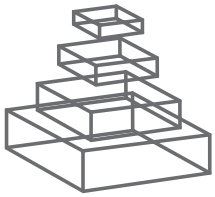
frontiers
RESEARCH TOPICS

RECENT ADVANCES IN GENOMIC AND GENETIC STUDIES IN THE ARCHAEA

Topic Editors
Frank T. Robb, Todd M. Lowe and
Zvi Kelman

frontiers in
MICROBIOLOGY



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-119-2
DOI 10.3389/978-2-88919-119-2

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

RECENT ADVANCES IN GENOMIC AND GENETIC STUDIES IN THE ARCHAEA

Topic Editors:
Frank T. Robb, Instute for Marine and Environmental Technology, USA
Todd M. Lowe, University of California, Santa Cruz, USA
Zvi Kelman, University of Maryland, USA

The accumulation of archaeal genomes has lagged significantly behind the Bacteria; however, in the last several years the coverage of the major phyla of Archaea has been significantly improved. There are now multiple genomes in several important genera such as Pyrobaculum, Sulfolobus, Thermococcus/Pyrococcus, Halobacterium, Methanosarcina, Methanopyrus and Methanocaldococcus. Comparative genomic studies are now under way, and in many cases there are several consortialmultilaboratory groups, such as the SulfoSys community, which have started to break into new systems biology initiatives. At the same time, access to streamlined genetic approaches in the genera Sulfolobus, Thermococcus, Methanosarcina, and Halobacterium/Haloferax has improved significantly and is leveraging the genomic information in the Archaea. The result has been that genome-driven studies of metabolism, DNA replication and repair, transcription and translation, and posttranslational processing have become more detailed and that basic research findings are burgeoning. The areas of global gene regulation, the roles of small RNAs and mechanisms of transcription and DNA replication will be focus areas in the guidelines of this Research Topic. Recently, insights into the unique characteristics of archaeal transcription and the ability to study the effects of mutation *in vivo* following knock-in gene replacement have resulted in incisive findings.

Table of Contents

frontiers in
MICROBIOLOGY

EDITORIAL
published: 20 December 2012
doi: 10.3389/fmicb.2012.00430

The modern "3G" age of archaeal molecular biology

Frank T. Robb[1,2], Todd M. Lowe[3] and Zvi Kelman[4,5]*

[1] *Institute of Marine and Environmental Technology, Baltimore, MD, USA*
[2] *University of Maryland School of Medicine, Baltimore, MD, USA*
[3] *Department of Biomolecular Engineering, University of California, Santa Cruz, CA, USA*
[4] *National Institute of Standards and Technology, Gaithersburg, MD, USA*
[5] *Institute for Bioscience and Biotechnology Research, Rockville, MD, USA*
**Correspondence: zkelman@umd.edu*

Edited by:
John R. Battista, Louisiana State University and A & M College, USA

Reviewed by:
John R. Battista, Louisiana State University and A & M College, USA

The publication of the genome of *Methanocaldococcus jannaschii* in 1996 (Bult et al., 1996), just the second-ever complete microbial sequencing feat, marked an exciting beginning for archaeal genomics. In spite of the auspicious start, progress in archaeal genomics has been slow relative to bacteria and viruses, in part due to sequencing funding priorities favoring microbes with clinical relevance. In the last 5 years, the precipitous drop in high throughput sequencing costs and increasingly automated genome annotation has caused the inventory of archaeal genomes to grow exponentially. There is now genome representation in all major phyla of the Archaea, with a number of popular genera (including Halobacterium, Methanocaldococcus, Methanosarcina, Pyrobaculum, Pyrococcus, Sulfolobus, and Thermococcus) represented by multiple species. With these sequence data, comparative genomic studies have played an important role in integrating a broader evolutionary perspective into archaeal research. Experimental genetics now commonly utilize multiple archaeal species in a modern systems approach, coinciding with an increasing upswing in international collaboration. Multilaboratory consortia, such as the SulfoSys community (Albers et al., 2009), have started to break into new systems biology initiatives. At the same time, streamlined genetic approaches in the genera Halobacterium/Haloferax, Methanosarcina, Sulfolobus, and Thermococcus, are leveraging the growing wealth of genomic information in the Archaea. The result has been genome-driven studies of metabolism, DNA replication and repair, transcription and translation, and post-translational processing.

This collection of 11 papers attempts to put this quiet maturation in archaeal molecular biology in context with a mix of overview articles (Atomi et al., 2012; Hileman and Santangelo, 2012; Kohler and Metcalf, 2012; Wagner et al., 2012) and original data (Bernick et al., 2012a,b; Iverson and Stedman, 2012; Mao and Grogan, 2012; Schut et al., 2012; Stroud et al., 2012; Szabo and Pohlschroder, 2012) highlighting the expansion of important areas enabled by the 3G's: genomics, genetics, and global collaboration. Research papers highlight the roles of small RNAs, CRISPR action, mechanisms of membrane secretion, hydrogen production, and DNA replication, while excellent technical overviews cover recently developed genetic methods within four major research communities utilizing Haloferax, Sulfolobus, Pyrococcus, or Thermococcus.

ACKNOWLEDGMENTS

The support of Air Force Office of Scientific Research Grant AFOSR FA9550-10-1-0272 to Frank T. Robb is gratefully acknowledged.

REFERENCES

Albers, S. V., Birkeland, N. K., Driessen, A. J., Gertig, S., Haferkamp, P., Klenk, H. P., et al. (2009). SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon Sulfolobus solfataricus under temperature variation. *Biochem. Soc. Trans.* 37, 58–64.

Atomi, H., Imanaka, T., and Fukui, T. (2012). Overview of the genetic tools in the Archaea. *Front. Microbio*. 3:337. doi: 10.3389/fmicb.2012.00337

Bernick, D. L., Cox, C. L., Dennis, P. P., and Lowe, T. M. (2012a). Comparative genomic and transcriptional analyses of CRISPR systems across the genus *Pyrobaculum*. *Front. Microbio*. 3:251. doi: 10.3389/fmicb.2012.00251

Bernick, D. L., Dennis, P. P., Lui, L. M., and Lowe, T. M. (2012b). Diversity of antisense and other non-coding RNAs in archaea revealed by comparative small RNA sequencing in four *Pyrobaculum* species. *Front. Microbio.* 3:231. doi: 10.3389/fmicb.2012.00231

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., et al. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073.

Hileman, T. H., and Santangelo, T. J. (2012). Genetics techniques for *Thermococcus kodakarensis*. *Front. Microbio.* 3:195. doi: 10.3389/fmicb.2012.00195

Iverson, E., and Stedman, K. (2012). A genetic study of SSV1, the prototypical fusellovirus. *Front. Microbio*. 3:200. doi: 10.3389/fmicb.2012.00200

Kohler, P. R. A., and Metcalf, W. W. (2012). Genetic manipulation of *Methanosarcina* spp. *Front. Microbio*. 3:259. doi: 10.3389/fmicb.2012.00259

Mao, D., and Grogan, D. W. (2012). Heteroduplex formation, mismatch resolution, and genetic sectoring during homologous recombination in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. *Front. Microbio.* 3:192. doi: 10.3389/fmicb.2012.00192

Schut, G. J., Nixon, W. J., Lipscomb, G. L., Scott, R. A., and Adams, M. W. W.

(2012). Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Front. Microbio.* 3:163. doi: 10.3389/fmicb.2012.00163

Stroud, A., Liddell, S., and Allers, T. (2012). Genetic and biochemical identification of a novel single-stranded DNA-binding complex in *Haloferax volcanii*. *Front. Microbio.* 3:224. doi: 10.3389/fmicb.2012.00224

Szabo, Z., and Pohlschroder, M. (2012) Diversity and subcellular distribution of archaeal secreted proteins. *Front. Microbio.* 3:207. doi: 10.3389/fmicb.2012.00207

Wagner, M., van Wolferen, M., Wagner, A., Lassak, K., Meyer, B. H., Reimann, J., et al. (2012). Versatile genetic tool box for the crenarchaeote *Sulfolobus acidocaldarius*. *Front. Microbio.* 3:214. doi: 10.3389/fmicb.2012.00214

Received: 12 November 2012; accepted: 06 December 2012; published online: 20 December 2012.

Citation: Robb FT, Lowe TM and Kelman Z (2012) The modern "3G" age of archaeal molecular biology. Front. Microbio. ***3****:430. doi: 10.3389/fmicb.2012.00430*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

REVIEW ARTICLE
published: 02 October 2012
doi: 10.3389/fmicb.2012.00337

Overview of the genetic tools in the Archaea

Haruyuki Atomi[1,2]*, Tadayuki Imanaka[2,3] and Toshiaki Fukui[4]

[1] *Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto, Japan*
[2] *JST, CREST, Sanbancho, Chiyoda-ku, Tokyo, Japan*
[3] *Department of Biotechnology, College of Life Sciences, Ritsumeikan University, Noji-Higashi, Kusatsu, Shiga, Japan*
[4] *Department of Bioengineering, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama, Japan*

Edited by:
Frank T. Robb, Institute of Marine and Environmental Technology, USA

Reviewed by:
Thijs Ettema, Uppsala University, Sweden
Imke Schroeder, University of California, Los Angeles, USA

****Correspondence:***
Haruyuki Atomi, Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Katsura, Nishikyo-ku, Kyoto 615-8510, Japan.
e-mail: atomi@sbchem.kyoto-u.ac.jp

This section provides an overview of the genetic systems developed in the Archaea. Genetic manipulation is possible in many members of the halophiles, methanogens, *Sulfolobus*, and Thermococcales. We describe the selection/counterselection principles utilized in each of these groups, which consist of antibiotics and their resistance markers, and auxotrophic host strains and complementary markers. The latter strategy utilizes techniques similar to those developed in yeast. However, Archaea are resistant to many of the antibiotics routinely used for selection in the Bacteria, and a number of strategies specific to the Archaea have been developed. In addition, examples utilizing the genetic systems developed for each group will be briefly described.

Keywords: Archaea, gene disruption, shuttle vectors, genetics, halophiles, methanogens, *Sulfolobus*, Thermococcales

INTRODUCTION

Genetic manipulation, designated here as the ability to introduce, remove, or modify genes in a given organism, is a vital tool to study gene function. Deleting or overexpressing a gene may lead to phenotypic changes that provide valuable clues in determining the physiological role of the gene. Random mutagenesis and the isolation of mutant strains, followed by screening for genes that complement the mutations is a classical strategy to identify groups of genes that are involved in a particular biological function. Genetic manipulation can also be used to engineer cells to improve or introduce a desired function in a cell. The tools necessary for genetic manipulation have been developed in a wide variety of eukaryotes and bacteria, including the yeast *Saccharomyces cerevisiae*, the Gram-negative bacterium *Escherichia coli* and Gram-positive bacterium *Bacillus subtilis*, all of which have been subject to genome-wide gene disruption projects (Giaever et al., 2002; Kobayashi et al., 2003; Baba et al., 2006).

Compared to eukaryotes and bacteria, the development of genetic systems in Archaea is still at a modest stage. Many archaeal species have been found to be resistant against conventional antibiotics utilized for selection in bacterial genetic systems. In addition, many archaeal species can be regarded as extremophiles, preferring growth conditions that greatly differ to those of the mesophilic, aerobic model microbes such as *S. cerevisiae*, *E. coli*, and *B. subtilis*, which adds some difficulty to establish efficient screening methods. For example, when developing a system for hyperthermophilic archaea, the (thermo)stability of the compounds used for selection must also be taken into account, and establishing techniques necessary for growing colonies at high temperatures (and in many cases under an anaerobic environment) are necessary. These factors and others have hampered the development of archaeal genetic systems in the past, but the number of archaea with genetic systems is now increasing at a steady rate. Among the Crenarchaeotes, genetic manipulation is possible in a number of species in the genus *Sulfolobus*. In the Euryarchaeota, genetic systems have been developed in a number of halophiles, methanogens, and members of the Thermococcales. This section will give an overview of the genetic systems developed in these archaeal species, focusing on the principles applied for transformant selection (summarized in **Table 1**) and some examples of gene disruption that have led to a better understanding of gene function. An in-depth description of the individual organisms and detailed methodology, along with a historical account on the development of these systems, are available in the literature (Whitman et al., 1997; Tumbula and Whitman, 1999; Allers and Mevarech, 2005; Rother and Metcalf, 2005; Berkner and Lipps, 2008; Wagner et al., 2009; Buan et al., 2011; Leigh et al., 2011).

HALOPHILES

SYSTEMS BASED ON ANTIBIOTIC RESISTANCE

Genetic systems have mainly been developed in *Halobacterium salinarum* and *Haloferax volcanii*. Systems based on both antibiotic resistance markers and auxotrophic selectable markers have been established. In terms of systems based on antibiotic resistance, novobiocin, which inhibits DNA gyrase, and mevinolin/simvastatin, which inhibits 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, are two antibiotics that have successfully been applied in halophiles belonging to the genera *Haloferax* and *Halobacterium*. DNA gyrase is a type II topoisomerase that introduces negative supercoils into DNA and whose function is essential for DNA synthesis. HMG-CoA reductase is one of the enzymes of the mevalonate

Table 1 | A simple summary of the selection strategies employed for genetic manipulation in the Archaea.

	Marker gene	Host requirements	Medium requirements	Demonstrated in			
				H	M	S	T
SELECTION CRITERION							
Novobiocin resistance	*gyrB* mutant	–	–	O	–	–	–
Mevinolin/simvastatin resistance	*hmgR* overexpression	–	–	O	–	O	O
Puromycin resistance	*pac*	–	–	–	O	–	–
Neomycin resistance	APH3′I/II	–	–	–	O	–	–
Hygromycin B resistance	Thermostable *hph* mutant	–	–	–	–	O	–
Butanol/benzyl alcohol resistance	*adh*	–	–	–	–	O	O
Uracil prototrophy	*pyrE*, *pyrF*	$pyrE^{-}$, $pyrF^{-}$, upp^{+}	Pyrimidine-free	O	–	O	O
Leucine prototrophy	*leuB*	$leuB^{-}$	Leucine-free	O	–	–	–
Tryptophan prototrophy	*trpE*, *trpAB*	$trpE^{-}$, $trpAB^{-}$	Tryptophan-free	O	–	–	O
Histidine prototrophy	*hisA*	$hisA^{-}$	Histidine-free	–	O	–	–
Lactose prototrophy	*lacS*	$lacS^{-}$	Lactose as major carbon/energy source	–	–	O	–
Agmatine prototrophy	*pdaD*	$pdaD^{-}$	Applicable with tryptone/yeast extract	–	–	–	O
COUNTERSELECTION CRITERION							
5-Fluoroorotic acid resistance	*pyrE*, *pyrF*	$pyrE^{-}$, $pyrF^{-}$	–	O	–	O	O
6-Azauracil/8-azahypoxanthine/8-aza-2,6-diaminopurine resistance	*hpt*	Resistant w/o *hpt*	–	–	O	–	–
6-Methylpurine resistance	*hpt*	Resistant w/o *hpt*	–	–	–	–	O

H, halophiles; M, methanogens; S, Sulfolobus; T, Thermococcales. Details are described in the text and referred publications.

pathway, which utilizes three molecules of acetyl-CoA to synthesize isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) (**Figure 1**). IPP and DMAPP are precursors for isoprenoid compounds, which are particularly important for the archaea as their membrane lipids utilize isoprenoid chains. A gene that encodes a novobiocin-resistant DNA gyrase was isolated from *Haloferax* strain Aa2.2 and has been used as a selection marker in developing *Hf. volcanii–E. coli* shuttle vectors (Holmes and Dyall-Smith, 1990; Holmes et al., 1991, 1994). Furthermore in *Hf. volcanii*, shotgun cloning of DNA from spontaneous mevinolin-resistant strains led to the isolation of DNA fragments that could transform *Hf. volcanii* to mevinolin resistance, enabling the construction of shuttle vectors (Lam and Doolittle, 1989; Blaseio and Pfeifer, 1990). An examination of various *Hf. volcanii* mevinolin-resistant mutants have revealed that resistance is brought about by either gene amplification or up-promoter mutations, both resulting in enhanced and excess production of HMG-CoA reductase (Lam and Doolittle, 1992). Shuttle vectors such as pWL102 have been shown to also be applicable in transforming members of the genus *Haloarcula* (Cline and Doolittle, 1992). Additional shuttle vectors and gene disruption systems are now available in several members of the *Haloarcula* and their application has been demonstrated (Zhou et al., 2004; Ozawa et al., 2005; Tu et al., 2005).

SYSTEMS BASED ON AUXOTROPHIC SELECTABLE MARKERS

The *ura3* (or *pyrF*) gene encoding orotidine-5′-monophosphate decarboxylase, an enzyme necessary for *de novo* pyrimidine biosynthesis, has been utilized as a selection marker in a number of halophilic archaea including *Hb. salinarum* NRC-1, *Haloferax mediterranei,* and *Haloarcula hispanica* (Liu et al., 2011). A host cell with a defect in *ura3*/*pyrF* can grow when uracil is added to the medium owing to the function of uracil phosphoribosyltransferase encoded by the *upp* gene. The *ura3*/*pyrF* system is especially convenient as it also allows counterselection. The addition of 5-fluoroorotic acid (5-FOA) to the medium prohibits growth of cells with an intact *ura3*/*pyrF* gene, as 5-FOA is converted to the toxic 5-fluorouridine 5′-phosphate and 5-fluorouracil (Boeke et al., 1984, 1987). These compounds inhibit DNA/RNA synthesis, with the latter known to inhibit thymidylate synthase, an enzyme necessary for thymidine synthesis. It is thus possible to specifically select cells that have lost a *pyrF* (or *pyrE*) gene by supplementing the medium with 5-FOA and a pyrimidine precursor such as uracil (**Figure 2**). In *Hb. salinarum*, a *ura3*/*pyrF* deletion strain was constructed using a mevinolin resistance marker, and the use of the *ura3*/*pyrF* gene as a counterselection marker has been extensively examined (Peck et al., 2000). Further improvements have enabled the use of *ura3*/*pyrF* as both a selection marker (uracil prototrophy) for initial plasmid integration, and as a counterselection marker (5-FOA resistance) for plasmid

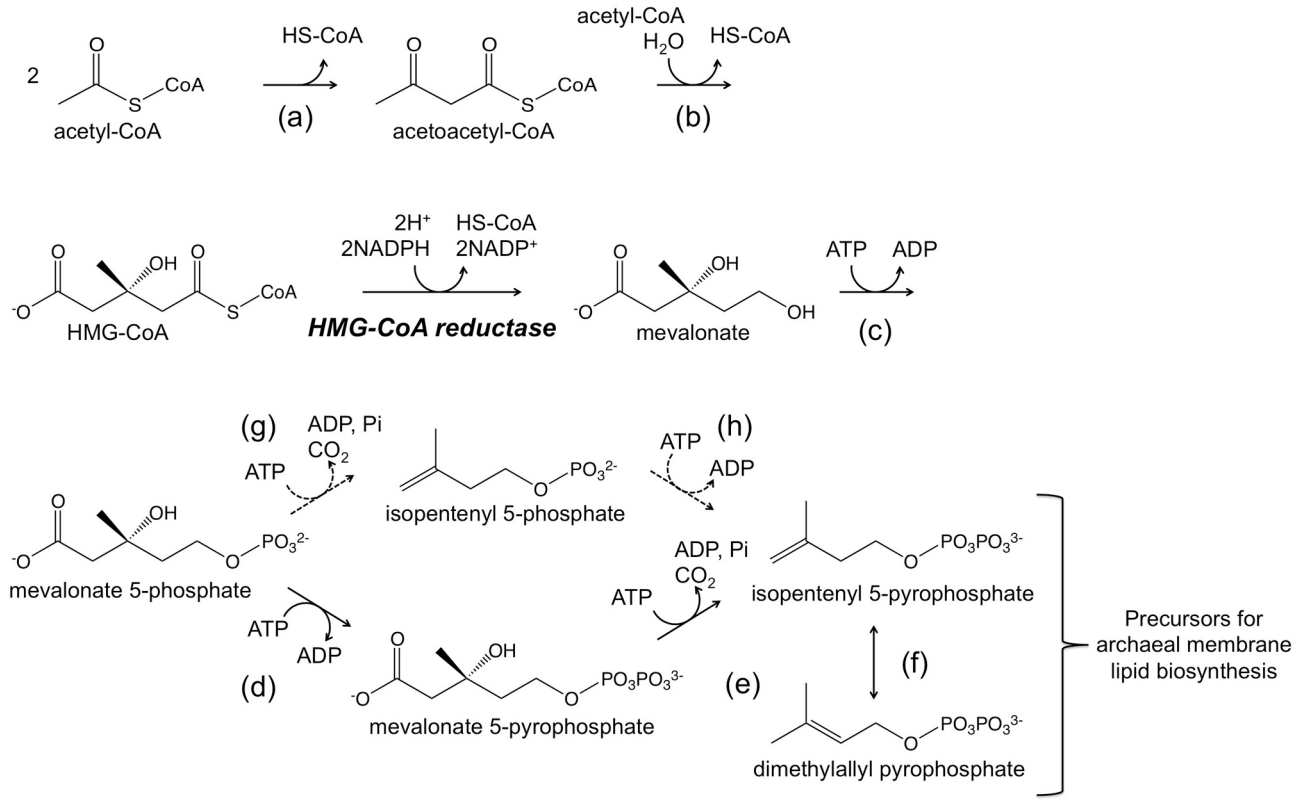


FIGURE 1 | A schematic illustration of the mevalonate pathway. The dotted arrows indicate a possible route adopted in the Archaea (Grochowski et al., 2006; Matsumi et al., 2011). Enzymes catalyzing the reactions are (a) acetoacetyl-CoA thiolase, (b) HMG-CoA synthase, HMG-CoA reductase (written), (c) mevalonate kinase, (d) mevalonate-5-phosphate kinase, (e) mevalonate-5-pyrophosphate decarboxylase, (f) isopentenyl-5-pyrophosphate isomerase, (g) mevalonate-5-phosphate decarboxylase, and (h) isopentenyl-5-phosphate kinase.

excision and gene deletion (Wang et al., 2004). The methodology has been successfully applied in disrupting and examining the arsenic resistance genes of this organism (Wang et al., 2004). The system has also been used to study the physiological roles of TATA binding proteins and transcription factor B proteins, whose genes are present in multiple copies on the genome (Coker and Dassarma, 2007). The methodology developed in *Hb. salinarum* can also be used in *Hf. mediterranei* and *Ha. hispanica*. The system was applied in deleting the phytoene synthase gene in both of these organisms (Liu et al., 2011).

Systems based on other auxotrophic selectable markers have been established in *Hf. volcanii*. The *ura5* (or *pyrE*) gene encoding orotate phosphoribosyltransferase, responsible for the reaction preceding that of the *ura3*/*pyrF* product, has been demonstrated to be applicable as both a selection and counterselection marker (Bitan-Banin et al., 2003). Although two genes (*pyrE1* and *pyrE2*) encoded proteins homologous with PyrE, *pyrE2* was the gene actually involved in pyrimidine biosynthesis. A Δ*pyrE2* strain was constructed and used as a host cell to disrupt the *cmi4* gene of *Hf. volcanii*. Systems based on selection markers involved in amino acid biosynthesis have also been developed in *Hf. volcanii* (Allers et al., 2004). The *leuB* gene, encoding 3-isopropylmalate dehydrogenase in the leucine biosynthesis pathway, and the *trpA* gene that encodes one of the two subunits of tryptophan synthase have been used for selection based on leucine and tryptophan prototrophy, respectively. A convenient system for gene expression has also been developed in *Hf. volcanii* (Allers et al., 2010). Use of the tryptophanase promoter of *Hf. volcanii* (p.*tna*) promoter, which is induced by tryptophan, allows conditional overexpression of the target gene. The genetic background of the host strain has also been modified to facilitate the purification of His-tagged proteins and relieving the need to passage DNA through an *E. coli dam* mutant.

APPLICATION OF THE GENETIC SYSTEMS IN HALOPHILES

Gene manipulation is routinely performed in the halophiles and an overwhelming amount of genetic examinations has been reported in the literature (Leigh et al., 2011; Soppa, 2011). This most likely reflects the fact that genetic systems were developed at a relatively early stage in the halophiles and the versatility of the genetic systems themselves, along with the mesophilic and aerobic lifestyles of these organisms. To mention only several of the most recent studies, in *Hb. salinarum*, a gradual inducible gene expression system has been developed (Kixmüller and Greie, 2012). It relies on the promoter of the potassium uptake system operon (P*kdp*), which responds to potassium cation concentrations in the medium. A workflow for genome-wide mapping of transcription factors from *Hb. salinarum* has also been reported (Wilbanks

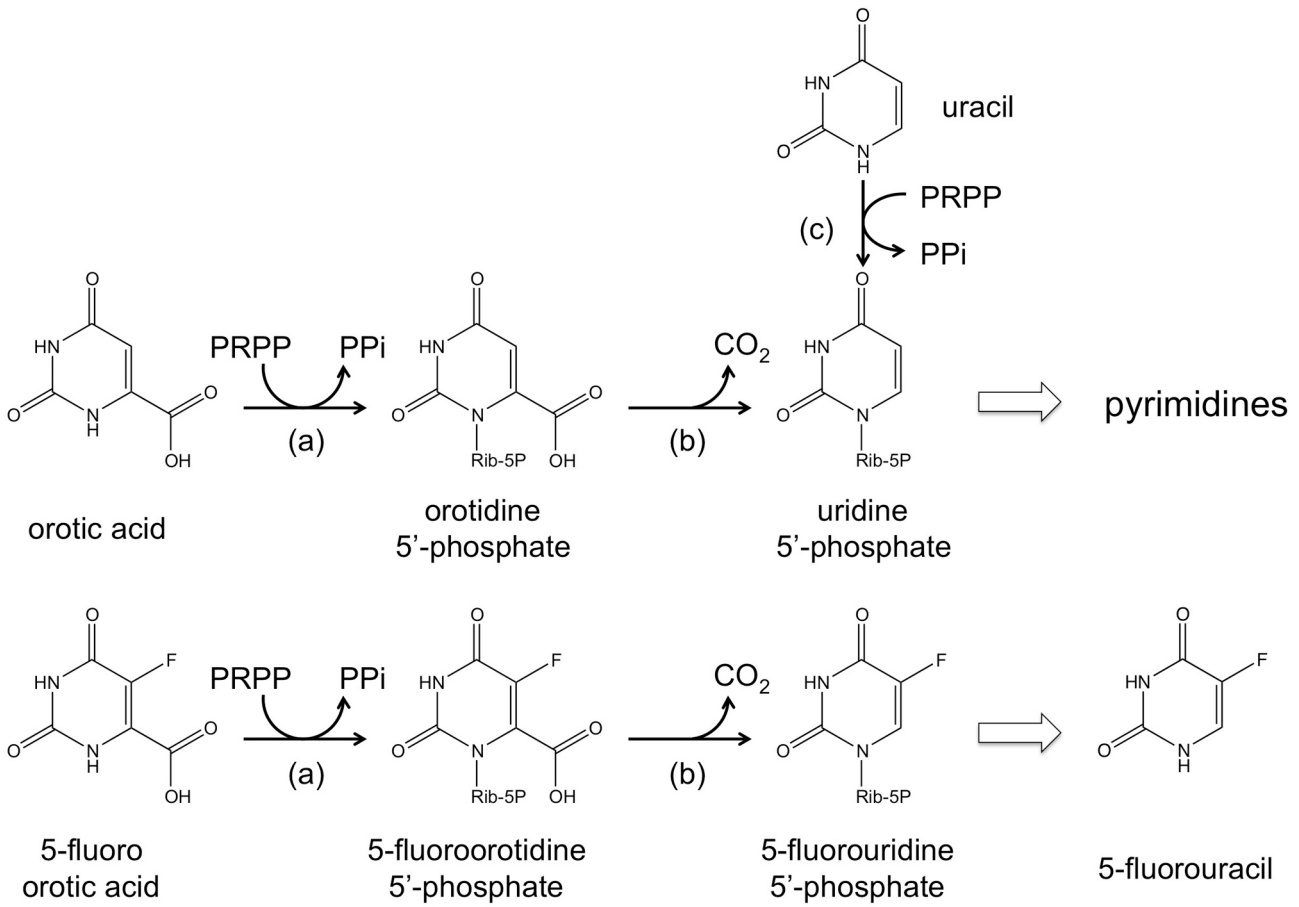


FIGURE 2 | A schematic illustration of the reactions catalyzed by (a) orotate phosphoribosyltransferase (*pyrE* gene product), (b) orotidine-5′-monophosphate decarboxylase (*pyrF* gene product), and (c) uracil phosphoribosyltransferase (*upp* gene product). The conversion from 5-fluoroorotic acid (5-FOA) to 5-fluorouridine 5′-phosphate and 5-fluorouracil is also shown.

et al., 2012). Target genes such as those encoding the general transcription factor TfbD and the specific transcription factor Bat were modified to incorporate a hemagglutinin tag at the C-termini of the proteins, and the epitopes were used for chromatin immunoprecipitation coupled with high-throughput sequencing. Another study has examined the regulation of bacteriorhodopsin, particularly the relationship between bacterioopsin and retinal biosynthesis (Dummer et al., 2011). The results suggest that bacterioopsin accumulation promotes the production of its cofactor retinal by inhibiting bacterioruberin biosynthesis. In *Hf. mediterranei*, a genetic approach was applied in examining the functions of polyhydroxyalkanoate granule-associated proteins (Cai et al., 2012). The PhaP protein in this organism was found to act as the predominant structure protein on the PHA granules. In *Hf. volcanii*, a conserved archaeal gene with sequence similarity with a tRNA 3′-processing endonuclease has been studied biochemically and genetically, suggesting that in contrary to its annotation, the gene is involved in membrane transport (Fischer et al., 2012). Another study identifies the enzyme responsible for reduction of the ω-position isoprene of dolichol phosphate in *Hf. volcanii* (Naparstek et al., 2012). Single-stranded DNA-binding proteins have also been genetically examined. Five genes that encode proteins homologous to replication protein A (RPA) from *Hf. volcanii* (RpaA1A2, RpaB1B2, RpaC) were analyzed, revealing the essentiality of RpaC and the functional relationship among RPA proteins in this archaeon (Skowyra and Macneill, 2012). A metabolic study identified the enzymes responsible for fructose metabolism. *Hf. volcanii* adopts a bacteria-like phosphoenolpyruvate-dependent phosphotransferase system that generates fructose 1-phosphate, which is further converted to trioses via fructose-1-phosphate kinase and a Class II fructose-1,6-bisphosphate aldolase (Pickl et al., 2012).

