J.O. contributed new reagents/analytic tools; and C.Z., A.P.R.P., R.L.W. and R.J.W. wrote the paper. All authors edited the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07379-4.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018