METHANOGENS

Genetic systems have been developed in a number of species in the genera *Methanococcus* and *Methanosarcina*. DNA-mediated transformation was first demonstrated in *Methanococcus voltae* (Bertani and Baresi, 1987). Most systems in the methanogens rely on antibiotic resistance for selection. Puromycin and the puromycin transacetylase (*pac*) gene from the bacterium *Streptomyces alboniger* (Gernhardt et al., 1990) and its derivatives

are often used as the antibiotic and the resistance marker gene, respectively, in methanogen genetics. Initial integration shuttle vectors using the *pac* gene were constructed and successfully used to transform *M. voltae* (Gernhardt et al., 1990; Patel et al., 1994) and *Methanococcus maripaludis* (Sandbeck and Leigh, 1991; Tumbula et al., 1994). Selection based on histidine auxotrophy/prototrophy using the *hisA* gene as a marker has also been demonstrated (Pfeifer et al., 1998). Counterselection methods have been developed (Moore and Leigh, 2005) based on the observation that *M. maripaludis* cells displaying growth were sensitive to the base analogs 6-azauracil and 8-azahypoxanthine (Bowen and Whitman, 1987; Ladapo and Whitman, 1990; Bowen et al., 1996; Kim and Whitman, 1999). Many of the principles and techniques developed in one methanogen have been shown to be applicable in other methanogen species.

In the *Methanosarcina*, efficient introduction of DNA is possible using liposome-mediated transformation (Metcalf et al., 1997). Replicating shuttle vectors were developed for *Methanosarcina acetivorans* and were also found to be applicable in a wide range of other *Methanosarcina* species including *Methanosarcina barkeri*, *Methanosarcina mazei*, and *Methanosarcina thermophila*. Gene disruption using the *pac* gene has been demonstrated in *M. acetivorans* and *M. mazei*. Markerless genetic exchange using the hypoxanthine phosphoribosyltransferase gene (*hpt*) was developed in *M. acetivorans* and also utilized in *M. barkeri*, with counterselection performed based on 8-aza-2,6-diaminopurine (8ADP) resistance (Pritchett et al., 2004; Rother and Metcalf, 2005; Welander and Metcalf, 2008; Buan et al., 2011). An *in vivo* transposon mutagenesis system has also been developed using a modified mariner-family transposable element originally derived from insect (Zhang et al., 2000).

APPLICATION OF THE GENETIC SYSTEMS IN METHANOGENS

Using these genetic systems, a number of genes in *M. mazei* Gö1, including those encoding the $GlnK_1$ protein and archaeal histone, have been disrupted. *glnK*$_1$ disruption revealed that $GlnK_1$ is not directly involved in the transcriptional regulation of nitrogen assimilation/fixation genes, but does play a role in growth under nitrogen limiting conditions (Ehlers et al., 2005). Disruption of the histone gene was not lethal, but resulted in impaired growth on methanol and trimethylamine, and increased sensitivity to UV light. A broad genome-wide defect in gene transcription was also observed (Weidenbach et al., 2008). In *M. acetivorans*, the *pylT* gene encoding the tRNA for pyrrolysine was disrupted. The disruptant did not show growth defects when grown on methanol or acetate, but could not grow on methylamines, consistent with the fact that the methyltransferases from this organism that are involved in methylamine-dependent methanogenesis possess pyrrolysine (Mahapatra et al., 2006). A genetic approach was also used to distinguish the physiological roles of two gene clusters on the *M. acetivorans* genome encoding an archaeal A_1A_0-ATPase and a bacterial F_1F_0-ATPase. A mutant disrupted of the latter gene cluster did not display growth defects, and intracellular ATP levels were identical to those in wild-type cells, indicating that the F_1F_0-ATPase is dispensible for growth in *M. acetivorans* (Saum et al., 2009). The four studies introduced here have all utilized the *pac* gene for selection of the gene disruptants.

For *M. voltae*, protoplasts can efficiently be transformed by natural or electroporation-mediated uptake of exogenous DNA (Patel et al., 1994). Liposome-mediated transformation has also been applied (Heinicke et al., 2004; Chaban et al., 2009). Gene disruption has been demonstrated on the selenium-free Vhc and Frc hydrogenase genes in order to examine the individual roles of four hydrogenase gene clusters (Berghöfer and Klein, 1995). Four genes encoding the chromatin proteins histone (*hstA*, *hstB*), histone-like protein (*hmvA*), and an Alba homolog (*AlbA*) have been individually disrupted, revealing their involvement in regulation of gene expression (Heinicke et al., 2004). A genetic approach has also been taken to study post-translational protein modification. For example, two genes designated as *aglC* and *aglK* were shown to be necessary for proper *N*-glycosylation in this organism. It was suggested that the two genes are involved in the biosynthesis or transfer of diacetylated glucuronic acid within the glycan structure (Chaban et al., 2009).

In *M. maripaludis*, integration shuttle vectors and methods for auxotroph isolation (see above), and transposon insertion mutagenesis (Blank et al., 1995) and random insertional mutagenesis (Kim and Whitman, 1999) were developed at an early stage (Whitman et al., 1997; Tumbula and Whitman, 1999; Leigh et al., 2011). In addition to puromycin, *M. maripaludis* was found to be sensitive to neomycin, and the use of aminoglycoside phosphotransferase genes APH3′I and APH3′II as selectable markers has been demonstrated (Argyle et al., 1996). Transformation methods have been optimized and are performed via a polyethylene glycol-mediated method (Tumbula et al., 1994). A shuttle vector that replicates in both *E. coli* and *M. maripaludis* was constructed based on the plasmid pURB500 from this archaeon (Tumbula et al., 1997). Using the histone promoter from *M. voltae*, vectors for overexpression of endogenous and heterologous genes have been developed (Gardner and Whitman, 1999). A genetic approach has been taken to examine a wide variety of biological functions in *M. maripaludis*. The mechanisms and regulation of nitrogen fixation has been extensively examined. Repressor binding sites of *nifH*, encoding the nitrogenase reductase component of the nitrogenase complex, have been identified using a *nifH* promoter-*lacZ* *in vivo* reporter system (Cohen-Kupiec et al., 1997). A similar sequence was found upstream of the glutamine synthetase gene (*glnA*) and shown to function in repression. The repressor protein, NrpR, was identified, and its gene disruption, along with *in vitro* binding experiments, clearly demonstrated its function as a DNA-binding transcriptional repressor that regulates genes involved in nitrogen assimilation (Lie and Leigh, 2003). Further studies have revealed how NrpR binds to specific operator sequences and how it is released from DNA by 2-oxoglutarate binding (Lie et al., 2005, 2010). Furthermore, mechanisms governing posttranslation regulation, namely ammonia switchoff, of nitrogenase have also been examined in detail (Kessler and Leigh, 1999; Kessler et al., 2001; Dodsworth and Leigh, 2006). Genetics have also been utilized to study the energy-conserving hydrogenases in *M. maripaludis*. Gene disruption of one of the two membrane-bound hydrogenase complexes, Ehb, has revealed that the complex is involved in anabolic CO_2 assimilation (Porat et al., 2006). Results of phenotypic analyses suggested that Ehb donates the electrons

necessary for aromatic amino acid biosynthesis from aryl acids via the function of indolepyruvate oxidoreductase (Major et al., 2010). In addition to studies in *M. voltae*, *N*-glycosylation has also been examined in *M. maripaludis*. A putative acetyltransferase gene was subjected to gene disruption, and the mutant cells were found to produce flagellin proteins with sizes corresponding to proteins with defects in glycosylation. In addition to flagellar filament assembly, defects in pilus anchoring were also observed, indicating that flagellum and pilus assembly are linked in their post-translational modification mechanisms (Vandyke et al., 2008). Further studies have identified multiple genes that are necessary for piliation and have also led to the identification of the protein that corresponds to the major pilin monomer, a protein whose gene resides outside of the gene cluster that had been predicted to harbor most of the genes related to pilus formation (Ng et al., 2011). A number of recent studies have examined mechanisms of selenocysteine (Sec) biosynthesis in various strains of *M. maripaludis* (Stock et al., 2010, 2011; Hohn et al., 2011). A selenophosphate synthetase homolog (*selD*) in *M. maripaludis* S2 could not be deleted unless a bacterial selenophosphate synthetase gene was present *in trans*, whereas disruption of the corresponding gene in *M. maripaludis* JJ was possible. Further genetic examination on the latter strain indicated that selenophosphate is the selenium donor in this strain (Stock et al., 2010). In another strain *M. maripaludis* Mm900, which is related to S2, *selD* disruption was possible. Whereas the ability to grow on formate was abolished, hydrogenotrophic growth was unaffected (Hohn et al., 2011). Interestingly, disruption of genes encoding phosphoseryl-tRNA^{Sec} kinase and phosphoseryl-tRNA:Sec-tRNA synthase was possible only when either *selD* was disrupted or if selenium-free hydrogenases were expressed. Detailed biochemical characterization of the gene disruption strains suggests a complex regulatory mechanism of Sec biosynthesis in *M. maripaludis*.

SULFOLOBUS

There are many natural genetic elements related to the crenarchaeal genus *Sulfolobus*, including viruses, cryptic plasmids, and transposons (Zillig et al., 1996, 1998; Prangishvili et al., 1998; Stedman et al., 2000). Several transformation systems based on these natural elements were developed in *Sulfolobus solfataricus* and *Sulfolobus acidocaldarius* at an early stage, followed by the establishment of gene manipulation systems based on useful selectable markers (Aagaard et al., 1996; Elferink et al., 1996; Berkner and Lipps, 2008; Wagner et al., 2009; Leigh et al., 2011). Transformation is now mainly carried out by electroporation. In *S. solfataricus*, selection is possible by using a strain with a deletion in *lacS*, which encodes a β-galactosidase. By using an intact *lacS* as a marker gene, transformants can be selected by their ability to grow in a minimal medium containing lactose (Worthington et al., 2003). LacS^+ colonies can be further identified by blue/white detection using X-Gal (Schelert et al., 2004). Selection based on resistance toward hygromycin B has also been reported using a gene encoding a thermostabilized hygromycin phosphotransferase from *E. coli* (Cannio et al., 1998). Another system is based on resistance toward butanol or benzyl alcohol using an alcohol dehydrogenase gene from *S. solfataricus* (Aravalli and Garrett, 1997). In *S. acidocaldarius* and *Sulfolobus islandicus*, host strains with defects in *pyrE*, *pyrF* or both are utilized, with intact *pyrE* and *pyrF* genes as selection markers (Deng et al., 2009; She et al., 2009; Wagner et al., 2009). This strategy has also been utilized in *S. solfataricus*. Based on these selection strategies, a wide range of *Sulfolobus–E. coli* shuttle vectors have been developed and are described in detail in the literature (Aravalli and Garrett, 1997; Stedman et al., 1999; Jonuscheit et al., 2003; Albers et al., 2006; Aucelli et al., 2006; Berkner et al., 2007; Berkner and Lipps, 2008).

APPLICATION OF THE GENETIC SYSTEMS IN *SULFOLOBUS*

Sulfolobus solfataricus

Gene disruption based on *lacS* selection has been utilized to examine a wide range of functions in *S. solfataricus*. Conditions for gene disruption have been carefully examined and optimized (Albers and Driessen, 2007). Genetic and biochemical examination has been performed on genes involved in mercury resistance, demonstrating that the *merR* gene product represses transcription of an operon that includes the mercuric reductase gene *merA* (Schelert et al., 2004, 2006). Another study demonstrated that the *copR* gene product is a transcriptional activator of genes encoding copper-transporting ATPase and copper-binding protein and is necessary for copper tolerance of *S. solfataricus* (Villafane et al., 2011). A Lrp-like regulator, Ss-LrpB, has been shown to act as an activator of genes including the pyruvate:ferredoxin oxidoreductase gene (Peeters et al., 2009). A genetic study has also been performed on a heat-shock-inducible ribonucleolytic toxin, VapC6, and its antitoxin VapB6. Analysis of disruption strains of these genes has identified possible targets of the ribonucleolytic activity (Maezato et al., 2011). Genetic manipulation is now also possible for the virus *Sulfolobus* turreted icosahedral virus (STIV) (Snyder et al., 2011; Wirth et al., 2011). An infectious clone of STIV was constructed, and gene disruptions of individual open reading frames and their effects on viral replication have been demonstrated.

Sulfolobus acidocaldarius

In *Sulfolobus acidocaldarius*, a series of small multicopy, non-integrative shuttle vectors have been developed and their use in overexpression of genes has been demonstrated (Berkner et al., 2007). Promoters for both constitutive and inducible gene expression have been examined. As for constitutive gene expression, the *sac7d* promoter led to the highest levels of β-galactosidase activity when various promoters were fused upstream of *lacS*. The *mal* promoter was the most suitable for induction, displaying a 17-fold increase upon addition of maltose or dextrin (Berkner et al., 2010). In terms of gene disruption, *pyrE*-deficient host cells have been used with an intact, heterologous *pyrE* from *S. solfataricus* to disrupt putative genes involved in UV photoproduct repair (Sakofsky et al., 2011). It should be noted that in this study, the lengths of the homologous regions flanking the selection marker were approximately 50 bp, introduced by PCR in the primer sequences, which may allow high-throughput gene disruption in a genome-wide scale. Another study clarified two *in vivo* activities of Y-family DNA polymerase in *S. acidocaldarius* (Sakofsky et al., 2012). One activity promotes slipped strand

events within simple repetitive sequences and the other promotes insertion of C opposite a potentially miscoding form of G, which may contribute in preventing G:C to T:A transversions. Genetics have contributed in the identification of sulfolobicins, antimicrobial proteins produced by *Sulfolobus* species (Ellen et al., 2011). Antimicrobial tests, protein separation, followed by MS led to the identification of candidate genes, and their disruption confirmed that two genes encoding secretion proteins corresponded to the sulfolobicin. Several studies report a genetic examination of genes involved in cell surface structure of *S. acidocaldarius*. Deletion of individual flagellin genes indicated that all genes were essential for flagellin assembly and that assembly proceeds through hierarchical protein interaction (Lassak et al., 2012). Another study has led to the identification of sulfoquinovose synthase, which is necessary for the synthesis of sulfoquinovose, a component of the *N*-linked glycans on the surface-layer glycoprotein of *S. acidocaldarius* (Meyer et al., 2011). Gene disruption confirmed this activity and also demonstrated the importance of *N*-glycosylation under conditions of increased salt concentrations. A complete genetic analysis of the three type IV pili-like structures in *S. acidocaldarius*, the flagellum, the UV-induced pili, and the adhesive pili, has also been reported (Henche et al., 2012). The effects of single, double, and triple deletion of the three structures on cell surface structure, surface attachment capability, motility, and biofilm formation were examined. It should be noted that this study utilizes cells expressing a codon adjusted, heat stable green fluorescent protein eCGP123. eCGP123 was used to distinguish strains within a biofilm generated from a mixture of strains with different gene deletions. Another study genetically demonstrates that the UV-inducible type IV pili are involved in intercellular, UV-inducible DNA exchange, a valuable mechanism to maintain chromosome integrity (Ajon et al., 2011).

Sulfolobus islandicus

In *Sulfolobus islandicus*, several genetic studies have been reported focusing on genes involved in DNA replication and maintenance of DNA topology. In one study, the topoisomerase III gene of *S. islandicus* was disrupted (Li et al., 2011a). Cells were viable but displayed various defects in chromosome distribution, cell size, and gene transcription. The results suggested that this enzyme plays an important role in chromosome segregation and maintenance of DNA topology for gene transcription. Another study addressed whether any of the three proliferating cell nuclear antigen (PCNA) on the *S. islandicus* genome are dispensable or not (Zhang et al., 2010). Disruption strains could not be isolated for any of the genes, and an improved knockout system has been described in order to carefully examine the essentiality of each gene. A recent study reported the development of a new gene disruption system in *S. islandicus* that is based on antibiotic resistance toward simvastatin (Zhang and Whitaker, 2012). The selectable marker gene was a construct promoting overexpression of the HMG-CoA reductase gene.

THERMOCOCCALES

Genetic systems have mainly been developed in *Thermococcus kodakarensis* and *Pyrococcus furiosus*. Gene disruption has also been demonstrated in *Thermococcus onnurineus* (Kim et al., 2010). Shuttle vectors are available for *Pyrococcus abyssi* (Lucas et al., 2002). In *T. kodakarensis*, gene disruption was accomplished by using host strains deleted of the *pyrF* and/or *trpE* genes, and selection with the corresponding intact marker gene (Sato et al., 2003, 2005). A system has also been developed based on the simvastatin/HMG-CoA reductase overexpression system (Matsumi et al., 2007). An improved system utilizing host cells that exhibit agmatine auxotrophy due to deletion of the arginine decarboxylase gene (Fukuda et al., 2008) has also been developed (Santangelo et al., 2010). This system allows selection in complex media and not only accelerates the gene disruption procedure, but should also contribute to the isolation of mutant cells that require nutrient-rich conditions for cell growth. Counterselection is performed with 5-FOA in the *pyrF* system (Sato et al., 2005), and counterselection in nutrient-rich medium is possible using a hypoxanthine–guanine phosphoribosyltransferase gene which, when present, results in 6-methylpurine sensitivity (Santangelo et al., 2010). In *Pyrococcus furiosus*, two transformation systems based on shuttle vectors that replicate in *P. furiosus* and *E. coli* have been developed. One is based on the shuttle vector system pYS2 from *P. abyssi*. The selectable marker is an HMG-CoA reductase overexpression cassette, and selection is based on resistance toward simvastatin (Waege et al., 2010). The other system is based on the *P. furiosus* chromosomal origin and utilizes the *pyrF* gene as a selectable marker in combination with the *P. furiosus* COM strain (Δ*pyrF*). The plasmids existed in a single copy in *P. furiosus*, and were stable without selective pressure for more than 100 generations (Farkas et al., 2011). The *pyrF* deletion strain *P. furiosus* COM1 can also be used as an efficient host for gene disruption (Lipscomb et al., 2011). This strain displays natural competence and a remarkable efficiency in DNA uptake, allowing marker replacement using linear as well as circular DNA. Using *pyrF* as the selectable marker, construction of markerless deletion mutants via counterselection by 5-FOA resistance has been demonstrated (Lipscomb et al., 2011). In a recent report, the limits of recombination efficiency of *P. furiosus* COM1 have been examined. It was found that marker replacement was possible with as few as 40 nucleotides of flanking homology to the target region (Farkas et al., 2012), which will surely facilitate genetic studies in this organism. Markerless deletion was utilized to disrupt the *trpAB* genes, encoding the two subunits of tryptophan synthase. The disruptant displayed tight tryptophan auxotrophy, and the wild-type *trpAB* genes could be used as a selectable marker in this strain (Farkas et al., 2012).

APPLICATION OF THE GENETIC SYSTEMS IN THERMOCOCCALES

Thermococcus kodakarensis and Thermococcus onnurineus

In *T. kodakarensis*, a wide range of genes has been disrupted in order to understand their physiological functions such as those involved in transcription and its regulation, DNA replication, and metabolism. The functions of individual transcription factors such as TFB1/2 (Santangelo et al., 2007), RNA polymerase subunits E and F (Hirata et al., 2008), the switch 3 loop of subunit B (Santangelo and Reeve, 2010) have been examined, along with sequences that can promote transcription termination (Santangelo et al., 2009). Deletion of transcription regulator genes, followed by transcriptome analysis, has led to

the identification of regulons and the function of these regulators (Kanai et al., 2007, 2010). In terms of DNA replication, a large number of proteins expected to be involved in DNA replication were His-tagged. Protein complexes were isolated and their components identified, revealing the various protein networks involved in DNA replication (Li et al., 2010). This led to the discovery of a novel GINS-associated nuclease, GAN (Li et al., 2011b). In other studies, genetic analyses of three *mcm* genes on the genome have revealed the essentiality/dispensability of the individual genes (Ishino et al., 2011; Pan et al., 2011). Disruption of the reverse gyrase gene indicated that the enzyme provides an advantage to cells when grown at temperatures of 80°C or higher (Atomi et al., 2004). In terms of metabolism, genes examined include those involved in glycolysis (Imanaka et al., 2006; Matsubara et al., 2011), gluconeogenesis (Sato et al., 2004), pentose metabolism (Orita et al., 2006), as well as coenzyme A (Yokooji et al., 2009), polyamine (Morimoto et al., 2010), and compatible solute biosynthesis (Borges et al., 2010). Gene disruption of three putative hydrogenase gene clusters and phenotypic analyses indicated that the cytosolic hydrogenase Hyh and the membrane-bound oxidoreductase complexes Mbh and Mbx are involved in H_2 consumption, H_2 generation, and H_2S generation, respectively (Kanai et al., 2011). Disruption of various pathways related to hydrogen production/consumption has clarified the reductant flux in *T. kodakarensis* and has demonstrated strategies to elevate hydrogen production in this organism (Santangelo et al., 2011). *T. kodakarensis* harbors two pairs of genes encoding chaperonins and prefoldins, and gene disruption has been performed to distinguish the functions of the individual proteins at different temperatures (Danno et al., 2008; Fujiwara et al., 2008; Gao et al., 2012). One of multiple homologs of NAD(P)H oxidase has been disrupted to examine its relationship with the oxygen sensitivity of this anaerobic archaeon (Kobori et al., 2012). Systems for gene overexpression, tagging, and protein secretion have been established (Santangelo et al., 2008; Mueller et al., 2009; Yokooji et al., 2009; Takemasa et al., 2011), in many cases relying on the selection/counterselection strategy utilizing the *pyrF* marker gene (**Figure 3**). In *T. onnurineus*, gene disruption has been demonstrated using the simvastatin/HMG-CoA

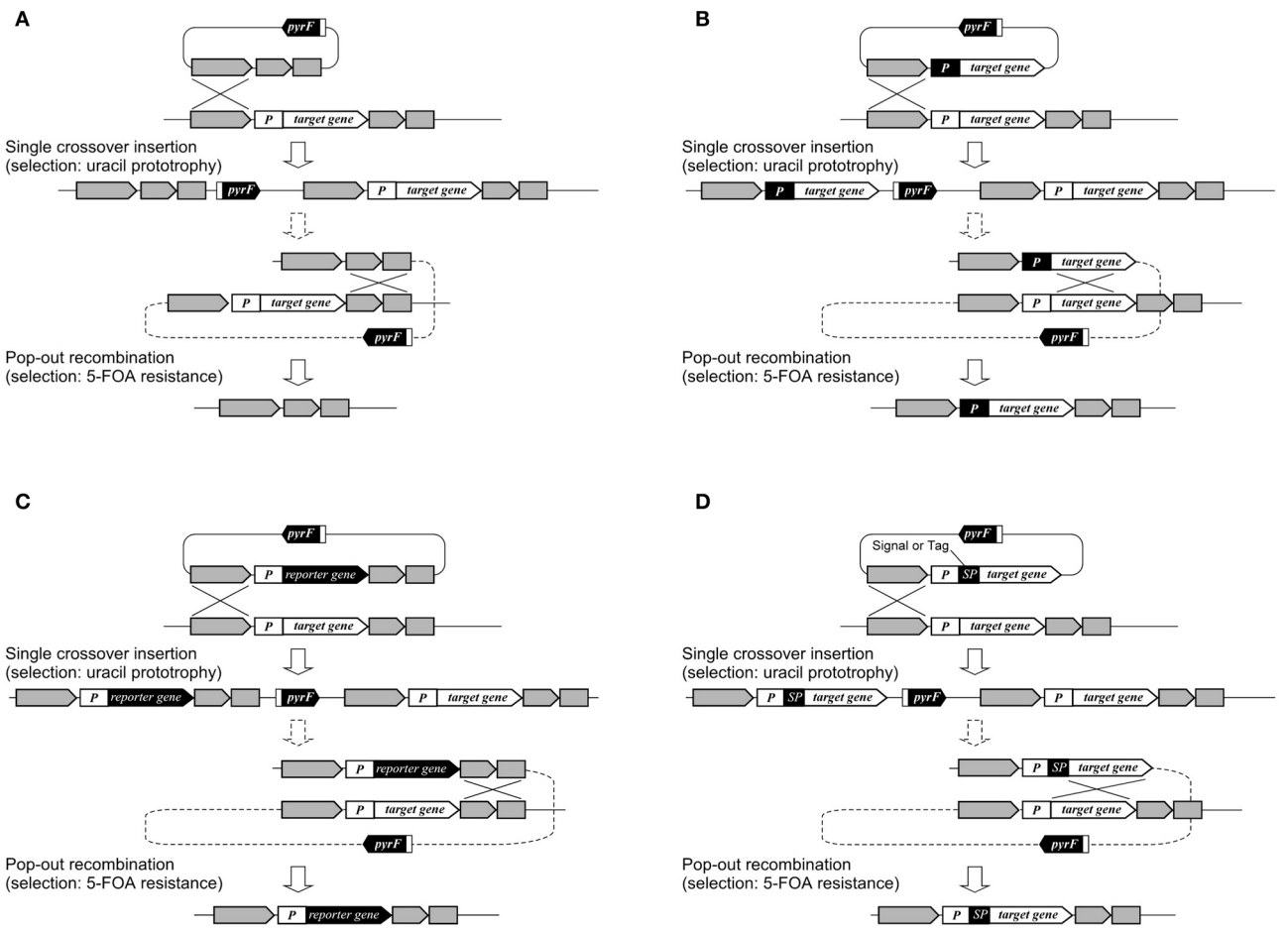


FIGURE 3 | Various applications using selection/counterselection marker genes. (A) Use in markerless gene disruption; **(B)** use in promoter exchange; **(C)** use in reporter gene integration; and **(D)** use in signal peptide (SP) or tag integration. The use of *pyrF* is shown, but other selection/counterselection marker genes can be applied in a similar manner in a wide range of Archaea.

reductase overexpression system. Gene disruption led to the identification of the gene cluster encoding formate hydrogen lyase, cation/proton antiporter and formate transporter, which were responsible for the growth of this organism on formate (Kim et al., 2010).

Pyrococcus furiosus

Using the shuttle vector pYS3, the RNA polymerase subunit D gene with a HisTag sequence was expressed with an inducible promoter deriving from the fructose-1,6-bisphosphatase gene from *P. furiosus*. This allowed a simple two-step purification of the thermostable RNA polymerase from this organism (Waege et al., 2010). In the process of developing the shuttle vector system based on the *P. furiosus* chromosomal origin, the minimum replication origin sequence required for autonomous plasmid replication in this organism has been identified (Farkas et al., 2011). Interestingly, the *cdc6/orc1* gene adjacent to *oriC* was not required in *cis* for replication of the shuttle vector in *P. furiosus*. The gene disruption system developed with *P. furiosus* COM1 has been successfully applied in disrupting individual or double gene disruptions of two cytoplasmic hydrogenase genes (Lipscomb et al., 2011). The system was further applied for detailed genetic studies on proteins related to elemental sulfur metabolism, membrane-bound oxidoreductase complex (Mbx), cytoplasmic coenzyme A-dependent NADPH:sulfur oxidoreductase (Nsr), and sulfur-induced protein A (SipA) (Bridger et al., 2011). The *mbx* disruptant displayed growth defects in the presence of sulfur, and little, if any, sulfide generation was observed, demonstrating that Mbx plays a critical role in elemental sulfur reduction and energy conservation in *P. furiosus*. Gene manipulation has also been used to overexpress the cytoplasmic [NiFe]-hydrogenase SHI. The promoter of the PF1399 gene, which encodes the S-layer protein, was fused upstream of the four-gene operon (PF0891–PF0894) encoding SHI. In the overexpression strain, a 20-fold higher SHI transcript level was observed, and moreover, a 100-fold higher amount of hydrogenase was obtained when compared with the highest homologous [NiFe]-hydrogenase system previously reported (Chandrayan et al., 2012). In another study, the lactate dehydrogenase gene from the moderately thermophilic *Caldicellulosiruptor bescii* was introduced into *P. furiosus* under the control of the cold-induced protein A (*cipA*, PF0190) promoter. Transcript levels of *cipA* in *P. furiosus* are 26-fold higher in cells grown at 72°C compared to those grown at 98°C. This recombinant strain, when grown at 98°C, ferments sugar to produce acetate and hydrogen as end products, as is the case of wild-type *P. furiosus*. When grown at 72°C, however, the strain generates lactate at concentrations up to 3 mM, demonstrating a temperature-dependent regulation of metabolism (Basen et al., 2012).

FUTURE PERSPECTIVES

The genetic systems developed in the halophiles, methanogens, *Sulfolobus*, and Thermococcales provide the tools to carry out sophisticated genetic analyses in these organisms (Leigh et al., 2011). With the abundance of genome sequence information, functional genomics in these organisms is a realistic approach. On the other hand, the Archaea comprise a diverse group of organisms, and there are still many interesting organisms that cannot be examined genetically. Genetics are limited to *Sulfolobus* in the Crenarchaeota, and considering the wealth of genome sequence information, the development of genetic systems in *Pyrobaculum* and *Thermoproteus* would greatly promote research in these genera. Genetic tools for *Ignicoccus* can be considered crucial to understand its unique relationship with *Nanoarchaeum equitans*. In the Euryarchaeota, the thermophilic/acidophilic Thermoplasmatales and the sulfate-reducing Archaeoglobales are major orders in which genetic tools are still not available. Another major group is the (hyper)thermophilic methanogens. Although much needs to be done, the variety of tools that have been developed will surely provide a basis to explore the possibilities of developing genetic systems in other archaea.

REFERENCES

Aagaard, C., Leviev, I., Aravalli, R. N., Forterre, P., Prieur, D., and Garrett, R. A. (1996). General vectors for archaeal hyperthermophiles: strategies based on a mobile intron and a plasmid. *FEMS Microbiol. Rev.* 18, 93–104.

Ajon, M., Fröls, S., Van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A. J., Grogan, D. W., Albers, S. V., and Schleper, C. (2011). UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817.

Albers, S., and Driessen, A. J. (2007). Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. *Archaea* 2, 145–149.

Albers, S. V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampe, R., Driessen, A. J., and Schleper, C. (2006). Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl. Environ. Microbiol.* 72, 102–111.

Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76, 1759–1769.

Allers, T., and Mevarech, M. (2005). Archaeal genetics - the third way. *Nat. Rev. Genet.* 6, 58–73.

Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes. *Appl. Environ. Microbiol.* 70, 943–953.

Aravalli, R. N., and Garrett, R. A. (1997). Shuttle vectors for hyperthermophilic archaea. *Extremophiles* 1, 183–191.

Argyle, J. L., Tumbula, D. L., and Leigh, J. A. (1996). Neomycin resistance as a selectable marker in *Methanococcus maripaludis*. *Appl. Environ. Microbiol.* 62, 4233–4237.

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

Aucelli, T., Contursi, P., Girfoglio, M., Rossi, M., and Cannio, R. (2006). A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*. *Nucleic Acids Res.* 34, e114.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 2006.0008.

Basen, M., Sun, J., and Adams, M. W. (2012). Engineering a hyperthermophilic archaeon for temperature-dependent product formation. *MBio* 3, e00053-00012.

Berghöfer, Y., and Klein, A. (1995). Insertional mutations in the hydrogenase *vhc* and *frc* operons encoding selenium-free hydrogenases in *Methanococcus voltae*. *Appl. Environ. Microbiol.* 61, 1770–1775.

Berkner, S., Grogan, D., Albers, S. V., and Lipps, G. (2007). Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren-)archaea. *Nucleic Acids Res.* 35, e88.

Berkner, S., and Lipps, G. (2008). Genetic tools for *Sulfolobus* spp.: vectors and first applications. *Arch. Microbiol.* 190, 217–230.

Berkner, S., Wlodkowski, A., Albers, S. V., and Lipps, G. (2010). Inducible and constitutive promoters for genetic systems in *Sulfolobus acidocaldarius*. *Extremophiles* 14, 249–259.

Bertani, G., and Baresi, L. (1987). Genetic transformation in the methanogen *Methanococcus voltae* PS. *J. Bacteriol.* 169, 2730–2738.

Bitan-Banin, G., Ortenberg, R., and Mevarech, M. (2003). Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene. *J. Bacteriol.* 185, 772–778.

Blank, C. E., Kessler, P. S., and Leigh, J. A. (1995). Genetics in methanogens: transposon insertion mutagenesis of a *Methanococcus maripaludis nifH* gene. *J. Bacteriol.* 177, 5773–5777.

Blaseio, U., and Pfeifer, F. (1990). Transformation of *Halobacterium halobium*: development of vectors and investigation of gas vesicle synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 87, 6772–6776.

Boeke, J. D., Lacroute, F., and Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Mol. Gen. Genet.* 197, 345–346.

Boeke, J. D., Trueheart, J., Natsoulis, G., and Fink, G. R. (1987). 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* 154, 164–175.

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

Bowen, T. L., Lin, W. C., and Whitman, W. B. (1996). Characterization of guanine and hypoxanthine phosphoribosyltransferases in *Methanococcus voltae*. *J. Bacteriol.* 178, 2521–2526.

Bowen, T. L., and Whitman, W. B. (1987). Incorporation of exogenous purines and pyrimidines by *Methanococcus voltae* and isolation of analog-resistant mutants. *Appl. Environ. Microbiol.* 53, 1822–1826.

Bridger, S. L., Clarkson, S. M., Stirrett, K., Debarry, M. B., Lipscomb, G. L., Schut, G. J., Westpheling, J., Scott, R. A., and Adams, M. W. (2011). Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus*. *J. Bacteriol.* 193, 6498–6504.

Buan, N., Kulkarni, G., and Metcalf, W. (2011). Genetic methods for *Methanosarcina* species. *Methods Enzymol.* 494, 23–42.

Cai, S., Cai, L., Liu, H., Liu, X., Han, J., Zhou, J., and Xiang, H. (2012). Identification of the haloarchaeal phasin (PhaP) that functions in polyhydroxyalkanoate accumulation and granule formation in *Haloferax mediterranei*. *Appl. Environ. Microbiol.* 78, 1946–1952.

Cannio, R., Contursi, P., Rossi, M., and Bartolucci, S. (1998). An autonomously replicating transforming vector for *Sulfolobus solfataricus*. *J. Bacteriol.* 180, 3237–3240.

Chaban, B., Logan, S. M., Kelly, J. F., and Jarrell, K. F. (2009). AglC and AglK are involved in biosynthesis and attachment of diacetylated glucuronic acid to the *N*-glycan in *Methanococcus voltae*. *J. Bacteriol.* 191, 187–195.

Chandrayan, S. K., McTernan, P. M., Hopkins, R. C., Sun, J., Jenney, F. E. Jr., and Adams, M. W. (2012). Engineering hyperthermophilic archaeon *Pyrococcus furiosus* to overproduce its cytoplasmic [NiFe]-hydrogenase. *J. Biol. Chem.* 287, 3257–3264.

Cline, S. W., and Doolittle, W. F. (1992). Transformation of members of the genus *Haloarcula* with shuttle vectors based on *Halobacterium halobium* and *Haloferax volcanii* plasmid replicons. *J. Bacteriol.* 174, 1076–1080.

Cohen-Kupiec, R., Blank, C., and Leigh, J. A. (1997). Transcriptional regulation in Archaea: *in vivo* demonstration of a repressor binding site in a methanogen. *Proc. Natl. Acad. Sci. U.S.A.* 94, 1316–1320.

Coker, J. A., and Dassarma, S. (2007). Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC Genet.* 8, 61.

Danno, A., Fukuda, W., Yoshida, M., Aki, R., Tanaka, T., Kanai, T., Imanaka, T., and Fujiwara, S. (2008). Expression profiles and physiological roles of two types of prefoldins from the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Mol. Biol.* 382, 298–311.

Deng, L., Zhu, H., Chen, Z., Liang, Y. X., and She, Q. (2009). Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles* 13, 735–746.

Dodsworth, J. A., and Leigh, J. A. (2006). Regulation of nitrogenase by 2-oxoglutarate-reversible, direct binding of a PII-like nitrogen sensor protein to dinitrogenase. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9779–9784.

Dummer, A. M., Bonsall, J. C., Cihla, J. B., Lawry, S. M., Johnson, G. C., and Peck, R. F. (2011). Bacterioopsin-mediated regulation of bacterioruberin biosynthesis in *Halobacterium salinarum*. *J. Bacteriol.* 193, 5658–5667.

Ehlers, C., Weidenbach, K., Veit, K., Deppenmeier, U., Metcalf, W. W., and Schmitz, R. A. (2005). Development of genetic methods and construction of a chromosomal *glnK1* mutant in *Methanosarcina mazei* strain Gö1. *Mol. Genet. Genomics* 273, 290–298.

Elferink, M. G., Schleper, C., and Zillig, W. (1996). Transformation of the extremely thermoacidophilic archaeon *Sulfolobus solfataricus* via a self-spreading vector. *FEMS Microbiol. Lett.* 137, 31–35.

Ellen, A. F., Rohulya, O. V., Fusetti, F., Wagner, M., Albers, S. V., and Driessen, A. J. (2011). The sulfolobicin genes of *Sulfolobus acidocaldarius* encode novel antimicrobial proteins. *J. Bacteriol.* 193, 4380–4387.

Farkas, J., Chung, D., Debarry, M., Adams, M. W., and Westpheling, J. (2011). Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon *Pyrococcus furiosus* needed for construction of a stable replicating shuttle vector. *Appl. Environ. Microbiol.* 77, 6343–6349.

Farkas, J., Stirrett, K., Lipscomb, G. L., Nixon, W., Scott, R. A., Adams, M. W., and Westpheling, J. (2012). Recombinogenic properties of the *Pyrococcus furiosus* COM1 strain enable rapid selection of targeted mutants. *Appl. Environ. Microbiol.* 78, 4669–4676.

Fischer, S., John Von Freyend, S., Sabag-Daigle, A., Daniels, C. J., Allers, T., and Marchfelder, A. (2012). Assigning a function to a conserved archaeal metallo-β-lactamase from *Haloferax volcanii*. *Extremophiles* 16, 333–343.

Fujiwara, S., Aki, R., Yoshida, M., Higashibata, H., Imanaka, T., and Fukuda, W. (2008). Expression profiles and physiological roles of two types of molecular chaperonins from the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Appl. Environ. Microbiol.* 74, 7306–7312.

Fukuda, W., Morimoto, N., Imanaka, T., and Fujiwara, S. (2008). Agmatine is essential for the cell growth of *Thermococcus kodakaraensis*. *FEMS Microbiol. Lett.* 287, 113–120.

Gao, L., Danno, A., Fujii, S., Fukuda, W., Imanaka, T., and Fujiwara, S. (2012). Indole-3-glycerol-phosphate synthase is recognized by a cold-inducible group II chaperonin in *Thermococcus kodakarensis*. *Appl. Environ. Microbiol.* 78, 3806–3815.

Gardner, W. L., and Whitman, W. B. (1999). Expression vectors for *Methanococcus maripaludis*: overexpression of acetohydroxyacid synthase and β-galactosidase. *Genetics* 152, 1439–1447.

Gernhardt, P., Possot, O., Foglino, M., Sibold, L., and Klein, A. (1990). Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. *Mol. Gen. Genet.* 221, 273–279.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A. P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K. D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kotter, P., Labonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C. Y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W., and Johnston, M. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391.

Grochowski, L. L., Xu, H., and White, R. H. (2006). *Methanocaldococcus jannaschii* uses a modified

mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J. Bacteriol.* 188, 3192–3198.

Heinicke, I., Müller, J., Pittelkow, M., and Klein, A. (2004). Mutational analysis of genes encoding chromatin proteins in the archaeon *Methanococcus voltae* indicates their involvement in the regulation of gene expression. *Mol. Genet. Genomics* 272, 76–87.

Henche, A. L., Koerdt, A., Ghosh, A., and Albers, S. V. (2012). Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. *Environ. Microbiol.* 14, 779–793.

Hirata, A., Kanai, T., Santangelo, T. J., Tajiri, M., Manabe, K., Reeve, J. N., Imanaka, T., and Murakami, K. S. (2008). Archaeal RNA polymerase subunits E and F are not required for transcription *in vitro*, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol. Microbiol.* 70, 623–633.

Hohn, M. J., Palioura, S., Su, D., Yuan, J., and Söll, D. (2011). Genetic analysis of selenocysteine biosynthesis in the archaeon *Methanococcus maripaludis*. *Mol. Microbiol.* 81, 249–258.

Holmes, M., Pfeifer, F., and Dyall-Smith, M. (1994). Improved shuttle vectors for *Haloferax volcanii* including a dual-resistance plasmid. *Gene* 146, 117–121.

Holmes, M. L., and Dyall-Smith, M. L. (1990). A plasmid vector with a selectable marker for halophilic archaebacteria. *J. Bacteriol.* 172, 756–761.

Holmes, M. L., Nuttall, S. D., and Dyall-Smith, M. L. (1991). Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. *J. Bacteriol.* 173, 3807–3813.

Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H., and Imanaka, T. (2006). Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.* 61, 898–909.

Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., Kanai, T., Atomi, H., and Ishino, Y. (2011). Biochemical and genetical analyses of the three *mcm* genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells* 16, 1176–1189.

Jonuscheit, M., Martusewitsch, E., Stedman, K. M., and Schleper, C. (2003). A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol.* 48, 1241–1252.

Kanai, T., Akerboom, J., Takedomi, S., Van De Werken, H. J., Blombach, F., Van Der Oost, J., Murakami, T., Atomi, H., and Imanaka, T. (2007). A global transcriptional regulator in *Thermococcus kodakaraensis* controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. *J. Biol. Chem.* 282, 33659–33670.

Kanai, T., Matsuoka, R., Beppu, H., Nakajima, A., Okada, Y., Atomi, H., and Imanaka, T. (2011). Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *J. Bacteriol.* 193, 3109–3116.

Kanai, T., Takedomi, S., Fujiwara, S., Atomi, H., and Imanaka, T. (2010). Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. *J. Biochem.* 147, 361–370.

Kessler, P. S., Daniel, C., and Leigh, J. A. (2001). Ammonia switch-off of nitrogen fixation in the methanogenic archaeon *Methanococcus maripaludis*: mechanistic features and requirement for the novel GlnB homologues, NifI(1) and NifI(2). *J. Bacteriol.* 183, 882–889.

Kessler, P. S., and Leigh, J. A. (1999). Genetics of nitrogen regulation in *Methanococcus maripaludis*. *Genetics* 152, 1343–1351.

Kim, W., and Whitman, W. B. (1999). Isolation of acetate auxotrophs of the methane-producing archaeon *Methanococcus maripaludis* by random insertional mutagenesis. *Genetics* 152, 1429–1437.

Kim, Y. J., Lee, H. S., Kim, E. S., Bae, S. S., Lim, J. K., Matsumi, R., Lebedinsky, A. V., Sokolova, T. G., Kozhevnikova, D. A., Cha, S. S., Kim, S. J., Kwon, K. K., Imanaka, T., Atomi, H., Bonch-Osmolovskaya, E. A., Lee, J. H., and Kang, S. G. (2010). Formate-driven growth coupled with H_2 production. *Nature* 467, 352–355.

Kixmüller, D., and Greie, J. C. (2012). Construction and characterization of a gradually inducible expression vector for *Halobacterium salinarum*, based on the *kdp* promoter. *Appl. Environ. Microbiol.* 78, 2100–2105.

Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P., Boland, F., Brignell, S. C., Bron, S., Bunai, K., Chapuis, J., Christiansen, L. C., Danchin, A., Debarbouille, M., Dervyn, E., Deuerling, E., Devine, K., Devine, S. K., Dreesen, O., Errington, J., Fillinger, S., Foster, S. J., Fujita, Y., Galizzi, A., Gardan, R., Eschevins, C., Fukushima, T., Haga, K., Harwood, C. R., Hecker, M., Hosoya, D., Hullo, M. F., Kakeshita, H., Karamata, D., Kasahara, Y., Kawamura, F., Koga, K., Koski, P., Kuwana, R., Imamura, D., Ishimaru, M., Ishikawa, S., Ishio, I., Le Coq, D., Masson, A., Mauel, C., Meima, R., Mellado, R. P., Moir, A., Moriya, S., Nagakawa, E., Nanamiya, H., Nakai, S., Nygaard, P., Ogura, M., Ohanan, T., O'Reilly, M., O'Rourke, M., Pragai, Z., Pooley, H. M., Rapoport, G., Rawlins, J. P., Rivas, L. A., Rivolta, C., Sadaie, A., Sadaie, Y., Sarvas, M., Sato, T., Saxild, H. H., Scanlan, E., Schumann, W., Seegers, J. F., Sekiguchi, J., Sekowska, A., Seror, S. J., Simon, M., Stragier, P., Studer, R., Takamatsu, H., Tanaka, T., Takeuchi, M., Thomaides, H. B., Vagner, V., Van Dijl, J. M., Watabe, K., Wipat, A., Yamamoto, H., Yamamoto, M., Yamamoto, Y., Yamane, K., Yata, K., Yoshida, K., Yoshikawa, H., Zuber, U., and Ogasawara, N. (2003). Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4678–4683.

Kobori, H., Ogino, M., Orita, I., Nakamura, S., Imanaka, T., and Fukui, T. (2012). Characterization of NADH oxidase/NADPH polysulfide oxidoreductase and its unexpected participation in oxygen sensitivity in an anaerobic hyperthermophilic archaeon. *J. Bacteriol.* 192, 5192–5202.

Ladapo, J., and Whitman, W. B. (1990). Method for isolation of auxotrophs in the methanogenic archaebacteria: role of the acetyl-CoA pathway of autotrophic CO_2 fixation in *Methanococcus maripaludis*. *Proc. Natl. Acad. Sci. U.S.A.* 87, 5598–5602.

Lam, W. L., and Doolittle, W. F. (1989). Shuttle vectors for the archaebacterium *Halobacterium volcanii*. *Proc. Natl. Acad. Sci. U.S.A.* 86, 5478–5482.

Lam, W. L., and Doolittle, W. F. (1992). Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*. *J. Biol. Chem.* 267, 5829–5834.

Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S. V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124.

Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol. Rev.* 35, 577–608.

Li, X., Guo, L., Deng, L., Feng, D., Ren, Y., Chu, Y., She, Q., and Huang, L. (2011a). Deletion of the topoisomerase III gene in the hyperthermophilic archaeon *Sulfolobus islandicus* results in slow growth and defects in cell cycle control. *J. Genet. Genomics* 38, 253–259.

Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz, J., Reeve, J. N., and Kelman, Z. (2011b). A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res.* 39, 6114–6123.

Li, Z., Santangelo, T. J., Čuboňová, L., Reeve, J. N., and Kelman, Z. (2010). Affinity purification of an archaeal DNA replication protein network. *MBio* 1, e00221-10.

Lie, T. J., Hendrickson, E. L., Niess, U. M., Moore, B. C., Haydock, A. K., and Leigh, J. A. (2010). Overlapping repressor binding sites regulate expression of the *Methanococcus maripaludis* $glnK_1$ operon. *Mol. Microbiol.* 75, 755–762.

Lie, T. J., and Leigh, J. A. (2003). A novel repressor of *nif* and *glnA* expression in the methanogenic archaeon *Methanococcus maripaludis*. *Mol. Microbiol.* 47, 235–246.

Lie, T. J., Wood, G. E., and Leigh, J. A. (2005). Regulation of *nif* expression in *Methanococcus maripaludis*: roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators. *J. Biol. Chem.* 280, 5236–5241.

Lipscomb, G. L., Stirrett, K., Schut, G. J., Yang, F., Jenney, F. E. Jr., Scott, R. A., Adams, M. W., and Westpheling, J. (2011). Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* 77, 2232–2238.

Liu, H., Han, J., Liu, X., Zhou, J., and Xiang, H. (2011). Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *J. Genet. Genomics* 38, 261–269.

Lucas, S., Toffin, L., Zivanovic, Y., Charlier, D., Moussard, H., Forterre,

P., Prieur, D., and Erauso, G. (2002). Construction of a shuttle vector for, and spheroplast transformation of, the hyperthermophilic archaeon *Pyrococcus abyssi*. *Appl. Environ. Microbiol.* 68, 5528–5536.

Maezato, Y., Daugherty, A., Dana, K., Soo, E., Cooper, C., Tachdjian, S., Kelly, R. M., and Blum, P. (2011). VapC6, a ribonucleolytic toxin regulates thermophilicity in the crenarchaeote *Sulfolobus solfataricus*. *RNA* 17, 1381–1392.

Mahapatra, A., Patel, A., Soares, J. A., Larue, R. C., Zhang, J. K., Metcalf, W. W., and Krzycki, J. A. (2006). Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine. *Mol. Microbiol.* 59, 56–66.

Major, T. A., Liu, Y., and Whitman, W. B. (2010). Characterization of energy-conserving hydrogenase B in *Methanococcus maripaludis*. *J. Bacteriol.* 192, 4022–4030.

Matsubara, K., Yokooji, Y., Atomi, H., and Imanaka, T. (2011). Biochemical and genetic characterization of the three metabolic routes in *Thermococcus kodakarensis* linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. *Mol. Microbiol.* 81, 1300–1312.

Matsumi, R., Atomi, H., Driessen, A. J., and Van Der Oost, J. (2011). Isoprenoid biosynthesis in Archaea–biochemical and evolutionary implications. *Res. Microbiol.* 162, 39–52.

Matsumi, R., Manabe, K., Fukui, T., Atomi, H., and Imanaka, T. (2007). Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. *J. Bacteriol.* 189, 2683–2691.

Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R., and Wolfe, R. S. (1997). A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2626–2631.

Meyer, B. H., Zolghadr, B., Peyfoon, E., Pabst, M., Panico, M., Morris, H. R., Haslam, S. M., Messner, P., Schäffer, C., Dell, A., and Albers, S. V. (2011). Sulfoquinovose synthase-an important enzyme in the *N*-glycosylation pathway of *Sulfolobus acidocaldarius*. *Mol. Microbiol.* 82, 1150–1163.

Moore, B. C., and Leigh, J. A. (2005). Markerless mutagenesis in *Methanococcus maripaludis* demonstrates roles for alanine dehydrogenase, alanine racemase, and alanine permease. *J. Bacteriol.* 187, 972–979.

Morimoto, N., Fukuda, W., Nakajima, N., Masuda, T., Terui, Y., Kanai, T., Oshima, T., Imanaka, T., and Fujiwara, S. (2010). Dual biosynthesis pathway for longer-chain polyamines in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *J. Bacteriol.* 192, 4991–5001.

Mueller, M., Takemasa, R., Schwarz, A., Atomi, H., and Nidetzky, B. (2009). "Short-chain" α-1, 4-glucan phosphorylase having a truncated N-terminal domain: functional expression and characterization of the enzyme from *Sulfolobus solfataricus*. *Biochim. Biophys. Acta* 1794, 1709–1714.

Naparstek, S., Guan, Z., and Eichler, J. (2012). A predicted geranylgeranyl reductase reduces the ω-position isoprene of dolichol phosphate in the halophilic archaeon, *Haloferax volcanii*. *Biochim. Biophys. Acta* 1821, 923–933.

Ng, S. Y., Wu, J., Nair, D. B., Logan, S. M., Robotham, A., Tessier, L., Kelly, J. F., Uchida, K., Aizawa, S., and Jarrell, K. F. (2011). Genetic and mass spectrometry analyses of the unusual type IV-like pili of the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 193, 804–814.

Orita, I., Sato, T., Yurimoto, H., Kato, N., Atomi, H., Imanaka, T., and Sakai, Y. (2006). The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 188, 4698–4704.

Ozawa, K., Harashina, T., Yatsunami, R., and Nakamura, S. (2005). Gene cloning, expression and partial characterization of cell division protein FtsZ1 from extremely halophilic archaeon *Haloarcula japonica* strain TR-1. *Extremophiles* 9, 281–288.

Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N., and Kelman, Z. (2011). *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res.* 39, 9671–9680.

Patel, G. B., Nash, J. H., Agnew, B. J., and Sprott, G. D. (1994). Natural and electroporation-mediated transformation of *Methanococcus voltae* protoplasts. *Appl. Environ. Microbiol.* 60, 903–907.

Peck, R. F., Dassarma, S., and Krebs, M. P. (2000). Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Mol. Microbiol.* 35, 667–676.

Peeters, E., Albers, S. V., Vassart, A., Driessen, A. J., and Charlier, D. (2009). Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase-encoding operon and permease genes. *Mol. Microbiol.* 71, 972–988.

Pfeifer, M., Bestgen, H., Bürger, A., and Klein, A. (1998). The *vhuU* gene encoding a small subunit of a selenium-containing [NiFe]-hydrogenase in *Methanococcus voltae* appears to be essential for the cell. *Arch. Microbiol.* 170, 418–426.

Pickl, A., Johnsen, U., and Schönheit, P. (2012). Fructose degradation in the haloarchaeon *Haloferax volcanii* involves bacterial type phosphoenolpyruvate-dependent phosphotransferase system, fructose-1-phosphate kinase and Class II fructose-1, 6-bisphosphate aldolase. *J. Bacteriol.* 194, 3088–3097.

Porat, I., Kim, W., Hendrickson, E. L., Xia, Q., Zhang, Y., Wang, T., Taub, F., Moore, B. C., Anderson, I. J., Hackett, M., Leigh, J. A., and Whitman, W. B. (2006). Disruption of the operon encoding Ehb hydrogenase limits anabolic CO_2 assimilation in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 188, 1373–1380.

Prangishvili, D., Albers, S. V., Holz, I., Arnold, H. P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A., Kristjansson, J. K., and Zillig, W. (1998). Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. *Plasmid* 40, 190–202.

Pritchett, M. A., Zhang, J. K., and Metcalf, W. W. (2004). Development of a markerless genetic exchange method for *Methanosarcina acetivorans* C2A and its use in construction of new genetic tools for methanogenic archaea. *Appl. Environ. Microbiol.* 70, 1425–1433.

Rother, M., and Metcalf, W. W. (2005). Genetic technologies for Archaea. *Curr. Opin. Microbiol.* 8, 745–751.

Sakofsky, C. J., Foster, P. L., and Grogan, D. W. (2012). Roles of the Y-family DNA polymerase Dbh in accurate replication of the *Sulfolobus* genome at high temperature. *DNA Repair (Amst.)* 11, 391–400.

Sakofsky, C. J., Runck, L. A., and Grogan, D. W. (2011). *Sulfolobus* mutants, generated via PCR products, which lack putative enzymes of UV photoproduct repair. *Archaea* 2011, ID: 864015.

Sandbeck, K. A., and Leigh, J. A. (1991). Recovery of an integration shuttle vector from tandem repeats in *Methanococcus maripaludis*. *Appl. Environ. Microbiol.* 57, 2762–2763.

Santangelo, T. J., Čuboňová, L., James, C. L., and Reeve, J. N. (2007). TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription *in vitro*. *J. Mol. Biol.* 367, 344–357.

Santangelo, T. J., Čuboňová, L., and Reeve, J. N. (2008). Shuttle vector expression in *Thermococcus kodakaraensis*: contributions of *cis* elements to protein synthesis in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 74, 3099–3104.

Santangelo, T. J., Čuboňová, L., and Reeve, J. N. (2010). *Thermococcus kodakarensis* genetics: TK1827-encoded β-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl. Environ. Microbiol.* 76, 1044–1052.

Santangelo, T. J., Čuboňová, L., and Reeve, J. N. (2011). Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* 81, 897–911.

Santangelo, T. J., Čuboňová, L., Skinner, K. M., and Reeve, J. N. (2009). Archaeal intrinsic transcription termination *in vivo*. *J. Bacteriol.* 191, 7102–7108.

Santangelo, T. J., and Reeve, J. N. (2010). Deletion of switch 3 results in an archaeal RNA polymerase that is defective in transcript elongation. *J. Biol. Chem.* 285, 23908–23915.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* 185, 210–220.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2005). Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.* 71, 3889–3899.

Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H., and Imanaka, T. (2004). Genetic evidence identifying the true gluconeogenic fructose-1, 6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.* 186, 5799–5807.

Saum, R., Schlegel, K., Meyer, B., and Muller, V. (2009). The F_1F_0 ATP

synthase genes in *Methanosarcina acetivorans* are dispensable for growth and ATP synthesis. *FEMS Microbiol. Lett.* 300, 230–236.

Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M., and Blum, P. (2004). Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J. Bacteriol.* 186, 427–437.

Schelert, J., Drozda, M., Dixit, V., Dillman, A., and Blum, P. (2006). Regulation of mercury resistance in the crenarchaeote *Sulfolobus solfataricus*. *J. Bacteriol.* 188, 7141–7150.

She, Q., Zhang, C., Deng, L., Peng, N., Chen, Z., and Liang, Y. X. (2009). Genetic analyses in the hyperthermophilic archaeon *Sulfolobus islandicus*. *Biochem. Soc. Trans.* 37, 92–96.

Skowyra, A., and Macneill, S. A. (2012). Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res.* 40, 1077–1090.

Snyder, J. C., Brumfield, S. K., Peng, N., She, Q., and Young, M. J. (2011). *Sulfolobus* turreted icosahedral virus c92 protein responsible for the formation of pyramid-like cellular lysis structures. *J. Virol.* 85, 6287–6292.

Soppa, J. (2011). Functional genomic and advanced genetic studies reveal novel insights into the metabolism, regulation, and biology of *Haloferax volcanii*. *Archaea* 2011, ID: 602408.

Stedman, K. M., Schleper, C., Rumpf, E., and Zillig, W. (1999). Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* 152, 1397–1405.

Stedman, K. M., She, Q., Phan, H., Holz, I., Singh, H., Prangishvili, D., Garrett, R., and Zillig, W. (2000). pING family of conjugative plasmids from the extremely thermophilic archaeon *Sulfolobus islandicus*: insights into recombination and conjugation in Crenarchaeota. *J. Bacteriol.* 182, 7014–7020.

Stock, T., Selzer, M., Connery, S., Seyhan, D., Resch, A., and Rother, M. (2011). Disruption and complementation of the selenocysteine biosynthesis pathway reveals a hierarchy of selenoprotein gene expression in the archaeon *Methanococcus maripaludis*. *Mol. Microbiol.* 82, 734–747.

Stock, T., Selzer, M., and Rother, M. (2010). *In vivo* requirement of selenophosphate for selenoprotein synthesis in archaea. *Mol. Microbiol.* 75, 149–160.

Takemasa, R., Yokooji, Y., Yamatsu, A., Atomi, H., and Imanaka, T. (2011). *Thermococcus kodakarensis* as a host for gene expression and protein secretion. *Appl. Environ. Microbiol.* 77, 2392–2398.

Tu, D., Blaha, G., Moore, P. B., and Steitz, T. A. (2005). Gene replacement in *Haloarcula marismortui:* construction of a strain with two of its three chromosomal rRNA operons deleted. *Extremophiles* 9, 427–435.

Tumbula, D. L., Bowen, T. L., and Whitman, W. B. (1997). Characterization of pURB500 from the archaeon *Methanococcus maripaludis* and construction of a shuttle vector. *J. Bacteriol.* 179, 2976–2986.

Tumbula, D. L., Makula, R. A., and Whitman, W. B. (1994). Transformation of *Methanococcus maripaludis* and identification of a *Pst* I-like restriction system. *FEMS Microbiol. Lett.* 121, 309–314.

Tumbula, D. L., and Whitman, W. B. (1999). Genetics of Methanococcus: possibilities for functional genomics in Archaea. *Mol. Microbiol.* 33, 1–7.

Vandyke, D. J., Wu, J., Ng, S. Y., Kanbe, M., Chaban, B., Aizawa, S., and Jarrell, K. F. (2008). Identification of a putative acetyltransferase gene, MMP0350, which affects proper assembly of both flagella and pili in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 190, 5300–5307.

Villafane, A., Voskoboynik, Y., Ruhl, I., Sannino, D., Maezato, Y., Blum, P., and Bini, E. (2011). CopR of *Sulfolobus solfataricus* represents a novel class of archaeal-specific copper-responsive activators of transcription. *Microbiology* 157, 2808–2817.

Waege, I., Schmid, G., Thumann, S., Thomm, M., and Hausner, W. (2010). Shuttle vector-based transformation system for *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 76, 3308–3313.

Wagner, M., Berkner, S., Ajon, M., Driessen, A. J., Lipps, G., and Albers, S. V. (2009). Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. *Biochem. Soc. Trans.* 37, 97–101.

Wang, G., Kennedy, S. P., Fasiludeen, S., Rensing, C., and Dassarma, S. (2004). Arsenic resistance in *Halobacterium* sp. strain NRC-1 examined by using an improved gene knockout system. *J. Bacteriol.* 186, 3187–3194.

Weidenbach, K., Gloer, J., Ehlers, C., Sandman, K., Reeve, J. N., and Schmitz, R. A. (2008). Deletion of the archaeal histone in *Methanosarcina mazei* Gö1 results in reduced growth and genomic transcription. *Mol. Microbiol.* 67, 662–671.

Welander, P. V., and Metcalf, W. W. (2008). Mutagenesis of the C1 oxidation pathway in *Methanosarcina barkeri*: new insights into the Mtr/Mer bypass pathway. *J. Bacteriol.* 190, 1928–1936.

Whitman, W. B., Tumbula, D. L., Yu, J. P., and Kim, W. (1997). Development of genetic approaches for the methane-producing archaebacterium *Methanococcus maripaludis*. *Biofactors* 6, 37–46.

Wilbanks, E. G., Larsen, D. J., Neches, R. Y., Yao, A. I., Wu, C. Y., Kjolby, R. A., and Facciotti, M. T. (2012). A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. *Nucleic Acids Res.* 40, e74.

Wirth, J. F., Snyder, J. C., Hochstein, R. A., Ortmann, A. C., Willits, D. A., Douglas, T., and Young, M. J. (2011). Development of a genetic system for the archaeal virus *Sulfolobus* turreted icosahedral virus (STIV). *Virology* 415, 6–11.

Worthington, P., Hoang, V., Perez-Pomares, F., and Blum, P. (2003). Targeted disruption of the α-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 185, 482–488.

Yokooji, Y., Tomita, H., Atomi, H., and Imanaka, T. (2009). Pantoate kinase and phosphopantothenate synthetase, two novel enzymes necessary for CoA biosynthesis in the Archaea. *J. Biol. Chem.* 284, 28137–28145.

Zhang, C., Guo, L., Deng, L., Wu, Y., Liang, Y., Huang, L., and She, Q. (2010). Revealing the essentiality of multiple archaeal *pcna* genes using a mutant propagation assay based on an improved knockout method. *Microbiology* 156, 3386–3397.

Zhang, C., and Whitaker, R. J. (2012). A broadly applicable gene knockout system for the thermoacidophilic Archaea *Sulfolobus islandicus* based on simvastatin selection. *Microbiology* 158, 1513–1522.

Zhang, J. K., Pritchett, M. A., Lampe, D. J., Robertson, H. M., and Metcalf, W. W. (2000). *In vivo* transposon mutagenesis of the methanogenic archaeon *Methanosarcina acetivorans* C2A using a modified version of the insect mariner-family transposable element Himar1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9665–9670.

Zhou, M., Xiang, H., Sun, C., and Tan, H. (2004). Construction of a novel shuttle vector based on an RCR-plasmid from a haloalkaliphilic archaeon and transformation into other haloarchaea. *Biotechnol. Lett.* 26, 1107–1113.

Zillig, W., Arnold, H. P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R., and Kristjansson, J. K. (1998). Genetic elements in the extremely thermophilic archaeon *Sulfolobus*. *Extremophiles* 2, 131–140.

Zillig, W., Prangishvilli, D., Schleper, C., Elferink, M., Holz, I., Albers, S., Janekovic, D., and Götz, D. (1996). Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. *FEMS Microbiol. Rev.* 18, 225–236.

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

Received: 13 May 2012; paper pending published: 05 June 2012; accepted: 01 September 2012; published online: 02 October 2012.

*Citation: Atomi H, Imanaka T and Fukui T (2012) Overview of the genetic tools in the Archaea. Front. Microbio. **3**:337. doi: 10.3389/fmicb.2012.00337*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

REVIEW ARTICLE
published: 08 June 2012
doi: 10.3389/fmicb.2012.00195

Genetics techniques for *Thermococcus kodakarensis*

Travis H. Hileman and Thomas J. Santangelo*

Department of Microbiology, Center for RNA Biology, Ohio State University, Columbus, OH, USA

Edited by:
Zvi Kelman, University of Maryland, USA

Reviewed by:
Yoshizumi Ishino, Kyushu University, Japan
James Chong, University of York, UK

****Correspondence:***
Thomas J. Santangelo, Department of Microbiology, Ohio State University, 290 Aronoff Building, 318 West 12th Avenue, Columbus, OH 43210, USA.
e-mail: santangelo.11@osu.edu

Thermococcus kodakarensis (*T. kodakarensis*) has emerged as a premier model system for studies of archaeal biochemistry, genetics, and hyperthermophily. This prominence is derived largely from the natural competence of *T. kodakarensis* and the comprehensive, rapid, and facile techniques available for manipulation of the *T. kodakarensis* genome. These genetic capacities are complemented by robust planktonic growth, simple selections, and screens, defined *in vitro* transcription and translation systems, replicative expression plasmids, *in vivo* reporter constructs, and an ever-expanding knowledge of the regulatory mechanisms underlying *T. kodakarensis* metabolism. Here we review the existing techniques for genetic and biochemical manipulation of *T. kodakarensis*. We also introduce a universal platform to generate the first comprehensive deletion and epitope/affinity tagged archaeal strain libraries.

Keywords: genetics, recombination, hyperthermophilic, archaea, *Thermococcales*, transcription

INTRODUCTION

Archaea are prevalent in many extreme environments but are also found in vast numbers in mesophilic marine and terrestrial environments where they contribute substantially to carbon, phosphorous, sulfur, and nitrogen cycles (Jarrell et al., 2011, references cited therein). Their importance in these global cycles is newly appreciated and demands a more complete understanding of archaeal encoded biochemistries and metabolic pathways. Advances have been limited due in large part to the difficulties associated with cultivating many archaea in the laboratory and the severe bottleneck resultant from the lack of genetic systems for most archaea. The dearth of genetic resources not only restricts our understanding of archaea in their natural environments, but also constrains the utility of this Domain for biomedical, biochemical, and industrial applications. This knowledge gap is perhaps most poignant for the hyperthermophilic archaea wherein the commercial utility of thermostable enzymes has been long recognized (Fujiwara et al., 1998; Hashimoto et al., 2001; Imanaka et al., 2001; Izumi et al., 2001; Hotta et al., 2002; Imanaka and Atomi, 2002; Cho et al., 2007; Griffiths et al., 2007; Blumer-Schuette et al., 2008; De Stefano et al., 2008; Bae et al., 2009; Kelly et al., 2009; Gaidamaviciute et al., 2010).

Within just the past decade, the recalcitrance of the archaea has dramatically declined due to advances in genetic techniques for select model organisms, and our understanding of archaeal physiology has resultantly exponentially increased (Tumbula and Whitman, 1999; Rother and Metcalf, 2005; Albers and Driessen, 2008; Santangelo et al., 2010; Leigh et al., 2011; Lipscomb et al., 2011). Essentially all barriers have been removed, and arguably the most complete set of genetic techniques has been developed for the globally abundant *Thermococcales* (Endoh et al., 2008; Santangelo and Reeve, 2010b; Bridger et al., 2011; Farkas et al., 2011; Takemasa et al., 2011). The model organism *Thermococcus kodakarensis* (*T. kodakarensis*; formally *Pyrococcus kodakaraensis* or *T. kodakaraensis*; Morikawa et al., 1994; Atomi et al., 2004b), for which the most complete suite of genetic techniques is available, is the subject of this review.

New avenues, based on advances in *T. kodakarensis* genetics, permit direct characterization of innumerable archaeal enzymes and their chemistries (Atomi et al., 2001, 2004a; Shiraki et al., 2003; Fukuda et al., 2004, 2008; Rashid et al., 2004; Sato et al., 2004, 2007; Imanaka et al., 2006; Murakami et al., 2006; Orita et al., 2006; Kanai et al., 2007, 2010, 2011; Danno et al., 2008; Fujiwara et al., 2008; Louvel et al., 2009; Yokooji et al., 2009; Borges et al., 2010; Kobori et al., 2010; Morimoto et al., 2010; Matsubara et al., 2011), provide industrially relevant alternative biofuel platforms(Kanai et al., 2005, 2011; Chou et al., 2008; Kim et al., 2010; Atomi et al., 2011; Santangelo et al., 2011; Bae et al., 2012; Davidova et al., 2012), unlock the largely untapped reservoir of archaeal encoded natural products (Atomi, 2005; Kim and Peeples, 2006; Littlechild, 2011; Matsumi et al., 2011; Sato and Atomi, 2011), and offer the opportunity to dissect eukaryotic-like information processing systems composed of minimal components (Yamamoto et al., 2003; Santangelo and Reeve, 2006, 2010a; Kanai et al., 2007; Santangelo et al., 2007, 2008a, 2009; Hirata et al., 2008; Dev et al., 2009; Yamaji et al., 2009; Fujikane et al., 2010; Li et al., 2010, 2011; Ishino et al., 2011; Nunoura et al., 2011; Pan et al., 2011; Santangelo and Artsimovitch, 2011).

Thermococcus kodakarensis is a marine, anaerobic, heterotrophic, hyperthermophilic (85°C), planktonic euryarchaeon, and thrives in medium supplemented with peptides, starch, or chitin, using S° or H^{+} as a terminal electron acceptor, generating H_2S or H_2, respectively. *T. kodakarensis* grows rapidly (doubling rate ∼40 min in rich media) to high cell densities and produces defined colonies on solid media, allowing overnight selections reliant on prototrophic markers (i.e., tryptophan, arginine, uracil, or agmatine) or antibiotic resistance (i.e., mevinolin, simvastatin) on defined or rich plates. Counter-selective procedures have been developed that facilitate repetitive modification of *T. kodakarensis*' small 2.08 Mb genome (52% GC; Fukui et al., 2005) that readily

incorporates exogenous DNA via homologous recombination with high efficiency (Sato et al., 2003). *T. kodakarensis* is naturally competent, requires no special techniques for transformation, accepts linear and circular DNAs, and via homologous recombination through short sequences of complementarity yields transformants at a frequency of ∼1 in 10^7 cells plated (∼100 transformants/10^9 cells/μg of transforming DNA). *T. kodakarensis* can support maintenance of autonomously replicating plasmids that provide ectopic expression platforms (Santangelo et al., 2008b), and a strong knowledge base of transcription regulation allows selective expression of native or introduced genes.

This multitude of selective markers and genetic techniques allow for essentially limitless genomic alterations including gene deletion, gene insertion, gene modification (allelic modifications as well as addition of affinity and epitope tags), promoter exchange, reporter gene expression, and any combination therein. Simple preservation techniques permit long-term frozen strain storage and the pace of strain construction is now limited only by molecular biology considerations. Here we briefly review existing genetic techniques available for *T. kodakarensis* and present a universally applicable platform and strategy developed to generate the first comprehensive archaeal strain collections.

GENETIC TECHNIQUES ALLOWING MODIFICATION OF THE *T. KODAKARENSIS* GENOME

Protocols for *T. kodakarensis* growth and genetic manipulation, as well as the underlying basis of each selection have recently been reviewed (Santangelo and Reeve, 2010b) and here we instead focus on the advantages and limitations of each technique and selective system. **Table 1** provides an overview of the genetic selections available and highlights the utility and limitations of each system. Each selective marker can be deployed as a single expression cassette, allowing the marker to be incorporated into and retained within the genome while coincidently modifying the genome in some manner, for example to generate a strain deleted for a specific gene (**Figure 1**). With the exception of the uracil-based marker (see below), use of any marker in isolation necessarily eliminates reuse of the same marker for any subsequent modification of the same genome. Individual markers have been employed consecutively to produce strains with multiple modifications, with each modification resulting in an additional marker retained in the genome of the final strain. More often, counter-selective pressures applied to recover the marker through a recombination based excision from the genome, allowing for markerless and repetitive modifications to be made to a single genome. Such counter-selective strategies are available for the uracil and 6-methyl purine (6MP) based markers, but only the uracil marker is functional for both positive- and counter-selection in isolation. The 6MP-based marker provides no useful positive selection, but can be paired with any of the positive selection cassettes to provide a two-gene cassette capable of positive selection into and counter-selective excision from the *T. kodakarensis* genome (**Figure 2**).

Regardless of the modification, all genetic manipulations are directed to specific loci through homology between the donor DNA and the *T. kodakarensis* chromosome, and recombination is efficient with ∼200 or more base pairs (bp) of sequence homology (**Figures 1** and **2**). Smaller regions of complementarity are also functional, but constraints of this manner are atypical of most transformations. Donor DNAs are most commonly circular DNAs that cannot autonomously replicate in *T. kodakarensis*, although linear DNA is also suitable for transformation. Expression cassettes containing a single selectable marker are most commonly flanked by sequences with homology to the locus of choice and transformants resultant from double-homologous recombination

Table 1 | Selective markers available for modification of *Thermococcus kodakarensis*.

Selectable marker	Gene(s)	Gene function	Strain (required genotype)	Advantages	Limitations/disadvantages	Reference
Uracil	TK2276	Orotidine-5′-phosphate decarboxylase	KU216 (Δ*pyrF*), KUW1 (Δ*pyrF*, Δ*trpE*)	Easily paired with 5-FOA-based counter-selection for markerless modifications	Uracil contamination yields high backgrounds; limited to minimal media; limited host range	Sato et al. (2003, 2005)
Tryptophan	TK0254	Large subunit of anthranilate synthase	KW128 (Δ*pyrF*; Δ*trpE::pyrF*)	Rigid selection requiring no media additions	Limited to minimal media; limited host range	Sato et al. (2005)
Arginine/ citrulline	PF0207, PF0208	Argininosuccinate synthase, argininosuccinate lyase	Any strain	No strain restrictions	Limited to minimal media; requires supplementation with citrulline	Santangelo and Reeve (2010b)
Agmatine	TK0149	Pyruvoyl-dependent arginine decarboxylase	TS559 (Δ*pyrF*; Δ*trpE::pyrF*, Δ*TK0664*, Δ*TK0149*)	Provides selective pressure in rich media	Limited host range	Santangelo and Reeve (2010b)
Simvastatin/ mevinolin	PF1848	HMG-CoA reductase	Any strain	Provides selective pressure in rich media; no strain restrictions	Spontaneous Sim/Mev resistance provides a high background	Matsumi et al. (2007), Santangelo et al. (2007)
6-methyl purine	TK0664	Hypoxanthine guanine phosphoribosyl-transferase	TS517 (ΔpyrF; ΔtrpE::pyrF, ΔTK0664)	Provides counter-selective pressure	Provides no positive selection; counter-selection requires minimal media	Santangelo et al. (2011)

into the *T. kodakarensis* genome are identified via diagnostic PCRs. Recombination of an entire circular donor DNA molecule into the genome via only a single homologous crossover does result in transformants that survive selective pressures, but under most circumstances such transformants are non-desirable and easily identified via diagnostic PCRs.

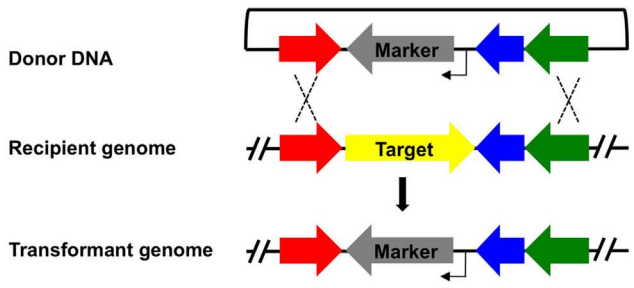


FIGURE 1 | Integration of a selection cassette into the *Thermococcus kodakarensis* genome provides a mechanism for genomic modification. A general scheme for integration of a selectable marker cassette (gene providing positive selection, gray; promoter, bent arrow) into the genome of a recipient strain. The resultant genome of the transformant results from two homologous recombination events (dashed lines) between flanking regions contained in both the donor DNA and targeted genomic locus. For simplicity, a single gene replacement of a target locus (yellow) by the selectable marker is shown, although different genomic alterations including allelic exchanges, promoter alterations, and additions of epitopes can similarly be introduced using the same technique. Each marker in **Table 1**, with the exception of 6MP, can be employed to generate strains that retain the marker at or near the modified locus.

PROTOTROPHIC SELECTIONS

ARGININE/CITRULLINE-BASED SELECTION

In contrast to several members of the *Thermococcales*, *T. kodakarensis* is an arginine auxotroph (Fukui et al., 2005). Introduction and expression of two genes from *P. furiosus*, PF0207 and PF0208 that encode argininosuccinate synthase and argininosuccinate lyase respectively, to the *T. kodakarensis* genome, or on a replicative plasmid (see below), were shown to provide any *T. kodakarensis* strain the capacity to combine citrulline and aspartate to form arginine, and thus provide arginine prototrophy. *T. kodakarensis* makes abundant aspartate, but cultures must be supplemented with citrulline for introduction of the exogenous *P. furiosus* genes to confer arginine prototrophy; citrulline supplementation in the absence of PF0207/PF0208 expression is insufficient to impart arginine prototrophy.

Arginine/citrulline-based selections are rigid, with no spontaneous arginine-prototrophic colony formation. Initial selection is limited to growth on minimal media (19 amino acids plus citrulline), but once confirmed, strains can be passage and plated on rich media without concern for retention of the marker; spontaneous excision has not been reported for any marker employed for *T. kodakarensis* genetic manipulations. The greatest advantage of arginine/citrulline-based selections is the lack of any genotypic strain requirements. This flexibility is unmatched by the other available prototrophic markers, and although the arginine/citrulline-based selection has only recently been developed, it is compatible with all other selections. The only potential disadvantage is the necessity to supplement defined media with citrulline, however addition is not of significant concern based on

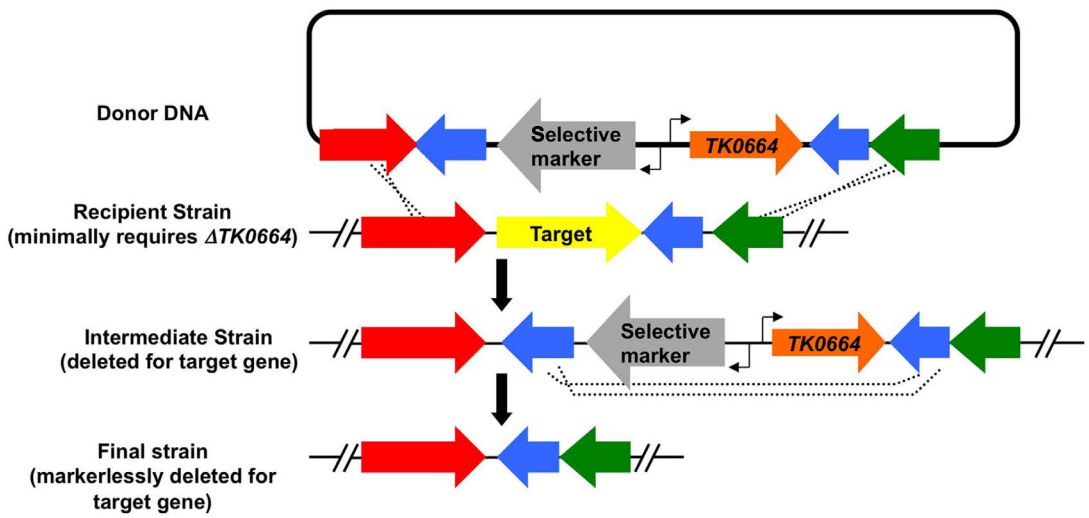


FIGURE 2 | Markerless-modification of the *Thermococcus kodakarensis* genome via sequential positive selection and subsequent counter-selection. Donor DNA, a hypothetical target locus in the recipient genome, the resultant intermediate, and final genomes are shown. The scheme diagrammed generates an intermediate strain wherein both the selectable (gray) and counter-selectable marker (TK0664; orange) are flanked by a direct repeat (cyan), and the intermediate strain is deleted for the target gene. Alternative donor DNAs can result in intermediate strains that retain the target gene, necessitating excision of the selectable markers in addition to the target gene to via recombination between direct repeats to generate the desired final markerless-deletion strain. The same protocol can be used with the uracil (*pyrF*) cassette that will serve as both the positive and counter-selective marker. Any combination of a positive selection cassette (**Table 1**) and TK0664-based 6MP counter-selection can be employed in a two-gene mechanism to introduce markerless modifications to the *T. kodakarensis* genome.

cost, availability, or stability, nor is addition of citrulline necessary for growth in rich media.

TRYPTOPHAN-BASED SELECTION

Only specific strains of *T. kodakarensis* are amenable to tryptophan-based selection, although these strains are widely available and selections based on tryptophan have been employed in the most diverse *T. kodakarensis* strain constructions (Sato et al., 2003, 2005; Imanaka et al., 2006). Most reported tryptophan auxotrophic strains are non-reverting as the result of an insertion in TK0254 (*trpE*), encoding the large subunit of anthranilate synthase. Tryptophan selection, like arginine/citrulline-based selections is rigid, and no spontaneous tryptophan-prototrophic colonies have been recovered in the absence of donor DNA. Initial selections are still limited to defined media, but in contrast to arginine/citrulline-based selections, the tryptophan-based selection does not require medium supplementation. Integration of the *trpE* cassette is similarly stable and does not require selective pressure once established, allowing growth of confirmed strains in rich media without concern for loss of the marker.

AGMATINE-BASED SELECTION

Polyamines serve as essential counter-ions in all Domains and are typically derived from the precursor agmatine (Morimoto et al., 2010). Agmatine is decarboxylated arginine, and although *T. kodakarensis* is an arginine auxotroph, it does encode a pyruvoyl-dependent arginine decarboxylase (TK0149; Fukuda et al., 2008). Deletion or inactivation of TK0149 results in the agmatine-dependent growth, allowing selections in specific strains wherein TK0149 was previously deleted (Santangelo and Reeve, 2010b). The major advantage of agmatine-based selections, despite their limited host range, is that agmatine auxotrophy is lethal even in rich media, thus initial selections to agmatine prototrophy can be performed on rich media. This pronounced difference from uracil-, arginine/citrulline-, or tryptophan-based prototrophic selections provides a means for selection of transformants overnight in contrast to the 3–4 days typically required for colony formation on defined media.

Agmatine-based selections are as rigid as those for tryptophan or arginine/citrulline, but do require strains deleted for TK0149 as well as supplementation of these strains with agmatine prior to transformation. These concerns are minor as agmatine is inexpensive and widely available. The selective pressure provided by the agmatine-marker is particularly useful for retention of replicative plasmids in rich media (see below).

URACIL-BASED SELECTION AND 5-FOA-BASED COUNTER-SELECTION

The first reported selections (Sato et al., 2003, 2005) that were established into useful genetic techniques were based on spontaneous resistance to 5-fluoro-orotic acid (5-FOA^R), a cytotoxic pyrimidine analog. 5-FOA^R strains were isolated and shown to contain mutations disrupting the sequence of, or limiting the expression of TK2138 or TK2276 (*pyrE* or *pyrF*, respectively), consistent with the conserved roles of *pyrE* and *pyrF* in pyrimidine metabolism (Sato et al., 2003). Strains containing non-reverting mutations of *pyrF* serve as the host for donor DNAs that introduce targeted gene disruptions at remote loci while restoring *pyrF* expression, and thus uracil prototrophy, from this same location.

Uracil-dependent techniques are limited to defined media, and while uracil auxotrophy/prototrophy is still used to isolate transformants, many media components contain trace or greater amounts of uracil that routinely complicate isolation of mutants using this method. The technology is retained in large part due to its simplicity, and perhaps more importantly, the ability to counter-select against *pyrF* function, allowing repetitive manipulation of the genome via an initial selection and sequent counter-selection (pop-in/pop-out) mechanism (**Figure 2**). By integrating a *pyrF* cassette while at the same time introducing duplicate sequences flanking *pyrF*, researchers can develop so-called intermediate strains with the desired phenotype. Exposure of this intermediate strain to 5-FOA most commonly results in 5-FOA^R colonies resultant from recombination between the direct repeats flanking *pyrF*, and thus excision of the marker from the chromosome while the desired modification is retained in the genome. The excision event produces restored uracil auxotrophy, allowing reuse of the uracil marker to generate a second, third, fourth, etc., genomic alteration. Proper planning allows for exquisite precision during initial integration and subsequent excision, permitting markerless deletions, or alternative genomic modifications.

ANTIBIOTIC-BASED SELECTIONS

Few antibiotics/antimicrobials have demonstrated efficacy against archaea, and of those, only a few are stable enough at high temperature to be employed for use with hyperthermophiles. Antibiotic selections remain highly desirable as they are typically robust, effective on many media, and provide continued selective pressure when transformants are transferred to liquid media. A class of statins, developed to inhibit cholesterol biosynthesis and typified by simvastatin (Sim) and mevinolin (Mev), is effective in the low μM range at limiting *T. kodakarensis* growth (Matsumi et al., 2007; Santangelo et al., 2008a). Archaeal strains spontaneously resistant to Sim/Mev (Sim/Mev^R) were shown to have mutations that mapped to the locus encoding 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase), and these mutations generally lead to overexpression of HMG-CoA reductase (Lam and Doolittle, 1992). It was hypothesized and subsequently shown that increased expression of, rather than modifications to, HMG-CoA reductase provided a level of resistance to Sim/Mev. Selections based on the introduction of an additional and highly expressed copy of HMG-CoA reductase, provided to generally increase the *in vivo* levels of HMG-CoA reductase soon followed, with the HMG-CoA reductase from *P. furiosus* employed to limit unwanted recombination between the native *T. kodakarensis* HMG-CoA reductase and the introduced copy of HMG-CoA reductase.

Selections based on Sim/Mev^R are advantageous in that they are not limited to a select genotype and provide selective pressure in all media allowing rapid selection on rich media. However, Sim/Mev-based selections are disreputably weak and are cost prohibitive on a large scale. Spontaneous Sim/Mev^R colonies are readily recovered, and this incidence presents real and significant challenges when working with slow growing strains that may be overwhelmed in liquid culture by spontaneous Sim/Mev^R cells. Plasmid maintenance based on Sim/Mev^R is possible, with the noted caveats of spontaneous resistance.

TWO-GENE SELECTION/COUNTER-SELECTION MODIFICATION OF THE *T. KODAKARENSIS* CHROMOSOME

The uracil-based selection and 5-FOA-based counter-selection are advantageous for generating strains with markerless modifications to the *T. kodakarensis* chromosome and for repeated manipulation of the same chromosome at multiple loci, but is hamstrung by the lack of rigidity in generating uracil free media. The uracil contamination of most commercially available medium components is sufficient to support weak growth of Δ*pyrF* strains over the course of several days.

To circumvent these concerns, a more rigid selection/counter-selection procedure was developed based on the sensitivity of all *T. kodakarensis* strains to the cytotoxic compound 6-methyl purine (6MP). *T. kodakarensis* encodes a complete purine biosynthetic pathway, but like many organisms also encodes a purine-scavenging pathway to recycle purines and nucleotides from the environment. 6MP is a nucleotide base analog that, once imported and converted to a modified nucleotide- or deoxynucleotide-triphosphate and incorporated into macromolecules, overwhelms DNA repair pathways and inhibits the information processing machinery. Spontaneous $6MP^R$ *T. kodakarensis* strains were isolated, and the mutations resulting in $6MP^R$ were mapped. All mutations were at the TK0664 locus, and it was subsequently shown that inactivation or deletion of TK0664, encoding a hypoxanthine guanine phosphoribosyltransferase, conferred $6MP^R$.

A 6MP-based counter-selection against TK0664 was established, but required pairing with an initial prototrophic or antibiotic-based positive selection to generate initial transformants. Combining any of the selection cassettes with a cassette encoding TK0664 provides a two-gene based selection/counter-selection procedure (**Figure 2**) that functionally mimics the single *pyrF*-based uracil prototrophy selection, 5-FOA counter-selection protocol. The limitations of this two-gene system include the necessity to introduce two expression cassettes during initial strain construction, the reasonable expense of 6MP, the very limited host range, and the requirement of a different host strain for each gene pair. These complications are largely outweighed by the efficiency of the system compared to the uracil-based selection/counter-selection, but are still laborious.

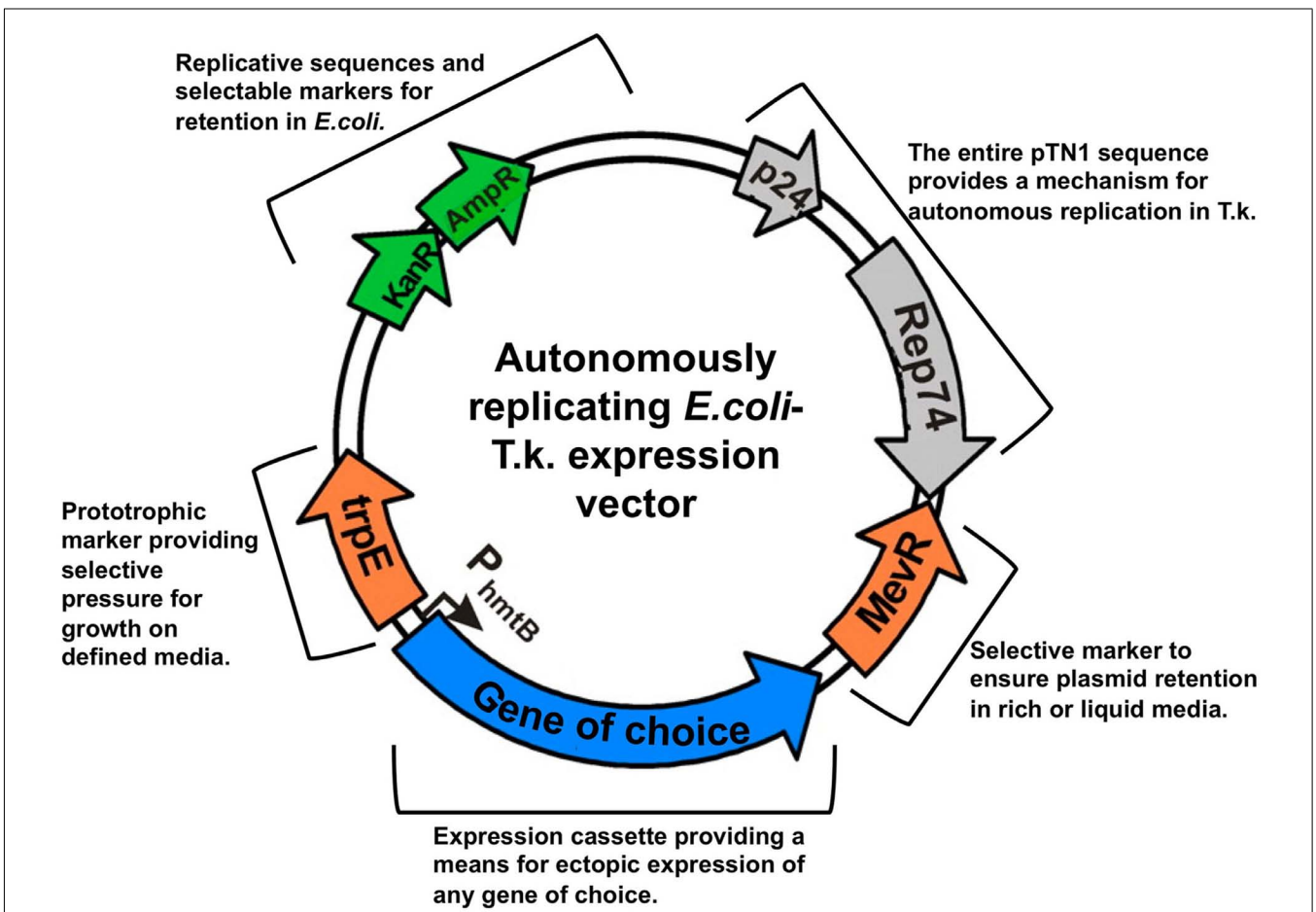


FIGURE 3 | Replicative vectors provide platforms for exogenous and ectopic expression in *Thermococcus kodakarensis*. T.k-*E. coli* shuttle vectors are the result of merging the entire *Thermococcus nautilus*-derived pTN1 sequence with a common *E. coli* plasmid (pCR2.1-Topo; Invitrogen). The marker additions provide the means to select *T. kodakarensis* transformants on any media, and the expression cassette provides a mechanism to ectopically express any gene of choice in *T. kodakarensis*.

REPLICATIVE EXPRESSION VECTORS

Thermococcus kodakarensis does not naturally contain any extrachromosomal elements, but *T. kodakarensis* does support replication of plasmids derived from other *Thermococcales*. *T. nautilus* was shown to contain three distinct plasmids (Soler et al., 2007), and the smallest of these, pTN1, was converted into an *E. coli*-*T. kodakarensis* shuttle vector (**Figure 3**; Santangelo et al., 2008b). Variants of this vector carrying most combinations of selective markers are available, and necessarily retain at least one marker (Sim/Mev or agmatine) that exerts selective pressure in rich media.

When these vectors are used to express essential genes, the chromosomal locus for the same gene can be modified or deleted, and viability may then be dependent on retention of the plasmid. Replicative vectors also permit protein expression under native conditions, at controlled levels, and allow *T. kodakarensis* to serve as a host of exogenous gene expression (Santangelo et al., 2008b, 2010; Takemasa et al., 2011). The latter may impart novel capacities to *T. kodakarensis*, or fulfill industrial needs for thermostable protein expression. These shuttle vectors can be easily modified in *E. coli* to introduce sequences encoding epitopes and affinity tags that facilitate downstream purification of the encoded products. Finally, collections of plasmids, for example, a plasmid library containing randomly mutagenized variants of a gene, can be quickly produced in *E. coli* and the efficiency of *T. kodakarensis* transformation supports introduction of this library for use in screens and selections.

A UNIVERSAL PLATFORM FOR CONSTRUCTION OF COMPREHENSIVE STRAIN COLLECTIONS

The existing genetic techniques for *T. kodakarensis* were recently combined into a universally applicable strategy to generate the first comprehensive strain libraries for any archaeon. Two *T. kodakarensis* strain libraries, one wherein each non-essential gene is individually deleted, and a second wherein each protein-encoding gene is modified to encode an epitope and affinity tagged isoform, are under construction. We outline some details of the streamlined procedure for construction of the library vectors and associated strains (**Figures 4–6**). This platform can be adapted to generate any *T. kodakarensis* strain of choice and provides an alternative to the often laborious molecular biology manipulations that are required to generate standard vectors for integration into, and subsequent excision from, the *T. kodakarensis* genome. The use of single crossover integrations provides for the possibility of excision of the plasmid to restore the TS559 genome without modification,

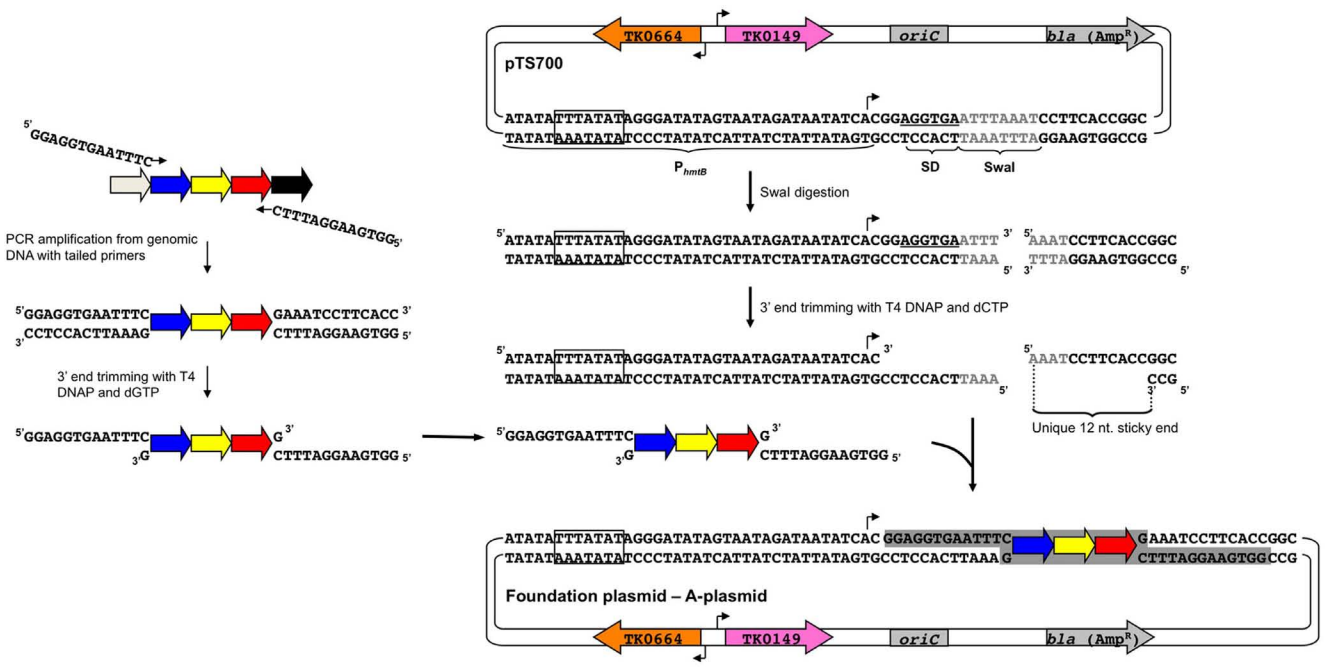


FIGURE 4 | Construction of foundation plasmids by a ligation-independent technique. (Left) Each target gene as well as ∼500–700 bp of both upstream and downstream adjacent DNA is amplified from *T. kodakarensis* chromosomal DNA with a pair of primers that introduce unique 13 bp terminal extensions to the amplicon. Incubation of the purified amplicon with T4 DNA polymerase (DNAP) and only dGTP leads to a 3′-5′ exonuclease recession of the 3′ ends to produce non-complementary 12-nt sticky ends. (Right) pTS700 is linearized through a unique *Swa*I site, producing 3′ ends that are similarly recessed with T4 DNAP and dCTP to yield sticky ends complementary to the sticky ends of each amplicon permitting directional cloning of the amplicon in a ligase-independent reaction. Amplicons for each *T. kodakarensis* gene, regardless of gene orientation, gene length, or sequence can be cloned in an identical manner to yield foundation plasmids (A-plasmids). The TATA-box of P_{hmtB} is boxed and the site of transcription initiation marked with a bent arrow; the Shine–Dalgarno sequence (SD) is underlined; the *Swa*I recognition sequence is shown in gray; the selectable and counter-selectable expression cassettes, TK0149 and TK0664, are shown in pink and orange, respectively; the plasmid origin of replication (*oriC*) and *E. coli* resistance cassette (*bla*) are shown in gray. The color scheme is conserved in **Figures 4–7**.

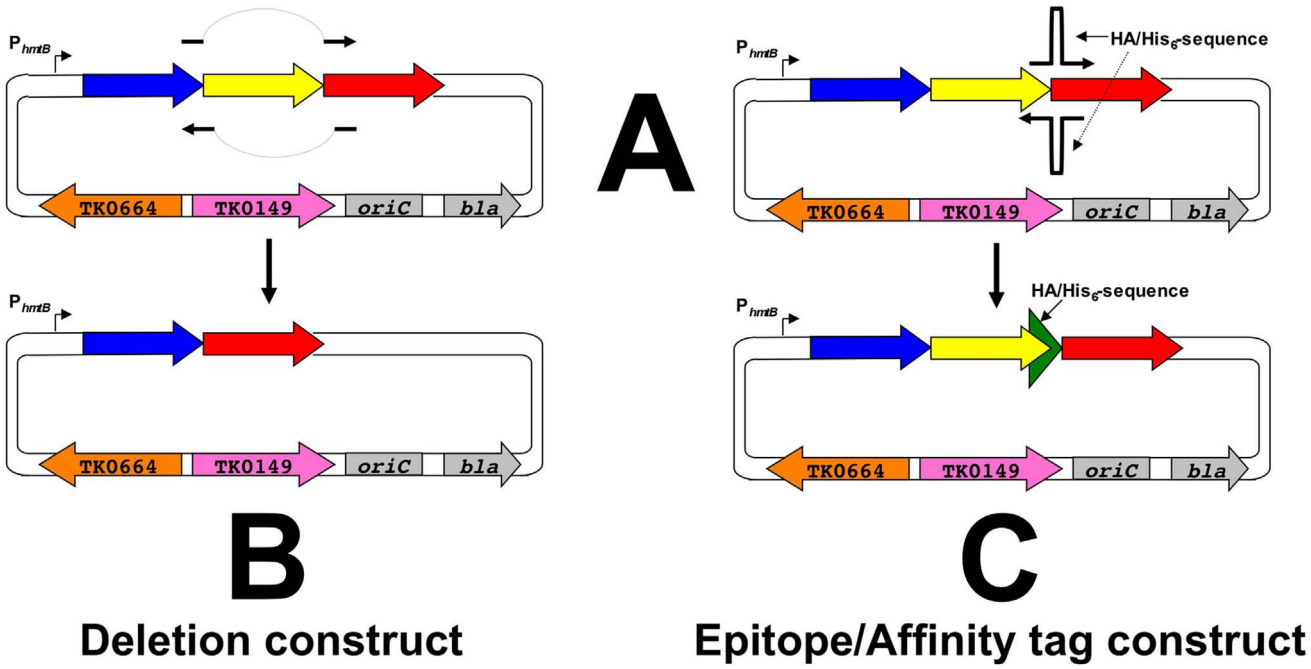


FIGURE 5 | Construction of deletion- and affinity-plasmids from the foundation plasmids. A-plasmids serve as templates for PCR (QuikChange; Agilent) wherein essentially the entire plasmid is replicated to generate B,C-plasmids respectively. B-plasmids result from primer pairs that are complementary, both upstream and downstream, of the target gene (yellow), whereas C-plasmids result from primer pairs that introduce 45 additional nucleotides, encoding both a His_6 and HA epitope tag, in-frame and immediately prior to the normal stop codon of the target gene. A simplified diagram of the plasmids is used for clarity only and each plasmid retains all the components highlighted in **Figure 4**. The larger green arrow depicts the sequence encoding the His_6-HA tag.

as would be necessary when targeting an essential gene. Continued recovery of only the restored TS559 genome provides a statistical measure that can be applied to determine essentiality of individual genes.

The collections are all based on a single *T. kodakarensis* strain and vector, but the technologies employed are nearly identical to the platforms described above. pTS700 provides the vehicle for introduction of donor DNA complementary to the *T. kodakarensis* genome and carries the selectable and counter-selectable markers facilitating the most rapid and rigid integration into and excision from the *T. kodakarensis* genome, but will not autonomously replicate in *T. kodakarensis* (**Figure 4**). The host strain, TS559, has the complementary genotype (Δ*pyrF*; Δ*trpE::pyrF*, Δ*TK0664*, Δ*TK0149*) for the selectable markers carried on pTS700. The strain libraries so constructed are markerless, and as such the resultant strains are isogenic while importantly permitting continued modification of the strains using the library of vectors so established. The strains are compatible with every replicative vector, further increasing their utility for complex genetic manipulations and their use in selections and screens to isolate strain variants of choice. The use of a single platform provides for economical and rapid construction of the necessary vectors to generate ∼4,600 unique strains.

Construction begins with amplification of a target gene (shown in yellow, **Figure 4**) with ∼500–700 bp of flanking DNA. This amplicon is cloned via a ligation-independent mechanism into pTS700 generating the initial vector termed an "A" plasmid. The ligation-independent mechanism permits all amplicons to be cloned using the same procedure, thus simplifying construction of ∼2,300 A-plasmids, one for each *T. kodakarensis* protein-encoding gene. Each A-plasmid serves as a foundation plasmid, from which additional vectors of choice can be generated. Two plasmid variants, termed "B-" and "C-plasmids," are typically generated from the A-plasmid that respectively provide the donor DNAs to generate the deletion and epitope/affinity tagged *T. kodakarensis* strains for each protein-encoding gene. A-plasmids can undergo additional or combinatorial modifications for construction of more unique strains (i.e., "Q-plasmids" contain modified promoters; "M-plasmids" contain allelic modifications), but for this review we concentrate on construction of the deletion and epitope/affinity tagged libraries.

B-plasmids are constructed in a single-step PCR-based procedure wherein the original target gene (yellow) is deleted from the plasmid, while leaving the flanking DNA that will target integration of the entire B-plasmid to the TS559 genome (**Figure 5**). Two separate initial integration events are possible for incorporation of the entire B-plasmid to the genome, and each generates an intermediate strain that, when 6MP-based counter-selective pressure is applied, can undergo an internal recombination event to either restore the original genome or produce a strain containing

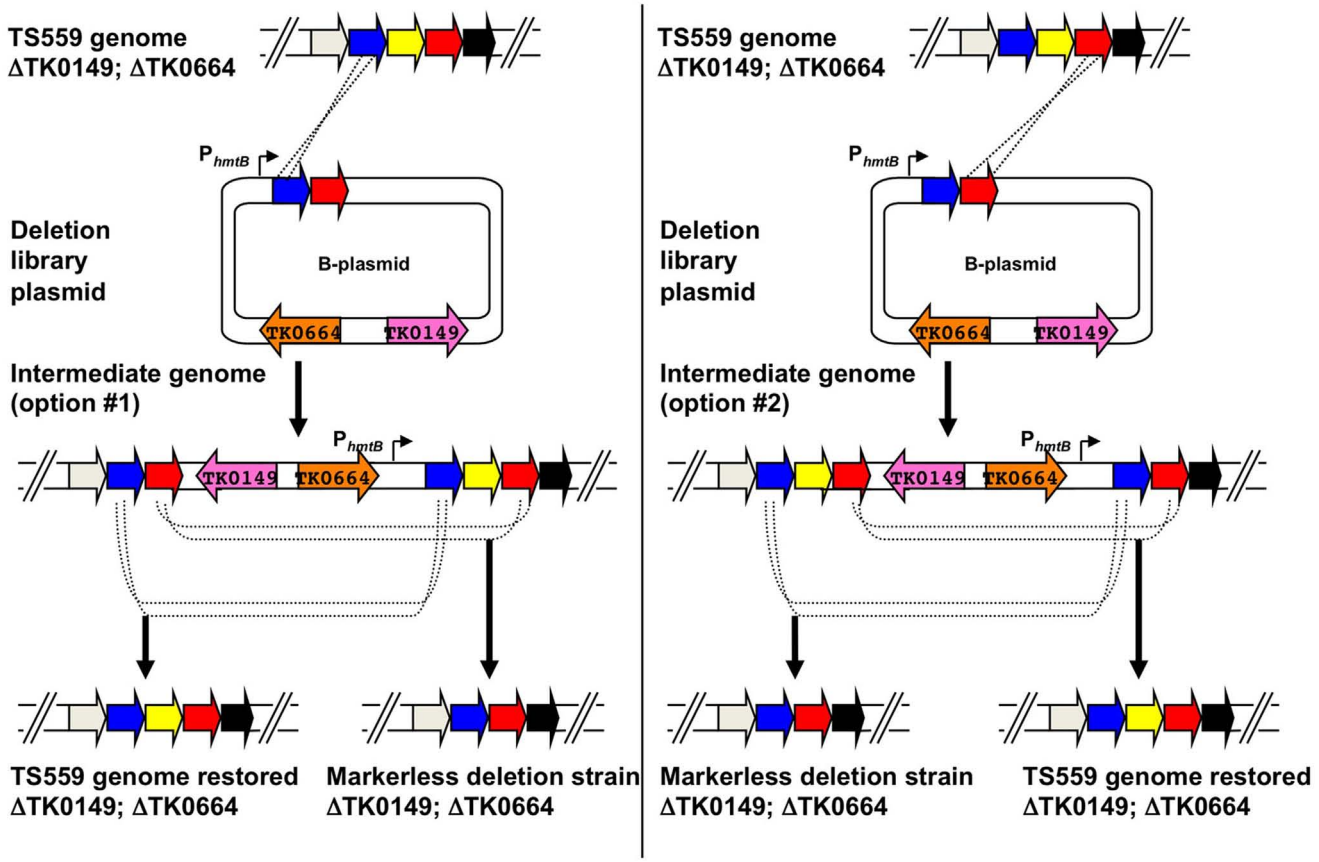


FIGURE 6 | Use of deletion plasmids (B-plasmids) to generate markerless knockouts in the genome of *Thermococcus kodakarensis*. A hypothetical region of the genome of *T. kodakarensis* strain TS559 is shown at top, with the left and right panels depicting the two possible integration events yielding agmatine-prototrophic intermediate strains from the diagramed B-plasmid. Both intermediate strains #1 and #2 contain direct repeats flanking the target locus, and dependent on recombination responsible for excision upon counter-selection with 6MP, either the original TS559 genome can be restored or the desired deletion genome generated.

a genome with the desired, targeted deletion (**Figure 6**). The ratio of WT versus desired deletion recovered $6MP^R$ strains provides a statistical measure identifying essential genes; failure to recover a strain with the desired deletion from a large number of $6MP^R$ final strains implies essentiality.

C-plasmids are similar constructed from A-plasmids via a PCR-based procedure that introduces the 45 bp encoding the His_6 and 9-amino acid hemagglutinin (HA; YPYDVPDYA) tag to the target gene, whilst retaining the entire original amplicon of the foundation plasmid (**Figures 5** and **7**). Both the His_6 and HA-tags have been used successful to facilitate protein identification and ease protein purification with minimal background. C-plasmids can similarly integrate and excise from the genome to generate the desired tagged strain or restore the TS559 genome.

The donor plasmids also contain the constitutively expressed P_{hmtB} promoter immediately upstream of the site wherein the initial amplicon is cloned to generate the A-plasmid. P_{hmtB} provides an expression platform for genes that may become separated from their promoter, as would be common for an integration event that disrupts an operon.

CONCLUSION

The genetic techniques and tools for *T. kodakarensis* provide a potent arsenal of mechanisms to precisely and repetitively modify the chromosome, as well as ectopically introduce and express exogenous or modified genes *in vivo*. The selectable markers and host strains developed to date provide the means for essentially any modification in any strain background. Although not the subject of this review, the developed reporter constructs (Santangelo et al., 2008a, 2010) and *in vitro* transcription (Santangelo et al., 2007) and translation systems (Endoh et al., 2006, 2007; Yamaji et al., 2009) using *T. kodakarensis* components provide complementary *in vivo* and *in vitro* platforms to dissect regulation of transcription and translation regulatory mechanisms.

The possibility of generating comprehensive strain libraries is now a reality, and a common platform is in place to speed construction of such libraries and provide the community the first comprehensive strain collections with the most isogenic background possible. Furthering their value, the plasmid and strain libraries so constructed retain value for additional strain modification, including strains containing multiple deletions or multiple genes with tagged isoforms. More specific and imaginative variants

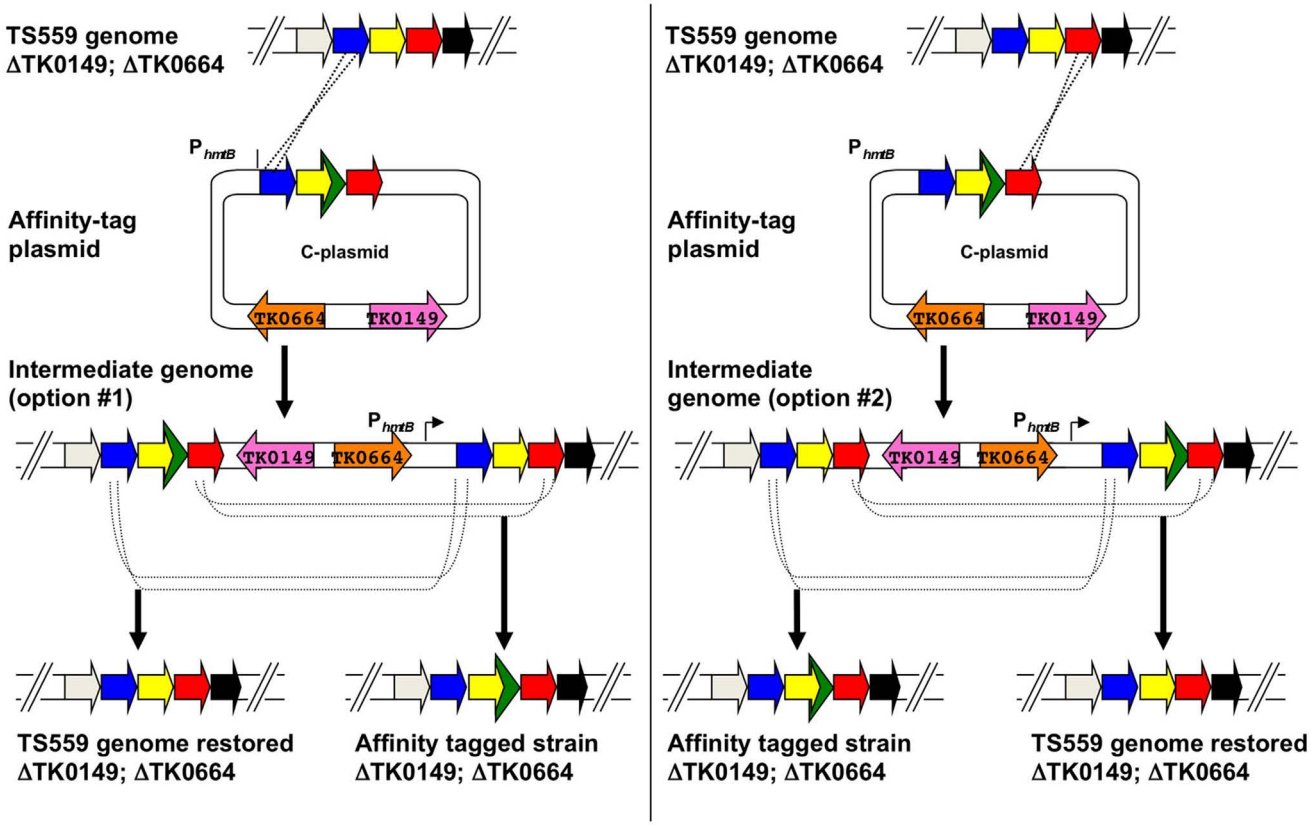


FIGURE 7 | Use of epitope/affinity tag plasmids (C-plasmids) to generate markerlessly tagged-strains of *Thermococcus kodakarensis*. A hypothetical region of the genome of *T. kodakarensis* strain TS559 is shown at top, with the left and right panels depicting the two possible integration events yielding agmatine-prototrophic intermediate strains from the diagramed C-plasmid. Both intermediate strains #1 and #2 contain direct repeats flanking the target locus, and dependent on recombination responsible for excision upon counter-selection with 6MP, either the original TS559 genome can be restored or the desired tag- and epitope-containing genome generated.

are quickly feasible through minor modification of the foundation plasmids for each protein-encoding gene; for example, strains with promoter alterations and allelic modifications can now be combined with select deletions, and these modifications can be made in backgrounds wherein other protein-encoding genes are tagged or modified. This depth of genetic flexibility will permit entire pathways to be introduced into, or deleted from, the *T. kodakarensis* genome, and provides a mechanism to probe biochemical pathways using complementation and screens and selections to isolate mutants with desired phenotypes.

The toolkit for *T. kodakarensis* genetics is impressive, but still incomplete. Specifically, transduction would further speed strain construction, but is not yet possible. The isolation of the first virus capable of replication in *T. kodakarensis* (Gorlas et al., 2012) may facilitate development of a transduction method. Replicative expression vectors currently all share the same origin (pTN1-based; Soler et al., 2007), and the development of additional shuttle vectors with complementary origins of replication derived from plasmids newly identified in *Thermococcal* species (Gonnet et al., 2011; Soler et al., 2011) would permit addition of several plasmids to the cell at once, or allow plasmid shuffle experiments as may be required when analyzing the function of essential genes. Constructs and promoters allowing easily regulated *in vivo* expression would be a welcome addition, providing a simply mechanism to turn gene expression off and on to monitor the effects of loss or gain of function on phenotype under dynamic conditions. Transformation efficiency should be improved, and insights may be garnered from the recently developed genetic system for the closely related organism *P. furiosus* wherein transformation efficiency is several orders of magnitude greater than that of *T. kodakarensis* (Lipscomb et al., 2011). These endeavors, as well as the ability to currently probe the form and function of each *T. kodakarensis* encoded gene, will undoubtedly continue to add to our knowledge and understanding of archaeal physiology, provide the basis for new technologies and production of commercial-relevant products, and underlie the basic research exploiting the component-simplified information processing machinery shared by archaea and eukaryotes.

ACKNOWLEDGMENTS

This work is supported by grants GM098176 and GM100329 from the National Institutes of Health to T.J.S.

REFERENCES

Albers, S. V., and Driessen, A. J. (2008). Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. *Archaea* 2, 145–149.

Atomi, H. (2005). Recent progress towards the application of hyperthermophiles and their enzymes. *Curr. Opin. Chem. Biol.* 9, 166–173.

Atomi, H., Ezaki, S., and Imanaka, T. (2001). Ribulose-1,5-bisphosphate carboxylase/oxygenase from *Thermococcus kodakarensis* KOD1. *Meth. Enzymol.* 331, 353–365.

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

Atomi, H., Fukui, T., Kanai, T., Morikawa, M., and Imanaka, T. (2004b). Description of *Thermococcus kodakarensis* sp nov, a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp KOD1. *Archaea* 1, 263–267.

Atomi, H., Sato, T., and Kanai, T. (2011). Application of hyperthermophiles and their enzymes. *Curr. Opin. Biotechnol.* 22, 618–626.

Bae, H., Kim, K. P., Lee, J. I., Song, J. G., Kil, E. J., Kim, J. S., and Kwon, S. T. (2009). Characterization of DNA polymerase from the hyperthermophilic archaeon *Thermococcus marinus* and its application to PCR. *Extremophiles* 13, 657–667.

Bae, S. S., Kim, T. W., Lee, H. S., Kwon, K. K., Kim, Y. J., Kim, M. S., Lee, J. H., and Kang, S. G. (2012). H2 production from CO, formate or starch using the hyperthermophilic archaeon, *Thermococcus* onnurineus. *Biotechnol. Lett.* 34, 75–79.

Blumer-Schuette, S. E., Kataeva, I., Westpheling, J., Adams, M. W., and Kelly, R. M. (2008). Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr. Opin. Biotechnol.* 19, 210–217.

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

Bridger, S. L., Clarkson, S. M., Stirrett, K., DeBarry, M. B., Lipscomb, G. L., Schut, G. J., Westpheling, J., Scott, R. A., and Adams, M. W. (2011). Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus*. *J. Bacteriol.* 193, 6498–6504.

Cho, Y., Lee, H. S., Kim, Y. J., Kang, S. G., Kim, S. J., and Lee, J. H. (2007). Characterization of a dUTPase from the hyperthermophilic archaeon *Thermococcus* onnurineus NA1 and its application in polymerase chain reaction amplification. *Mar. Biotechnol.* 9, 450–458.

Chou, C. J., Jenney, F. E. Jr., Adams, M. W., and Kelly, R. M. (2008). Hydrogenesis in hyperthermophilic microorganisms: implications for biofuels. *Metab. Eng.* 10, 394–404.

Danno, A., Fukuda, W., Yoshida, M., Aki, R., Tanaka, T., Kanai, T., Imanaka, T., and Fujiwara, S. (2008). Expression profiles and physiological roles of two types of prefoldins from the hyperthermophilic archaeon *Thermococcus kodakarensis*. *J. Mol. Biol.* 382, 298–311.

Davidova, I. A., Duncan, K. E., Perez-Ibarra, B. M., and Suflita, J. M. (2012). Involvement of thermophilic archaea in the biocorrosion of oil pipelines. *Environ. Microbiol.* doi: 10.1111/j.1462-2920.2012.02721.x. [Epub ahead of print].

De Stefano, L., Vitale, A., Rea, I., Staiano, M., Rotiroti, L., Labella, T., Rendina, I., Aurilia, V., Rossi, M., and D'Auria, S. (2008). Enzymes and proteins from extremophiles as hyperstable probes in nanotechnology: the use of D-trehalose/D-maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis* for sugars monitoring. *Extremophiles* 12, 69–73.

Dev, K., Santangelo, T. J., Rothenburg, S., Neculai, D., Dey, M., Sicheri, F., Dever, T. E., Reeve, J. N., and Hinnebusch, A. G. (2009). Archaeal aIF2B interacts with eukaryotic translation initiation factors eIF2alpha and eIF2Balpha: implications for aIF2B function and eIF2B regulation. *J. Mol. Biol.* 392, 701–722.

Endoh, T., Kanai, T., and Imanaka, T. (2007). A highly productive system for cell-free protein synthesis using a lysate of the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Appl. Microbiol. Biotechnol.* 74, 1153–1161.

Endoh, T., Kanai, T., and Imanaka, T. (2008). Effective approaches for the production of heterologous proteins using the *Thermococcus kodakarensis*-based translation system. *J. Biotechnol.* 133, 177–182.

Endoh, T., Kanai, T., Sato, Y. T., Liu, D. V., Yoshikawa, K., Atomi, H., and Imanaka, T. (2006). Cell-free protein synthesis at high temperatures using the lysate of a hyperthermophile. *J. Biotechnol.* 126, 186–195.

Farkas, J., Chung, D., DeBarry, M., Adams, M. W., and Westpheling, J. (2011). Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon *Pyrococcus furiosus* needed for construction of a stable replicating shuttle vector. *Appl. Environ. Microbiol.* 77, 6343–6349.

Fujikane, R., Ishino, S., Ishino, Y., and Forterre, P. (2010). Genetic analysis of DNA repair in the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Genet. Syst.* 85, 243–257.

Fujiwara, S., Aki, R., Yoshida, M., Higashibata, H., Imanaka, T., and Fukuda, W. (2008). Expression profiles and physiological roles of two types of molecular chaperonins from the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Appl. Environ. Microbiol.* 74, 7306–7312.

Fujiwara, S., Takagi, M., and Imanaka, T. (1998). Archaeon *Pyrococcus kodakaraensis* KOD1: application and evolution. *Biotechnol. Annu. Rev.* 4, 259–284.

Fukuda, W., Fukui, T., Atomi, H., and Imanaka, T. (2004). First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *J. Bacteriol.* 186, 4620–4627.

Fukuda, W., Morimoto, N., Imanaka, T., and Fujiwara, S. (2008). Agmatine is essential for the cell growth of *Thermococcus kodakarensis*. *FEMS Microbiol. Lett.* 287, 113–120.

Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S., and Imanaka, T. (2005). Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* 15, 352–363.

Gaidamaviciute, E., Tauraite, D., Gagilas, J., and Lagunavicius, A. (2010). Site-directed chemical modification of archaeal *Thermococcus litoralis* Sh1B DNA polymerase: acquired ability to read through template-strand uracils. *Biochim. Biophys. Acta* 1804, 1385–1393.

Gonnet, M., Erauso, G., Prieur, D., and Le Romancer, M. (2011). pAMT11, a novel plasmid isolated from a *Thermococcus* sp strain closely related to the virus-like integrated element TKV1 of the *Thermococcus kodakarensis* genome. *Res. Microbiol.* 162, 132–143.

Gorlas, A., Koonin, E. V., Bienvenu, N., Prieur, D., and Geslin, C. (2012). TPV1, the first virus isolated from the hyperthermophilic genus *Thermococcus*. *Environ. Microbiol.* 14, 503–516.

Griffiths, K., Nayak, S., Park, K., Mandelman, D., Modrell, B., Lee, J., Ng, B., Gibbs, M. D., and Bergquist, P. L. (2007). New high fidelity polymerases from *Thermococcus* species. *Protein Expr. Purif.* 52, 19–30.

Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001). Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus* kodakaraensis KOD1. *J. Mol. Biol.* 306, 469–477.

Hirata, A., Kanai, T., Santangelo, T. J., Tajiri, M., Manabe, K., Reeve, J. N., Imanaka, T., and Murakami, K. S. (2008). Archaeal RNA polymerase subunits E and F are not required for transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive. *Mol. Microbiol.* 70, 623–633.

Hotta, Y., Ezaki, S., Atomi, H., and Imanaka, T. (2002). Extremely stable and versatile carboxylesterase from a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 68, 3925–3931.

Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H., and Imanaka, T. (2006). Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.* 61, 898–909.

Imanaka, T., and Atomi, H. (2002). Catalyzing "hot" reactions: enzymes from hyperthermophilic archaea. *Chem. Rec.* 2, 149–163.

Imanaka, T., Fukui, T., and Fujiwara, S. (2001). Chitinase from *Thermococcus kodakarensis* KOD1. *Meth. Enzymol.* 330, 319–329.

Ishino, S., Fujino, S., Tomita, H., Ogino, H., Takao, K., Daiyasu, H., Kanai, T., Atomi, H., and Ishino, Y. (2011). Biochemical and genetical analyses of the three mcm genes from the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *Genes Cells* 16, 1176–1189.

Izumi, M., Fujiwara, S., Shiraki, K., Takagi, M., Fukui, K., and Imanaka, T. (2001). Utilization of immobilized archaeal chaperonin for enzyme stabilization. *J. Biosci. Bioeng.* 91, 316–318.

Jarrell, K. F., Walters, A. D., Bochiwal, C., Borgia, J. M., Dickinson, T., and Chong, J. P. (2011). Major players on the microbial stage: why archaea are

important. *Microbiology* 157(Pt 4), 919–936.

Kanai, T., Akerboom, J., Takedomi, S., van de Werken, H. J., Blombach, F., van der Oost, J., Murakami, T., Atomi, H., and Imanaka, T. (2007). A global transcriptional regulator in *Thermococcus kodakarensis* controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. *J. Biol. Chem.* 282, 33659–33670.

Kanai, T., Imanaka, H., Nakajima, A., Uwamori, K., Omori, Y., Fukui, T., Atomi, H., and Imanaka, T. (2005). Continuous hydrogen production by the hyperthermophilic archaeon, *Thermococcus kodakarensis* KOD1. *J. Biotechnol.* 116, 271–282.

Kanai, T., Matsuoka, R., Beppu, H., Nakajima, A., Okada, Y., Atomi, H., and Imanaka, T. (2011). Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *J. Bacteriol.* 193, 3109–3116.

Kanai, T., Takedomi, S., Fujiwara, S., Atomi, H., and Imanaka, T. (2010). Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, *Thermococcus kodakarensis*. *J. Biochem.* 147, 361–370.

Kelly, R. M., Dijkhuizen, L., and Leemhuis, H. (2009). Starch and alpha-glucan acting enzymes, modulating their properties by directed evolution. *J. Biotechnol.* 140, 184–193.

Kim, J. W., and Peeples, T. L. (2006). Screening extremophiles for bioconversion potentials. *Biotechnol. Prog.* 22, 1720–1724.

Kim, Y. J., Lee, H. S., Kim, E. S., Bae, S. S., Lim, J. K., Matsumi, R., Lebedinsky, A. V., Sokolova, T. G., Kozhevnikova, D. A., Cha, S. S., Kim, S. J., Kwon, K. K., Imanaka, T., Atomi, H., Bonch-Osmolovskaya, E. A., Lee, J. H., and Kang, S. G. (2010). Formate-driven growth coupled with H(2) production. *Nature* 467, 352–355.

Kobori, H., Ogino, M., Orita, I., Nakamura, S., Imanaka, T., and Fukui, T. (2010). Characterization of NADH oxidase/NADPH polysulfide oxidoreductase and its unexpected participation in oxygen sensitivity in an anaerobic hyperthermophilic archaeon. *J. Bacteriol.* 192, 5192–5202.

Lam, W. L., and Doolittle, W. F. (1992). Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*. *J. Biol. Chem.* 267, 5829–5834.

Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol. Rev.* 35, 577–608.

Li, Z., Pan, M., Santangelo, T. J., Chemnitz, W., Yuan, W., Edwards, J. L., Hurwitz, J., Reeve, J. N., and Kelman, Z. (2011). A novel DNA nuclease is stimulated by association with the GINS complex. *Nucleic Acids Res.* 39, 6114–6123.

Li, Z., Santangelo, T. J., Cuboňová, L., Reeve, J. N., and Kelman, Z. (2010). Affinity purification of an archaeal DNA replication protein network. *MBio* 1, e00221-10.

Lipscomb, G. L., Stirrett, K., Schut, G. J., Yang, F., Jenney, F. E. Jr., Scott, R. A., Adams, M. W., and Westpheling, J. (2011). Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* 77, 2232–2238.

Littlechild, J. A. (2011). Thermophilic archaeal enzymes and applications in biocatalysis. *Biochem. Soc. Trans.* 39, 155–158.

Louvel, H., Kanai, T., Atomi, H., and Reeve, J. N. (2009). The Fur iron regulator-like protein is cryptic in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *FEMS Microbiol. Lett.* 295, 117–128.

Matsubara, K., Yokooji, Y., Atomi, H., and Imanaka, T. (2011). Biochemical and genetic characterization of the three metabolic routes in *Thermococcus kodakarensis* linking glyceraldehyde 3-phosphate and 3-phosphoglycerate. *Mol. Microbiol.* 81, 1300–1312.

Matsumi, R., Atomi, H., Driessen, A. J., and van der Oost, J. (2011). Isoprenoid biosynthesis in archaea – biochemical and evolutionary implications. *Res. Microbiol.* 162, 39–52.

Matsumi, R., Manabe, K., Fukui, T., Atomi, H., and Imanaka, T. (2007). Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. *J. Bacteriol.* 189, 2683–2691.

Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T., and, Imanaka, T. (1994). Purification and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. *Appl. Environ. Microbiol.* 60, 4559–4566.

Morimoto, N., Fukuda, W., Nakajima, N., Masuda, T., Terui, Y., Kanai, T., Oshima, T., Imanaka, T., and Fujiwara, S. (2010). Dual biosynthesis pathway for longer-chain polyamines in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *J. Bacteriol.* 192, 4991–5001.

Murakami, T., Kanai, T., Takata, H., Kuriki, T., and Imanaka, T. (2006). A novel branching enzyme of the GH-57 family in the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *J. Bacteriol.* 188, 5915–5924.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G. J., Hattori, M., Kanai, A., Atomi, H., Takai, K., and Takami, H. (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223.

Orita, I., Sato, T., Yurimoto, H., Kato, N., Atomi, H., Imanaka, T., and Sakai, Y. (2006). The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakarensis*. *J. Bacteriol.* 188, 4698–4704.

Pan, M., Santangelo, T. J., Li, Z., Reeve, J. N., and Kelman, Z. (2011). *Thermococcus kodakarensis* encodes three MCM homologs but only one is essential. *Nucleic Acids Res.* 39, 9671–9680.

Rashid, N., Kanai, T., Atomi, H., and Imanaka, T. (2004). Among multiple phosphomannomutase gene orthologues, only one gene encodes a protein with phosphoglucomutase and phosphomannomutase activities in *Thermococcus kodakarensis*. *J. Bacteriol.* 186, 6070–6076.

Rother, M., and Metcalf, W. W. (2005). Genetic technologies for Archaea. *Curr. Opin. Microbiol.* 8, 745–751.

Santangelo, T. J., and Artsimovitch, I. (2011). Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* 9, 319–329.

Santangelo, T. J., Cubonová, L., James, C. L., and Reeve, J. N. (2007). TFB1 or TFB2 is sufficient for *Thermococcus kodakarensis* viability and for basal transcription in vitro. *J. Mol. Biol.* 367, 344–357.

Santangelo, T. J., Cubonová, L., Matsumi, R., Atomi, H., Imanaka, T., and Reeve, J. N. (2008a). Polarity in archaeal operon transcription in *Thermococcus kodakarensis*. *J. Bacteriol.* 190, 2244–2248.

Santangelo, T. J., Cubonová, L., and Reeve, J. N. (2008b). Shuttle vector expression in *Thermococcus kodakarensis*: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 74, 3099–3104.

Santangelo, T. J., Cubonová, L., and Reeve, J. N. (2010). *Thermococcus kodakarensis* genetics: TK1827-encoded beta-glycosidase, new positive-selection protocol, and targeted and repetitive deletion technology. *Appl. Environ. Microbiol.* 76, 1044–1052.

Santangelo, T. J., Cubonová, L., and Reeve, J. N. (2011). Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* 81, 897–911.

Santangelo, T. J., Cubonová, L., Skinner, K. M., and Reeve, J. N. (2009). Archaeal intrinsic transcription termination in vivo. *J. Bacteriol.* 191, 7102–7108.

Santangelo, T. J., and Reeve, J. N. (2006). Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *J. Mol. Biol.* 355, 196–210.

Santangelo, T. J., and Reeve, J. N. (2010a). Deletion of switch 3 results in an archaeal RNA polymerase that is defective in transcript elongation. *J. Biol. Chem.* 285, 23908–23915.

Santangelo, T. J., and Reeve, J. N. (2010b). "Genetic tools and manipulations of the hyperthermophilic heterotrophic archaeon *Thermococcus kodakarensis*," in *Extremophiles Handbook*, ed. K. Horikoshi (Tokyo: Springer), 567–582.

Sato, T., and Atomi, H. (2011). Novel metabolic pathways in Archaea. *Curr. Opin. Microbiol.* 14, 307–314.

Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315, 1003–1006.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *J. Bacteriol.* 185, 210–220.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2005). Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Appl. Environ. Microbiol.* 71, 3889–3899.

Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H., and Imanaka, T. (2004). Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakarensis* and other hyperthermophiles. *J. Bacteriol.* 186, 5799–5807.

Shiraki, K., Tsuji, M., Hashimoto, Y., Fujimoto, K., Fujiwara, S., Takagi, M., and Imanaka, T. (2003). Genetic, enzymatic, and structural analyses of phenylalanyl-tRNA synthetase from *Thermococcus kodakarensis* KOD1. *J. Biochem.* 134, 567–574.

Soler, N., Gaudin, M., Marguet, E., and Forterre, P. (2011). Plasmids, viruses and virus-like membrane vesicles from Thermococcales. *Biochem. Soc. Trans.* 39, 36–44.

Soler, N., Justome, A., Quevillon-Cheruel, S., Lorieux, F., Le Cam, E., Marguet, E., and Forterre, P. (2007). The rolling-circle plasmid pTN1 from the hyperthermophilic archaeon *Thermococcus* nautilus. *Mol. Microbiol.* 66, 357–370.

Takemasa, R., Yokooji, Y., Yamatsu, A., Atomi, H., and Imanaka, T. (2011). *Thermococcus kodakarensis* as a host for gene expression and protein secretion. *Appl. Environ. Microbiol.* 77, 2392–2398.

Tumbula, D. L., and Whitman, W. B. (1999). Genetics of *Methanococcus*: possibilities for functional genomics in archaea. *Mol. Microbiol.* 33, 1–7.

Yamaji, K., Kanai, T., Nomura, S. M., Akiyoshi, K., Negishi, M., Chen, Y., Atomi, H., Yoshikawa, K., and Imanaka, T. (2009). Protein synthesis in giant liposomes using the in vitro translation system of *Thermococcus kodakarensis*. *IEEE Trans. Nanobioscience* 8, 325–331.

Yamamoto, T., Matsuda, T., Sakamoto, N., Matsumura, H., Inoue, T., Morikawa, M., Kanaya, S., and Kai, Y. (2003). Crystallization and preliminary X-ray analysis of TBP-interacting protein from the hyperthermophilic archaeon Thermococcus kodakaraensis strain KOD1. *Acta Crystallogr. D Biol. Crystallogr.* 59, 372–374.

Yokooji, Y., Tomita, H., Atomi, H., and Imanaka, T. (2009). Pantoate kinase and phosphopantothenate synthetase, two novel enzymes necessary for CoA biosynthesis in the archaea. *J. Biol. Chem.* 284, 28137–28145.

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

Received: 22 March 2012; paper pending published: 15 April 2012; accepted: 13 May 2012; published online: 08 June 2012.

*Citation: Hileman TH and Santangelo TJ (2012) Genetics techniques for Thermococcus kodakarensis. Front. Microbio. **3**:195. doi: 10.3389/fmicb.2012.00195*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

REVIEW ARTICLE
published: 24 July 2012
doi: 10.3389/fmicb.2012.00259

Genetic manipulation of *Methanosarcina* spp.

Petra R. A. Kohler and William W. Metcalf*

Department of Microbiology, B103 Chemical and Life Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Edited by:
Frank T. Robb, University of California, USA

Reviewed by:
Jonathan H. Badger, J. Craig Venter Institute, USA
Paul Blum, University of Nebraska, USA
Isaac Cann, University of Illinois, USA

***Correspondence:**
William W. Metcalf, Department of Microbiology, University of Illinois at Urbana-Champaign, B103 C&LSL, 601 S. Goodwin, Urbana, IL 61801-3763, USA.
e-mail: metcalf@illinois.edu

The discovery of the third domain of life, the Archaea, is one of the most exciting findings of the last century. These remarkable prokaryotes are well known for their adaptations to extreme environments; however, Archaea have also conquered moderate environments. Many of the archaeal biochemical processes, such as methane production, are unique in nature and therefore of great scientific interest. Although formerly restricted to biochemical and physiological studies, sophisticated systems for genetic manipulation have been developed during the last two decades for methanogenic archaea, halophilic archaea and thermophilic, sulfur-metabolizing archaea. The availability of these tools has allowed for more complete studies of archaeal physiology and metabolism and most importantly provides the basis for the investigation of gene expression, regulation and function. In this review we provide an overview of methods for genetic manipulation of *Methanosarcina* spp., a group of methanogenic archaea that are key players in the global carbon cycle and which can be found in a variety of anaerobic environments.

Keywords: *Methanosarcina*, genetic manipulation, mutagenesis, markerless deletion, genetypic complementation, gene expression

INTRODUCTION

Life on earth, as currently perceived by mankind, is encompassed in three major domains Archaea, Eukarya and Bacteria. The recognition of Archaea as a distinct phylogenetic lineage is a fairly recent discovery (Woese et al., 1990). Initially it was assumed that Archaea were strictly anaerobic and mostly extremophiles, inhabiting environmental niches hostile to most other organisms, such as submarine volcanic vents, solfataric hot springs, or soda lakes. It is known today that Archaea are ubiquitous, represent a significant portion of the global biomass and play important roles in global ecosystems and biochemical cycling (Delong and Pace, 2001; Jarrell et al., 2011).

Archaea share metabolic and physiologic features with Eukarya and Bacteria, but they are also unique in many ways (Jarrell et al., 2011; Jun et al., 2011; White, 2011). In many cases Archaea are uniquely positioned to carry out biochemical reactions that are of significant interest for industrial and biotechnological applications. Prominent examples are: methanogenesis, the reduction of CO_2 to methane; or the application of Bacteriorhodopsin, a light driven proton pump, in solar cells and radiation sensors (Thavasi et al., 2009; Ahmadi and Yeow, 2011; De Vrieze et al., 2012).

Biochemical, structural and physiological studies have provided significant insight into the third domain of life over the last three decades (Cavicchioli, 2011). However, our understanding of Archaea still lags behind our knowledge of Eukarya and Bacteria. From a scientific and biotechnological point of view this gap in knowledge needs to be filled urgently. An important step toward achieving this goal was the development of systems for genetic manipulation of members of the methanogenic archaea, the halophilic archaea and the thermophilic, sulfur-metabolizing archaea (Rother and Metcalf, 2005; Buan et al., 2011; Leigh et al., 2011).

The study of methanogenic Archaea has a long-standing history because of their central role in the global carbon cycle and their potential application in biofuel production (Fox et al., 1977; Jarrell et al., 2011; De Vrieze et al., 2012). Methanogenic metabolism converts a limited number of one-carbon (C-1) compounds and acetic acid to methane through a series of coenzyme-bound intermediates in a process that drives the generation of ATP (Thauer et al., 2008). Methanogenic Archaea are represented in five orders, the *Methanococcales*, the *Methanosarcinales*, the *Methanobacteriales*, the *Methanomicrobiales* and the *Methanopyrales* (Liu and Whitman, 2008). Genetic systems are available for organisms in the first two orders (Rother and Metcalf, 2005; Leigh et al., 2011). Three classes of methanogens have been proposed based on physiological, biochemical and genomic traits (Anderson et al., 2009). Class I methanogens, *Methanococcales*, *Methanobacteriales* and *Methanopyrales*, and class II methanogens, *Methanomicrobiales*, are usually hydrogenotrophic; they use H_2/CO_2 and sometimes formate as substrates for methanogenesis (Anderson et al., 2009). Class III methanogens, *Methanosarcinales* metabolize a variety of substrates, H_2/CO_2, C-1 compounds, such as methylamines, methylsulfides or methanol, and acetate in four different methanogenic pathways (Thauer et al., 2008; Anderson et al., 2009). Three organisms, *Methanosarcina bakeri* Fusaro and *Methanosarcina acetivorans* C2A and *Methanosarcina mazei* Gö1 are well established in methanogenesis research and the complete genome sequences of all three organisms are publicly available. The major difference between these organisms lies within their ability to utilize methanogenic substrates,

M. bakeri Fusaro and *M. mazei* Gö1 use all known substrates, whereas *M. acetivorans* C2A lacks the ability to grow on H_2 and CO_2 (Thauer et al., 2008; Guss et al., 2009; Kulkarni et al., 2009).

In the past 15 years, a number of techniques that allow the study of gene function *in vivo* have been developed for these *Methanosarcina* species. This review focuses on the genetic manipulation of *Methanosarcina* spp. and provides an overview about established methods including, random and targeted mutagenesis, complementation, and reporter gene fusions.

TRANSCRIPTION, TRANSLATION, AND DNA REPAIR IN ARCHAEA

The transcription, translation, and DNA repair systems play an important role during genetic manipulation and therefore need to be considered for method development. The proteins involved in DNA repair facilitate the incorporation of cloned DNA into the genome of the target organisms, while the expression of cloned genes and selective markers is driven by the transcription and translation system.

Like bacteria, known Archaea store their genetic information on a circular chromosome and in many cases on additional plasmids (Keeling et al., 1994; Koonin and Wolf, 2008). In addition, transcriptional units often comprise operons (Brown et al., 1989). Nevertheless, the basal archaeal transcription apparatus including the 11- to 13-subunit DNA-dependent RNA Polymerase (RNAP) is closely related to the eukaryotic Pol II system and some rRNAs and tRNAs were found to be characteristic for Archaea (Woese et al., 1978; Jun et al., 2011).

The archaeal promoter is very similar to that of eukaryotes and includes a TATA box element located approximately 30 bp upstream of the transcriptional start (Hausner et al., 1991; Jun et al., 2011). At least two factors, homologues of the eukaryotic TATA-binding protein (TBP) and Transcription Factor II B (TFIIB), are required for transcription initiation. TFIIB interacts with the B recognition element (BRE), a purine rich sequence located directly upstream of the TATA-box (Jun et al., 2011). The majority of the transcriptional regulators identified to date are homologous to known bacterial regulators and transcriptional regulation generally seems to follow the bacterial model. This is especially evident from the mechanisms of actions described for transcriptional repressors that bind DNA close to the promoter and either occlude the TATA box and the BRE element or inhibit the recruitment of the RNAP (Bell, 2005; Jun et al., 2011; Malys and McCarthy, 2011). However, archaeal histones were shown to also be involved in transcription regulation similar to what is known for Eukaryotes (Reeve, 2003; Jun et al., 2011).

Archaeal mRNAs do not posses a 5′ cap structure and long poly-A tails (Brown and Reeve, 1985, 1986; Hennigan and Reeve, 1994). Translation can occur in a Shine-Dalgarno-dependent or -independent fashion; mRNAs lacking a Shine-Dalgarno sequence are either leaderless or are led by a 5′ untranslated region (Dennis, 1997; Malys and McCarthy, 2011).

DNA repair is achieved through double stranded break repair and homologous recombination. Mechanistically homologous recombination follows the eukaryotic model and the key enzymes involved, Mre11, RadA, and Rad50, are also found in Eukaryotes (White, 2011).

CULTIVATION OF *Methanosarcina* spp.

Successfully working with any organism requires the ability to satisfy their nutritional and environmental needs. *Methanosarcina* spp. and methanogens, in general, are very sensitive to oxygen. All experimental manipulations and the cultivation of strains must be performed under strict anaerobic conditions. Media are commonly bicarbonate buffered minimal media with a pH from 6.8 to 7.0 and must have a redox potential of at least –300 mV to keep *Methanosarcina* metabolically functional (Balch et al., 1979; Sowers et al., 1993; Metcalf et al., 1998).

DNA DELIVERY, EXCHANGEABLE PROMOTERS, SELECTABLE, AND COUNTERSELECTABLE MARKERS

The successful genetic manipulation of any organism depends on four basic requirements: (1) The ability to grow clonal colonies from single cells, (2) a DNA delivery system, (3) promoters for the expression of cloned genes, and (4) selectable genetic markers.

Methanosarcina often grow in multicellular packets containing from a few, up to tens of thousands of cells. Aggregates of *Methanosarcina* cells are encapsulated by methanochondroitin, a positively charged exopolysaccharide composed of *N*-acetyl-D-galactosamine, galacturonic- and glucuronic-acid that interacts with the S-layer, a proteinaceous matrix that is found immediately adjacent to the cell membrane (Kreisl and Kandler, 1986; Ellen et al., 2010). Cell aggregates represent a physical barrier for exogenous DNA and therefore impede genetic manipulation. The production of methanochondroitin seems to be environmentally regulated and it is not produced under conditions of high osmolarity, resulting in unicellular growth (Sowers et al., 1993). Based on this discovery high salt media were developed that allow the growth of clonal populations from single cells and render *Methanosarcina* more easily accessible for genetic manipulations (Sowers et al., 1993).

Liposome- and polyethylene glycol (PEG)-mediated transformation methods have been developed for *Methanosarcina* (Metcalf et al., 1997; Oelgeschlager and Rother, 2009). The former achieves high transformation frequencies, up to 2×10^8 transformants per μg DNA, representing about 20% of the CFU for *M. acetivorans*. Frequencies in other *Methanosarcina* spp. are slightly lower but still sufficient for most purposes (Metcalf et al., 1997; Ehlers et al., 2005).

Selectable or testable phenotypes are mandatory to monitor DNA up-take or exchange. The basic requirement, to achieve the establishment of a tractable phenotype through a selectable or counterselectable marker, is a reliable gene expression system. Two exchangeable transcription systems are routinely used for the expression of cloned genes in *Methanosarcina* spp. Both systems comprise a strong constitutive promoter, p*mcr* or p*mcrB*, that drives the transcription of the methyl-reductase operons in *Methanococcus voltae* and *Methanosarcina barkeri* Fusaro, respectively, and its corresponding ribosomal binding sites (RBS) and transcriptional terminators (Metcalf et al., 1997; Zhang et al., 2000).

Two antibiotic resistance markers have been developed for *Methanosarcina*. Resistance to the protein synthesis inhibitor Puromycin is easily achievable, by introducing the *pac* (puromycin transacetylase) gene from *Streptomyces alboniger* controlled by p*mcr* (Gernhardt et al., 1990; Metcalf et al., 1997). A second selectable marker, that codes for resistance to pseudomonic acid was created by mutagenesis of the isoleucyl-tRNA synthetase gene (*ileS12*) from *M. barkeri* Fusaro (Boccazzi et al., 2000).

A counterselectable marker system for the construction of *M. acetivorans* C2A and *M. barkeri* Fusaro mutants was developed based on the deletion of the *hpt* gene, encoding a hypoxanthine phosphoribosyltransferase (Pritchett et al., 2004; Guss et al., 2008). The Hpt protein is part of the purine salvage pathway and catalyzes the phosphorylation of various purines to the corresponding monophosphate. Toxic purine analogs like 8-aza-2,6-diaminopurine (8ADP) also serve as substrate for Hpt. Incorporation of these toxic bases into DNA can be lethal (Bowen and Whitman, 1987; Bowen et al., 1996). A *M. acetivorans* C2A Δ*hpt* strain is approximately 35-fold more resistant to 8ADP than the wild type; 8ADP sensitivity is restored upon reintroduction of the *hpt* gene. This phenotype has proven useful for the creation of unmarked deletion mutants and therefore a series of Δ*hpt* parental strains were constructed (see below; Pritchett et al., 2004; Guss et al., 2008).

SHUTTLE VECTORS

A series of autonomously replicating *Escherichia coli*/*Methanosarcina* shuttle vectors have been developed based on the native *M. acetivorans* pC2A plasmid that is present in about six copies per cell (Sowers and Gunsalus, 1988; Metcalf et al., 1997). These constructs contain a *pac* cassette for selection and the pC2A replicon for replication in *Methanosarcina*. The plasmids can be used for a variety of *Methanosarcina* spp. The β-lactamase gene enables selection and *ori*R6Kγ allows replication in *E. coli*. The variety of plasmids provide different multi cloning sites and some include the *lacZ* α gene for blue and white screening to allow facile identification of recombinant plasmids in *E. coli* (Metcalf et al., 1997).

The shuttle plasmid vectors gave rise to a series of molecular tools that make forward and reverse genetics possible in *Methanosarcina*. All plasmids for the genetic manipulation of *Methanosarcina* are designed to be maintained in appropriate *E. coli* hosts. Ampicillin- (*bla*), chloramphenicol- (*cat*) or kanamycin-resistance (*aph*) genes serve as selective markers. Replication and copy number control are dependent on either the high copy number control pMB1*ori*, the medium copy *pir*-dependent *ori*R6Kγ or the single copy number maintenance *oriS* in combination with the inducible high copy number maintenance *oriV*-TrfA system (Metcalf et al., 1997; Zhang et al., 2000; Pritchett et al., 2004; Guss et al., 2008).

FORWARD GENETICS

A transposon system derived from the *mariner* transposable element H*imar1*, which transposes to random sites in the genome at high frequency, is available for mutagenesis in *M. acetivorans* C2A (Zhang et al., 2000). The transposition of *mariner* elements is solely dependent on their cognate transposases and does not require host factors, hence they can be used for *in vivo* mutagenesis of a wide variety of eukaryotic and prokaryotic organisms (Lampe et al., 1996; Plasterk, 1996; Tosi and Beverley, 2000). The modified mini-H*imar1* transposon carries elements for selection in *Methanosarcina* (*pac* cassette) and in *E. coli* (*aph*) as well as the *ori*R6Kγ, for identification of the mutated gene through cloning of the insertion in *E. coli.* A suicide vector serves as the delivery plasmid. The *mariner* transposase (*tnp*) gene, controlled by the p*mcrB*, is not part of the transposable element, but resides on the plasmid and therefore perishes with the vector. This construct assures the stability of the insert, since the Tnp mediated transposition is fully reversible (Zhang et al., 2000).

REVERSE GENETICS

Integration of a cloned DNA fragment into the chromosome of *Methanosarcina* through homologous recombination was first achieved by Conway De Macario et al. (1996). Gene replacement or disruption can be facilitated by simply introducing the linearized cloning vector pBluescript (Stratagene) containing the appropriate homologous *Methanosarcina* DNA fragment and selective marker (Rother et al., 2005). As described above, only two selective markers, the *pac* and *ileS12* genes, are available for *Methanosarcina*. Thus, markerless deletions are preferred, because this allows repetitive use of both markers.

The first unmarked mutant constructed was a *M. acetivorans* C2A Δ*hpt* strain that served as the parent for subsequent deletion strains, since the loss of a functional *hpt* gene provides a convenient counter selective system (Pritchett et al., 2004). This principle was extended in a series of *M. acetivorans* C2A and *M. bakeri* Fusaro strains engineered to harbor a highly efficient site-specific recombination system at the Δ*hpt* locus that allows integration of genes into the chromosome (Guss et al., 2008).

This system utilizes the host factor independent ΦC31 integrase of the *Streptomyces* bacteriophage ΦC31, encoded by the *int* gene, and the ΦC31 phage integration sites (Thorpe and Smith, 1998; Guss et al., 2008). The Int protein catalyzes site-specific recombination between the ΦC31 site at the Δ*hpt* locus (*attB* or *attP*) and the corresponding ΦC31 (*attP* or *attB*) of a plasmid (**Figure 1**). The reaction is unidirectional and results in the stabile integration of the entire plasmid. This allows the constitutive expression of the ΦC31 integrase from p*mcrB* without destabilizing the insert. Some constructs contain an artificial *tetR-int* operon, expressed from p*mcrB* (Guss et al., 2008). In general, the *tetR* gene codes for the TetR transcriptional repressor that binds the *tetO* operator of the target promoter and prevents transcription. TetR releases the promoter upon binding tetracycline and gene expression is initiated. This tetracycline regulated promoter system allows the tight regulation of cloned genes and is used for different applications in *Methanosarcina* (see below; Beck et al., 1982; Guss et al., 2008).

MARKERLESS EXCHANGE

Two methods, employing different mechanisms, have been established for routine use to generate unmarked mutants in *Methanosarcina* in the Δ*hpt* background (Pritchett et al., 2004;

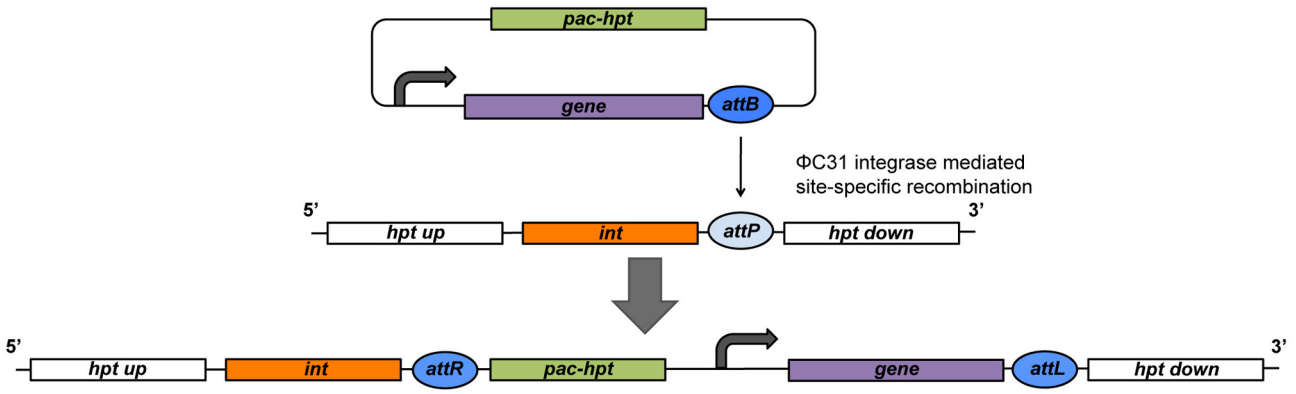


FIGURE 1 | Schematic of a site-specific recombination event at the engineered Δ*hpt* locus of the *Methanosarcina* chromosome. The ΦC31 integrase (*int*) is constitutively expressed at the Δ*hpt* locus that also harbors a ΦC31 attachment site (*attB* in this figure). The integration plasmid contains the gene of interest controlled by a p*mcrB* derivative (curved arrow) and a ΦC31 attachment site (*attP* in this figure). *pac*, gene for puromycin *N*-acetyltransferase; *hpt*, gene for hypoxanthine phosphoribosyltransferase.

Welander and Metcalf, 2008). Both methods rely on homologous recombination, using deletion constructs consisting of either the 5′- and 3′- sequences of the gene, resulting in gene disruption, or the flanking up- and downstream sequences, resulting in gene deletion (Pritchett et al., 2004; Welander and Metcalf, 2008).

The first method makes use of the *pac* selectable marker and *hpt* counter selectable marker and relies on an unstable merodiploid intermediate state (**Figure 2**; Pritchett et al., 2004). Plasmid pMP44 does not replicate in *Methanosarcina* and carries the *hpt* gene in addition to the *pac* cassette. Derivatives of pMP44 containing a deletion construct are used to transform the *Methanosarcina* Δ*hpt* parental strain to puromycin resistance and 8ADP sensitivity during the first recombination event. The plasmid integrates into the chromosome at either the up- or downstream homologous regions, resulting in an unstable merodiploid, which is resolved by a successive second recombination event in the absence of selective pressure. This event removes the vector backbone including the *pac* and *hpt* genes and renders the progeny puromycin sensitive and 8ADP resistant (**Figure 2**). The second step gives rise to two types of 8ADP-resistant progeny, half the offspring will retain the desired mutation and the other half will retain the wild type locus, provided that the target gene is not essential and there is no difference in growth rate between the mutant and the wild type strain. Mutants are then identified via PCR screening or phenotypic testing.

The second established method is based on creating a marked mutant that allows subsequent removal of the selective and counter selective marker (**Figure 3A**; Rother and Metcalf, 2005; Welander and Metcalf, 2008). Plasmid pJK301 carries an artificial *pac-hpt* operon, flanked by two Flp recombinase recognition sites (FRT). Two multi cloning sites for the cloning of the desired homologous regions are located directly up- and downstream of the FRT-*pac-hpt*-FRT cassette. The plasmid, carrying the deletion construct needs to be linearized before transformation to ensure that the wild type gene is replaced or disrupted with the FRT-*pac-hpt*-FRT cassette through a double-recombination event. The resulting marked mutant is puromycin resistant and 8ADP sensitive. Introducing plasmid pMR55 that expresses the *Saccharomyces* Flp recombinase creates an unmarked, puromycin sensitive and 8ADP resistant mutant. The Flp recombinase recognizes the FRT sites and excises the *pac-hpt* cassette, leaving behind a "FRT scar." The Flp recombinase system is highly effective, but fully reversible (Huang et al., 1997; Schweizer, 2003). Thus, the *flp* gene is transiently expressed and pMR55 does not replicate in *Methanosarcina*, in order to stabilize the construct.

GENOTYPIC COMPLEMENTATION AND REPORTER GENE FUSIONS

The respective wild type gene(s) can be reintroduced into *Methanosarcina* mutants through either plasmid based complementation or single integration into the chromosome.

GENOTYPIC COMPLEMENTATION VIA MULTICOPY PLASMIDS AND SINGLE COPY INTEGRATION

Multicopy expression can be achieved by using the *E. coli*/*Methanosarcina* shuttle vector pWM321 (Metcalf et al., 1997; Zhang et al., 2002). The plasmid offers a large multi cloning site but no promoter to drive the expression of the gene of interest and no RBS. A promoter-RBS-gene fusion needs to be cloned into pWM321. This provides the opportunity to tailor the construct to whatever is required by either choosing the native promoter of the gene, a promoter of known strength, or a tetracycline regulated promoter. The desired gene can either be cloned with its native RBS or placed under the control of the RBS of the methyl-reductase operon from *M. barkeri* Fusaro (p*mcrB* RBS).

GENOTYPIC COMPLEMENTATION THROUGH SINGLE COPY INTEGRATION

A number of plasmids were designed to express genes of interest from either a constitutive or tetracycline regulated promoter

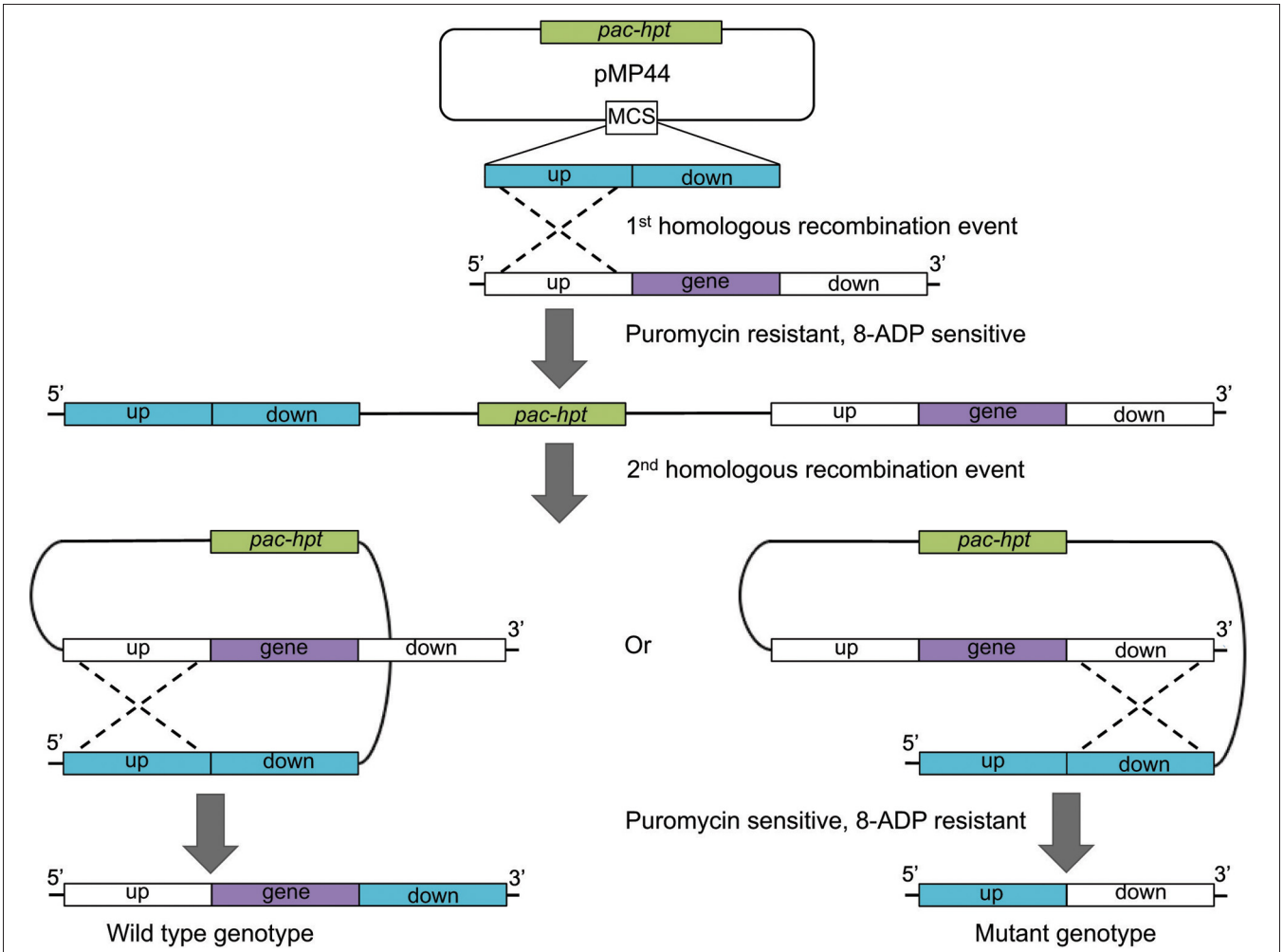


FIGURE 2 | Schematic of the construction of an unmarked deletion mutant strain using pMP44. Plasmid pMP44 carries a deletion construct comprising the homologous up- and downstream regions of the target gene. The circular plasmid is transformed into a *Methanosarcina* Δ*hpt* strain. A homologous recombination event (dotted line) creates a puromycin resistant and 8-aza-2,6-diaminopurine (8-ADP) sensitive merodiploid intermediate. The resolution of the instable intermediate either reconstitutes the wild type genotype or results in the deletion of the target gene, both events restore puromycin sensitivity and 8-ADP resistance. *pac*, gene for puromycin *N*-acetyltransferase; *hpt*, gene for hypoxanthine phosphoribosyltransferase.

(Guss et al., 2008). The p*mcrB* promoter achieves the highest level of constitutive expression. Lower constitutive expression levels can be obtained from the tetracycline regulated promoters, p*mcrB*(tetO1), p*mcrB*(tetO3), and p*mcrB*(tetO4), which decrease in promoter strength, respectively; if the plasmids are introduced into a host that does not express the *tetR* gene from the Δ*hpt* locus. All plasmids provide the p*mcrB* RBS and the p*mcrB* terminator and carry the FRT-*pac-hpt*-FRT cassette as well as a ΦC31 phage integration site. They do not replicate in *Methanosarcina* but integrate at the Δ*hpt* locus through site-specific recombination. Mutants that have successfully inserted the desired plasmid into the chromosome can be identified based on puromycin resistance and 8ADP sensitivity.

Another useful feature of the single copy expression vectors is one of the λ attachment sites (λ *attA* or λ *attB*) that can be used for retrofitting with other plasmids via a commercially available λ integrase system (Guss et al., 2008). This system can be used to turn the single copy expression vectors into autonomous, multicopy *Methanosarcina* plasmids through site-specific recombination with plasmid pAMG40 that carries the pC2A replicon. A *Methanosarcina* Δ*hpt* strain lacking the *int* gene and the ΦC31 integration site needs to be used as host for the expression plasmid:pAMG40 constructs to avoid integration into the chromosome.

GENE EXPRESSION STUDIES, TESTING FOR GENE ESSENTIALITY AND PROMOTER SWAP

REPORTER GENE FUSIONS

The widely used reporter gene *uidA*, encoding the β-glucuronidase from *E. coli*, is functional in *Methanosarcina*. Plasmid pAB79 was designed to allow the construction of transcriptional or translational fusions to *uidA* (**Figure 4**). Stable single copy fusions are advantageous for expression studies; pAB79

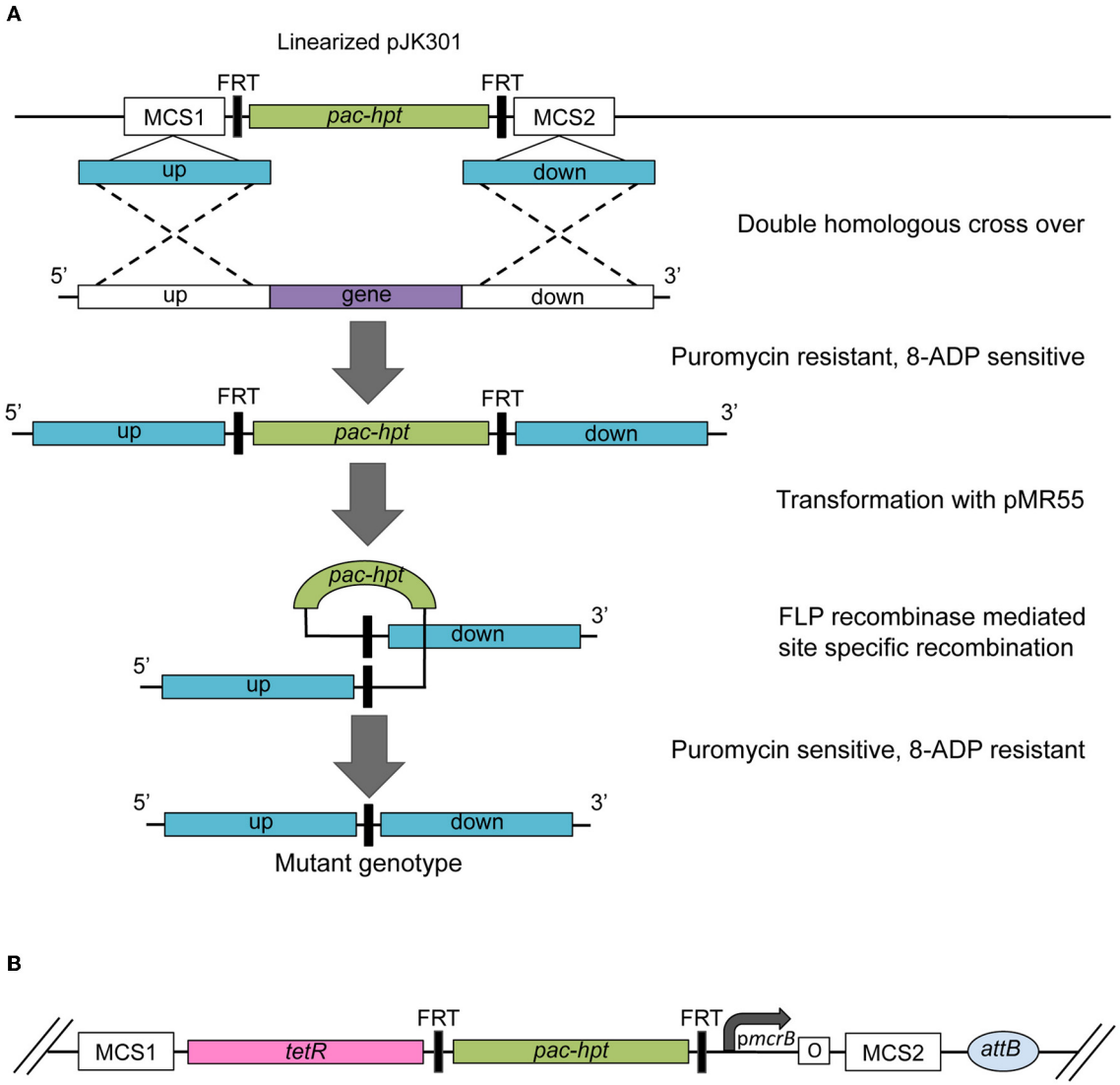


FIGURE 3 | (A) Schematic of the construction of an unmarked deletion mutant strain using pMPJK301 and pMR55. The *pac-hpt* cassette of plasmid pMJK301 is flanked by the homologous up- and downstream regions of the target gene. The linearized plasmid is transformed into a *Methanosarcina* Δ*hpt* strain. A double homologous recombination event (dotted line) creates a puromycin resistant and 8-aza-2,6-diaminopurine (8-ADP) marked deletion mutant. Successive transformation with plasmid pMR55 results in the removal of the *pac-hpt* cassette via site-specific recombination between the FRT sites mediated by the Flp recombinase expressed from pMR55. The resulting mutant strain is puromycin sensitive and 8-ADP resistant. **(B)** Schematic of a plasmid used for promoter swaps. The p*mcrB* is represented by a bend arrow and the Tet operator by the letter "o" enclosed in a box. *pac*, gene for puromycin *N*-acetyltransferase; *hpt*, gene for hypoxanthine phosphoribosyltransferase; *tetR*, gene for the TetR transcriptional repressor.

derivatives therefore integrate into the chromosome at the Δ*hpt* locus via ΦC31 mediated recombination (Guss et al., 2008).

For a transcriptional fusion the (putative) promoter of interest is cloned as an AscI/BamH1 fragment immediately upstream of tandem translational stop codons. The stop codons are followed by the further downstream located p*mcrB* RBS and *uidA* gene. In this construct transcription initiation is dependent on the cloned fragment, but translation initiation of the β-glucuronidase is based on the strong and effective p*mcrB* RBS provided by pAB79. The tandem stop codons guarantee no translational read through from the cloned fragment into the *uidA* gene (**Figure 4**).

To create a translational fusion the desired promoter is cloned as an AscI/NdeI fragment and replaces the tandem stop codons and the p*mcrB* RBS completely. The promoter's native RBS and other translational elements are maintained by converting the start codon of the respective gene into the NdeI site used for cloning. Transcriptional and translational initiation are dependent on the elements provided by the cloned fragment in this construct (**Figure 4**; Guss et al., 2008).

TESTING FOR GENE ESSENTIALITY AND PROMOTER SWAP

The methods for gene deletion discussed in the preceding sections cannot be employed to study the function of essential genes.

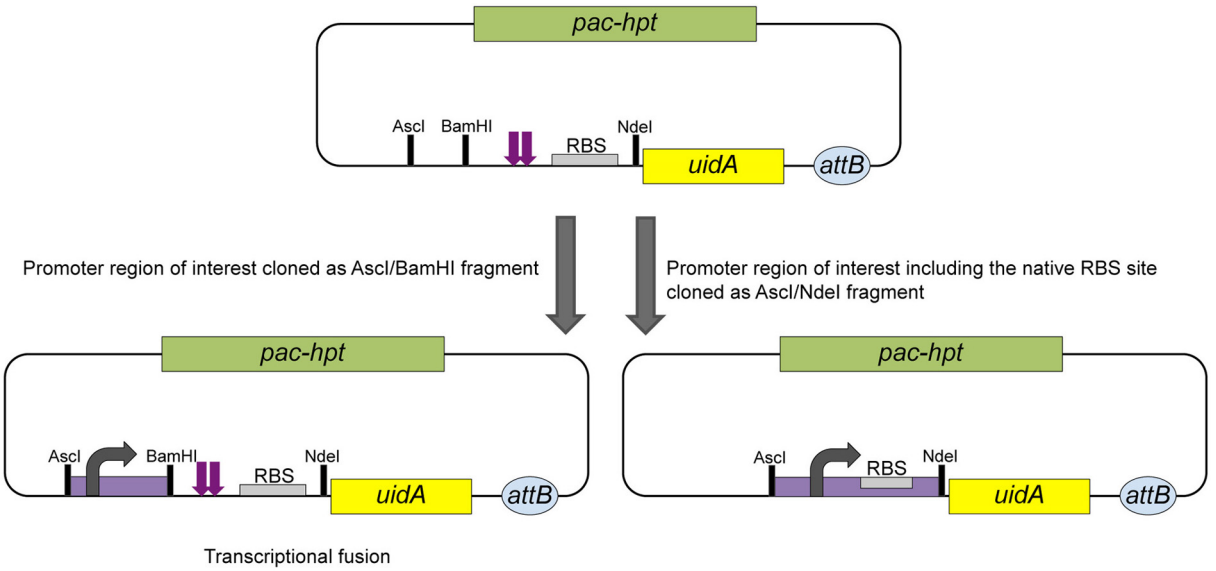


FIGURE 4 | Schematic of the construction of transcriptional and translational reporter gene fusions using pAB79. Transcriptional fusions can be achieved by cloning the promoter region of interest as an AscI/BamHI fragment into pAB79. In this construct transcription is initiated from the cloned promoter (bend arrow). During translation no read through can occur from the cloned fragment into the uidA gene (β-glucuronidase) due the to tandem stopcodons (purple arrows). Translation of the β-glucuronidase is initiated from the p*mcrB* RBS (gray box). To create a translational fusion the promoter of interest, including its translational elements, is cloned as an AscI/NdeI fragment. *pac*, gene for puromycin *N*-acetyltransferase; *hpt*, gene for hypoxanthine phosphoribosyltransferase; *attB*, ΦC31 attachment site.

Instead, a promoter swap can be used to create a conditional mutant as a way to investigate gene essentiality. In this method, the native promoter of the target gene is replaced with one of the tetracycline regulated promoters. A series of plasmids is available that are similar in design to the deletion vector pJK301 that carry the FRT-*pac*-*hpt*-FRT cassette for selection (**Figure 3B**; Guss et al., 2008). A *tetR* cassette is located upstream and a tetracycline dependent promoter downstream of the marker cassette. Two multi cloning sites flank this construct allowing the subsequent cloning of the upstream homologous region of the targeted promoter and the 5′ homologous region of the gene of interest. The promoter swap relies on the same mechanisms as the construction of a deletion mutant using the pJK301/pMR55 system (**Figure 3A**). If the targeted gene is essential the resulting mutant strain will only be viable in the presence of tetracycline.

CONCLUSIONS AND FUTURE DIRECTIONS

The availability of the genetic tools described above has greatly improved our knowledge of the methanogenic process in *Methanosarcina* species. Recent genetic studies have provided a better understanding of the C1 oxidation/reduction pathway and the energy-conserving electron transport chain of *M. bakeri* Fusaro (Welander and Metcalf, 2008; Kulkarni et al., 2009). It was also possible to characterize the biological roles of isozymes involved in different methanogenic pathways of *M. acetivorans* C2A, such as methanol and methylamine specific methyltransferases as well as cytoplasmic and membrane-bound heterosulfide reductases (Bose et al., 2009; Buan and Metcalf, 2010). The hydrogenases required for hydrogenotrophic methanogenesis are conserved in *M. bakeri* Fusaro and *M. acetivorans* C2A, cis-acting mutations were identified in the promoter regions of the respective genes of the latter that prevent gene expression and render the strain incapable of using H_2/CO_2 as growth substrates (Galagan et al., 2002; Guss et al., 2009). The ability to genetically manipulate *Methanosarcina* has not only contributed to a deeper understanding of methanogenesis, but has paved the way for the study of other cellular and metabolic functions such as transcriptional and post-transcriptional gene regulation (Bose et al., 2006, 2009; Bose and Metcalf, 2008; Opulencia et al., 2009) and mechanisms for synthesis and insertion of the 22nd genetically encoded amino acid, pyrrolysine, into methylamine methyltransferases of *Methanosarcina* (Mahapatra et al., 2006, 2007).

Ongoing and future research employing genetic methods in combination with the classical physiological and biochemical approaches will undoubtedly expand our understanding of methanogenic archaea even further. In addition, the ability to genetically manipulate *Methanosarcina* spp. opens up the possibility to metabolically engineer strains for the production of biogas from organic materials. Methane, one of the major end products of methanogenesis, represents a clean burning and renewable energy source. *Methanosarcina* spp. are among the most promising candidates for routine use in biofuel production, because among the methanogenic archaea *Methanosarcina* have the broadest range in substrates for methanogenesis and exhibit the highest tolerance to environmental stresses (De Vrieze et al., 2012). Engineering efforts to further enhance stress tolerance, optimize the efficiency of substrate usage and broaden the substrate range of *Methanosarcina* spp. could significantly improve their performance in industrial

settings like anaerobic digesters, and optimize methane yields. For example, the introduction of a gene coding for a broad-specificity esterase from *Pseudomonas veronii* into *M. acetivornas* has resulted in a strain that efficiently converts methyl acetate and methyl propionate to methane (Lessner et al., 2010).

ACKNOWLEDGMENTS

The authors acknowledge the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy (DE-FG02-02ER15296) and the National Science Foundation (MCB1022462) for support during the preparation of this review.

REFERENCES

Ahmadi, M., and Yeow, J. T. (2011). Fabrication and characterization of a radiation sensor based on bacteriorhodopsin. *Biosens. Bioelectron.* 26, 2171–2176.

Anderson, I., Ulrich, L. E., Lupa, B., Susanti, D., Porat, I., Hooper, S. D., Lykidis, A., Sieprawska-Lupa, M., Dharmarajan, L., Goltsman, E., Lapidus, A., Saunders, E., Han, C., Land, M., Lucas, S., Mukhopadhyay, B., Whitman, W. B., Woese, C., Bristow, J., and Kyrpides, N. (2009). Genomic characterization of *Methanomicrobiales* reveals three classes of methanogens. *PLoS ONE* 4:e5797. doi: 10.1371/journal.pone.0005797

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

Beck, C. F., Mutzel, R., Barbe, J., and Muller, W. (1982). A multifunctional gene (*tetR*) controls Tn10-encoded tetracycline resistance. *J. Bacteriol.* 150, 633–642.

Bell, S. D. (2005). Archaeal transcriptional regulation–variation on a bacterial theme? *Trends Microbiol.* 13, 262–265.

Boccazzi, P., Zhang, J. K., and Metcalf, W. W. (2000). Generation of dominant selectable markers for resistance to pseudomonic acid by cloning and mutagenesis of the *ileS* gene from the archaeon *Methanosarcina barkeri* Fusaro. *J. Bacteriol.* 182, 2611–2618.

Bose, A., Kulkarni, G., and Metcalf, W. W. (2009). Regulation of putative methyl-sulphide methyltransferases in *Methanosarcina acetivorans* C2A. *Mol. Microbiol.* 74, 227–238.

Bose, A., and Metcalf, W. W. (2008). Distinct regulators control the expression of methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A. *Mol. Microbiol.* 67, 649–661.

Bose, A., Pritchett, M. A., Rother, M., and Metcalf, W. W. (2006). Differential regulation of the three methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A. *J. Bacteriol.* 188, 7274–7283.

Bowen, T. L., Lin, W. C., and Whitman, W. B. (1996). Characterization of guanine and hypoxanthine phosphoribosyltransferases in *Methanococcus voltae*. *J. Bacteriol.* 178, 2521–2526.

Bowen, T. L., and Whitman, W. B. (1987). Incorporation of exogenous purines and pyrimidines by *Methanococcus voltae* and isolation of analog-resistant mutants. *Appl. Environ. Microbiol.* 53, 1822–1826.

Brown, J. W., Daniels, C. J., and Reeve, J. N. (1989). Gene structure, organization, and expression in Archaebacteria. *Crit. Rev. Microbiol.* 16, 287–338.

Brown, J. W., and Reeve, J. N. (1985). Polyadenylated, noncapped RNA from the archaebacterium *Methanococcus vannielii*. *J. Bacteriol.* 162, 909–917.

Brown, J. W., and Reeve, J. N. (1986). Polyadenylated RNA isolated from the archaebacterium *Halobacterium halobium*. *J. Bacteriol.* 166, 686–688.

Buan, N., Kulkarni, G., and Metcalf, W. (2011). Genetic methods for *Methanosarcina* species. *Methods Enzymol.* 494, 23–42.

Buan, N. R., and Metcalf, W. W. (2010). Methanogenesis by *Methanosarcina acetivorans* involves two structurally and functionally distinct classes of heterodisulfide reductase. *Mol. Microbiol.* 75, 843–853.

Cavicchioli, R. (2011). Archaea–timeline of the third domain. *Nat. Rev. Microbiol.* 9, 51–61.

Conway De Macario, E., Guerrini, M., Dugan, C. B., and Macario, A. J. (1996). Integration of foreign DNA in an intergenic region of the archaeon *Methanosarcina mazei* without effect on transcription of adjacent genes. *J. Mol. Biol.* 262, 12–20.

De Vrieze, J., Hennebel, T., Boon, N., and Verstraete, W. (2012). *Methanosarcina*: the rediscovered methanogen for heavy duty biomethanation. *Bioresour. Technol.* 112, 1–9.

Delong, E. F., and Pace, N. R. (2001). Environmental diversity of Bacteria and Archaea. *Syst. Biol.* 50, 470–478.

Dennis, P. P. (1997). Ancient ciphers: translation in Archaea. *Cell* 89, 1007–1010.

Ehlers, C., Weidenbach, K., Veit, K., Deppenmeier, U., Metcalf, W. W., and Schmitz, R. A. (2005). Development of genetic methods and construction of a chromosomal *glnK1* mutant in *Methanosarcina mazei* strain Gö1. *Mol. Genet. Genomics* 273, 290–298.

Ellen, A. F., Zolghadr, B., Driessen, A. M., and Albers, S. V. (2010). Shaping the archaeal cell envelope. *Archaea* 2010, 608243.

Fox, G. E., Magrum, L. J., Balch, W. E., Wolfe, R. S., and Woese, C. R. (1977). Classification of methanogenic Bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. U.S.A.* 74, 4537–4541.

Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., Fitzhugh, W., Calvo, S., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., Dearellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R. D., Cann, I., Graham, D. E., Grahame, D. A., Guss, A. M., Hedderich, R., Ingram-Smith, C., Kuettner, H. C., Krzycki, J. A., Leigh, J. A., Li, W., Liu, J., Mukhopadhyay, B., Reeve, J. N., Smith, K., Springer, T. A., Umayam, L. A., White, O., White, R. H., Conway De Macario, E., Ferry, J. G., Jarrell, K. F., Jing, H., Macario, A. J., Paulsen, I., Pritchett, M., Sowers, K. R., Swanson, R. V., Zinder, S. H., Lander, E., Metcalf, W. W., and Birren, B. (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* 12, 532–542.

Gernhardt, P., Possot, O., Foglino, M., Sibold, L., and Klein, A. (1990). Construction of an integration vector for use in the Archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. *Mol. Gen. Genet.* 221, 273–279.

Guss, A. M., Kulkarni, G., and Metcalf, W. W. (2009). Differences in hydrogenase gene expression between *Methanosarcina acetivorans* and *Methanosarcina barkeri*. *J. Bacteriol.* 191, 2826–2833.

Guss, A. M., Rother, M., Zhang, J. K., Kulkarni, G., and Metcalf, W. W. (2008). New methods for tightly regulated gene expression and highly efficient chromosomal integration of cloned genes for *Methanosarcina* species. *Archaea* 2, 193–203.

Hausner, W., Frey, G., and Thomm, M. (1991). Control regions of an archaeal gene. A TATA box and an initiator element promote cell-free transcription of the tRNA(Val) gene of *Methanococcus vannielii*. *J. Mol. Biol.* 222, 495–508.

Hennigan, A. N., and Reeve, J. N. (1994). mRNAs in the methanogenic Archaeon *Methanococcus vannielii*: numbers, half-lives and processing. *Mol. Microbiol.* 11, 655–670.

Huang, L. C., Wood, E. A., and Cox, M. M. (1997). Convenient and reversible site-specific targeting of exogenous DNA into a bacterial chromosome by use of the FLP recombinase: the FLIRT system. *J. Bacteriol.* 179, 6076–6083.

Jarrell, K. F., Walters, A. D., Bochiwal, C., Borgia, J. M., Dickinson, T., and Chong, J. P. (2011). Major players on the microbial stage: why Archaea are important. *Microbiology* 157, 919–936.

Jun, S. H., Reichlen, M. J., Tajiri, M., and Murakami, K. S. (2011). Archaeal RNA polymerase and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* 46, 27–40.

Keeling, P. J., Charlebois, R. L., and Doolittle, W. F. (1994). Archaebacterial genomes: eubacterial form and eukaryotic content. *Curr. Opin. Genet. Dev.* 4, 816–822.

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of Bacteria and Archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719.

Kreisl, P., and Kandler, O. (1986). Chemical structure or the cell wall polymer of *Methanosarcina*. *System Appl. Microbiol.* 7, 293–299.

Kulkarni, G., Kridelbaugh, D. M., Guss, A. M., and Metcalf, W. W. (2009). Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of *Methanosarcina barkeri*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15915–21590.

Lampe, D. J., Churchill, M. E., and Robertson, H. M. (1996). A purified mariner transposase is sufficient to mediate transposition *in vitro*. *EMBO J.* 15, 5470–5479.

Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol. Rev.* 35, 577–608.

Lessner, D. J., Lhu, L., Wahal, C. S., and Ferry, J. G. (2010). An engineered methanogenic pathway derived from the domains Bacteria and Archaea. *MBio* 1, e00243–10.

Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic Archaea. *Ann. N.Y. Acad. Sci.* 1125, 171–789.

Mahapatra, A., Patel, A., Soares, J. A., Larue, R. C., Zhang, J. K., Metcalf, W. W., and Krzycki, J. A. (2006). Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine. *Mol. Microbiol.* 59, 56–66.

Mahapatra, A., Srinivasan, G., Richter, K. B., Meyer, A., Lienard, T., Zhang, J. K., Zhao, G., Kang, P. T., Chan, M., Gottschalk, G., Metcalf, W. W., and Krzycki, J. A. (2007). Class I and class II lysyl-tRNA synthetase mutants and the genetic encoding of pyrrolysine in *Methanosarcina* spp. *Mol. Microbiol.* 64, 1306–1318.

Malys, N., and McCarthy, J. E. (2011). Translation initiation: variations in the mechanism can be anticipated. *Cell. Mol. Life Sci.* 68, 991–1003.

Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R., and Wolfe, R. S. (1997). A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. U.S.A.* 94, 2626–2631.

Metcalf, W. W., Zhang, J. K., and Wolfe, R. S. (1998). An anaerobic, intrachamber incubator for growth of *Methanosarcina* spp. on methanol-containing solid media. *Appl. Environ. Microbiol.* 64, 768–770.

Oelgeschlager, E., and Rother, M. (2009). *In vivo* role of three fused corrinoid/methyl transfer proteins in *Methanosarcina acetivorans*. *Mol. Microbiol.* 72, 1260–1272.

Opulencia, R. B., Bose, A., and Metcalf, W. W. (2009). Physiology and posttranscriptional regulation of methanol:coenzyme M methyltransferase isozymes in *Methanosarcina acetivorans* C2A. *J. Bacteriol.* 191, 6928–6935.

Plasterk, R. H. (1996). The Tc1/mariner transposon family. *Curr. Top. Microbiol. Immunol.* 204, 125–143.

Pritchett, M. A., Zhang, J. K., and Metcalf, W. W. (2004). Development of a markerless genetic exchange method for *Methanosarcina acetivorans* C2A and its use in construction of new genetic tools for methanogenic Archaea. *Appl. Environ. Microbiol.* 70, 1425–1433.

Reeve, J. N. (2003). Archaeal chromatin and transcription. *Mol. Microbiol.* 48, 587–598.

Rother, M., Boccazzi, P., Bose, A., Pritchett, M. A., and Metcalf, W. W. (2005). Methanol-dependent gene expression demonstrates that methyl-coenzyme M reductase is essential in *Methanosarcina acetivorans* C2A and allows isolation of mutants with defects in regulation of the methanol utilization pathway. *J. Bacteriol.* 187, 5552–5559.

Rother, M., and Metcalf, W. W. (2005). Genetic technologies for Archaea. *Curr. Opin. Microbiol.* 8, 745–751.

Schweizer, H. P. (2003). Applications of the *Saccharomyces cerevisiae* Flp-FRT system in bacterial genetics. *J. Mol. Microbiol. Biotechnol.* 5, 67–77.

Sowers, K. R., Boone, J. E., and Gunsalus, R. P. (1993). Disaggregation of *Methanosarcina* spp. and growth as single cells at elevated osmolarity. *Appl. Environ. Microbiol.* 59, 3832–3839.

Sowers, K. R., and Gunsalus, R. P. (1988). Plasmid DNA from the acetotrophic methanogen *Methanosarcina acetivorans*. *J. Bacteriol.* 170, 4979–4982.

Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). Methanogenic Archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* 6, 579–591.

Thavasi, V., Lazarova, T., Filipek, S., Kolinski, M., Querol, E., Kumar, A., Ramakrishna, S., Padros, E., and Renugopalakrishnan, V. (2009). Study on the feasibility of bacteriorhodopsin as bio-photosensitizer in excitonic solar cell: a first report. *J. Nanosci. Nanotechnol.* 9, 1679–1687.

Thorpe, H. M., and Smith, M. C. (1998). *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5505–5510.

Tosi, L. R., and Beverley, S. M. (2000). *cis* and trans factors affecting Mos1 mariner evolution and transposition *in vitro*, and its potential for functional genomics. *Nucleic Acids Res.* 28, 784–790.

Welander, P. V., and Metcalf, W. W. (2008). Mutagenesis of the C1 oxidation pathway in *Methanosarcina barkeri*: new insights into the Mtr/Mer bypass pathway. *J. Bacteriol.* 190, 1928–1936.

White, M. F. (2011). Homologous recombination in the Archaea: the means justify the ends. *Biochem. Soc. Trans.* 39, 15–19.

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

Woese, C. R., Magrum, L. J., and Fox, G. E. (1978). Archaebacteria. *J. Mol. Evol.* 11, 245–251.

Zhang, J. K., Pritchett, M. A., Lampe, D. J., Robertson, H. M., and Metcalf, W. W. (2000). *In vivo* transposon mutagenesis of the methanogenic archaeon *Methanosarcina acetivorans* C2A using a modified version of the insect mariner-family transposable element Himar1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9665–9670.

Zhang, J. K., White, A. K., Kuettner, H. C., Boccazzi, P., and Metcalf, W. W. (2002). Directed mutagenesis and plasmid-based complementation in the methanogenic archaeon *Methanosarcina acetivorans* C2A demonstrated by genetic analysis of proline biosynthesis. *J. Bacteriol.* 184, 1449–1454.

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

Received: 03 May 2012; paper pending published: 21 May 2012; accepted: 03 July 2012; published online: 24 July 2012.
*Citation: Kohler PRA and Metcalf WW (2012) Genetic manipulation of Methanosarcina spp.. Front. Microbio. **3**:259. doi: 10.3389/fmicb.2012.00259*
This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 01 May 2012
doi: 10.3389/fmicb.2012.00163

Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon *Pyrococcus furiosus*

Gerrit J. Schut[1†], William J. Nixon[1†], Gina L. Lipscomb[1], Robert A. Scott[1,2] and Michael W. W. Adams[1]*

[1] *Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA*
[2] *Department of Chemistry, University of Georgia, Athens, GA, USA*

Edited by:
Frank T. Robb, University of California, USA

Reviewed by:
Rudolf Kurt Thauer, Max Planck Society – Institute for Terrestrial Microbiology, Germany
John Van Der Oost, Wageningen University, Netherlands

***Correspondence:**
Michael W. W. Adams, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA.
e-mail: adams@bmb.uga.edu

[†]Gerrit J. Schut and William J. Nixon have contributed equally to this work.

Pyrococcus furiosus grows optimally near 100˚C by fermenting carbohydrates to produce hydrogen (H_2) or, if elemental sulfur (S^0) is present, hydrogen sulfide instead. It contains two cytoplasmic hydrogenases, SHI and SHII, that use NADP(H) as an electron carrier and a membrane-bound hydrogenase (MBH) that utilizes the redox protein ferredoxin. We previously constructed deletion strains lacking SHI and/or SHII and showed that they exhibited no obvious phenotype. This study has now been extended to include biochemical analyses and growth studies using the ΔSHI and ΔSHII deletion strains together with strains lacking a functional MBH (ΔmbhL). Hydrogenase activity in cytoplasmic extracts of various strains demonstrate that SHI is responsible for most of the cytoplasmic hydrogenase activity. The ΔmbhL strain showed no growth in the absence of S^0, confirming the hypothesis that, in the absence of S^0, MBH is the only enzyme that can dispose of reductant (in the form of H_2) generated during sugar oxidation. Under conditions of limiting sulfur, a small but significant amount of H_2 was produced by the ΔmbhL strain, showing that SHI can produce H_2 from NADPH *in vivo*, although this does not enable growth of ΔmbhL in the absence of S^0. We propose that the physiological function of SHI is to recycle H_2 and provide a link between external H_2 and the intracellular pool of NADPH needed for biosynthesis. This likely has a distinct energetic advantage in the environment, but it is clearly not required for growth of the organism under the usual laboratory conditions. The function of SHII, however, remains unknown.

Keywords: hydrogenase, energy metabolism, sulfur, ferredoxin, *Pyrococcus furiosus*, thermophile, anaerobe

INTRODUCTION

Hydrogen gas (H_2) plays an important role in anaerobic metabolism as the majority of anaerobes contain the enzyme hydrogenase responsible for the reversible interconversion of molecular hydrogen, protons, and electrons. Hydrogenases can be grouped into three classes based on the metal composition of their active site: [NiFe]-hydrogenases, [FeFe]-hydrogenases, and the more recently defined [Fe]-hydrogenases, so far restricted to certain methanogenic organisms (Vignais et al., 2001; Shima and Thauer, 2007). Organisms in the bacterial domain contain both [NiFe]- and [FeFe]-hydrogenases, while archaeal organisms are known to utilize [NiFe]- and [Fe]-hydrogenases (Vignais and Billoud, 2007). Almost all of the anaerobic archaea contain one or more [NiFe]-hydrogenases, implying that H_2 metabolism plays an important role in the extreme environments in which many of these organisms are found (Vignais and Billoud, 2007).

Pyrococcus furiosus is a well-studied hyperthermophile belonging to the order Thermococcales. It grows optimally at 100˚C by peptide and carbohydrate fermentation with H_2, organic acids, and CO_2 being the main fermentation products (Fiala and Stetter, 1986; Kengen and Stams, 1994; Driskill et al., 1999; Adams et al., 2001). *P. furiosus* degrades glucose via a modified Embden–Meyerhof glycolytic pathway that utilizes the low potential electron carrier protein ferredoxin (Fd) in place of NAD for all oxidative steps, resulting in no overall substrate level ATP yield (Kengen et al., 1994; de Vos et al., 1998; Verhees et al., 2004). Because of the low redox potential of Fd (−480 mV), H_2 production (−420 mV) with Fd is thermodynamically favorable (Park et al., 1991; Smith et al., 1995; Hagedoorn et al., 1998). Therefore, all reducing equivalents generated in sugar metabolism can be disposed of as H_2, with production of up to 4 mol of H_2 per mol glucose oxidized (Verhees et al., 2004). The enzyme responsible for H_2 formation is a unique membrane-bound hydrogenase (MBH) complex, which uses the energy from this exergonic reaction to create an ion gradient across the membrane that drives ATP synthesis, resulting in the generation of an estimated 0.3 mol of ATP per mol H_2 (Sapra et al., 2003). This system is one of the simplest forms of respiration. Besides MBH, *P. furiosus* also contains two cytoplasmic hydrogenases (SHI and SHII), and these homologous enzymes each consist of four subunits. Based on kinetic studies, SHI and SHII are proposed to be involved in H_2 recycling to provide NADPH for biosynthesis (Ma and Adams, 2001b; van Haaster et al., 2008).

Pyrococcus furiosus can utilize carbohydrates for growth in either the presence or absence of elemental sulfur (S^0; Fiala and Stetter, 1986; Adams et al., 2001). In the presence of S^0, *P. furiosus* undergoes a major metabolic shift to utilize S^0 as an electron

acceptor, thereby switching production from H_2 to H_2S. When S^0 becomes available, there is a dramatic decrease in the expression of all three hydrogenase operons correlated with a decrease in hydrogenase activity, and concomitantly, there is a large increase in expression of genes related to S^0 reduction, such as those encoding the membrane-bound oxidoreductase (MBX) as well as the NAD(P)H-linked sulfur reductase (NSR; Adams et al., 2001; Schut et al., 2007). The response to S^0 is mediated at least in part by the redox switch containing transcriptional regulator SurR (Lipscomb et al., 2009; Yang et al., 2010). MBX is highly homologous to MBH and is thought to function analogously to MBH by using Fd as an electron acceptor and conserving energy via formation of an ion gradient (Silva et al., 2000; Schut et al., 2007; Bridger et al., 2011). NSR has been proposed to reduce S^0 synergistically with MBX in which MBX can provide NADPH using Fd as electron donor (Schut et al., 2007; Bridger et al., 2011).

In a previous study it was shown that strains containing deletions of either SHI or SHII or both have no growth phenotype under the conditions tested, suggesting that other enzyme(s), independent of H_2, can guide electrons from carbohydrate oxidation to feed the NADPH pool (Lipscomb et al., 2011). This study has now been extended to include strains containing a deletion of the active subunit of MBH (mbhL) either alone or in combination with deletions of the genes encoding both soluble hydrogenases. Herein, we address the physiological function of all three hydrogenases in *P. furiosus* both in the absence and presence of sufficient and limiting concentrations of S^0.

MATERIALS AND METHODS

STRAINS AND GROWTH CONDITIONS

The *P. furiosus* strains used or constructed in the study are listed in **Table 1**. Growth medium was prepared as previously described (Adams et al., 2001), and contained maltose (5 g/L) as the primary carbon source, supplemented with 0.5 g/L yeast extract, with or without addition of S^0 at 0.5 or 2 g/L. Growth experiments were carried out in biological duplicates in 150-mL serum bottles with 75 mL medium, with incubation at 95°C and shaking at 150 rpm. Cultures for cell-free extracts were grown in a 20-L custom fermenter, and cytoplasmic fractions were prepared by ultra-centrifugation as described previously (Adams et al., 2001).

CONSTRUCTION OF GENE DELETIONS

Gene splicing by overlap extension and PCR (SOE–PCR; Horton et al., 1989) was used to construct a PCR product containing 0.5 kb regions upstream and downstream of *mbhL* on either side of a genetic marker. The destination strains (COM1 and ΔSHIΔSHII) contain a deletion of the *pyrF* gene; therefore, a cassette containing *pyrF* expressed by the PEP synthase promoter (123 bp) was used as the marker. The 500-bp flanking regions were amplified from wild-type gDNA using the two primer sets WN011 (GTCATAAAACTAAATGATGAGCATTTGACTTCATTTCTTCTCCCTC), WN013 (TTGGAGAAGAGAATTGCCCAAC), and WN012 (AGAATGGAGCTCAAGATAAATATGAAAATTGTATATGGAGTTATTGG), WN014 (AGACATCAACACACTGCTTACAC). The deletion construct for *mbhL* was transformed into *P. furiosus* COM1 and ΔSHIΔSHII selecting uracil prototrophy on solid defined medium as previously described (Lipscomb et al., 2011). To obtain *mbhL* deletions, solid medium was supplemented with 8 mM polysulfide (equivalent to approximately 1 g/L S^0) or 2 mM polysulfide with solid S^0 powder sprinkled on the surface of the plate. The *pyrF* gene deleted in the COM1, ΔSHI, ΔSHII, and ΔSHIΔSHII strains was restored to the wild-type by transformation with a PCR product of the wild-type *pyrF* locus containing ∼0.5 kb on either side of the gene. This was done to ensure all strains were uracil prototrophic so growth could be achieved in identical media. DNA was extracted from transformants as previously described (Lipscomb et al., 2011) and screened for deletion by PCR amplification of the locus using primers outside the homologous flanking regions used to construct the deletions. Transformants were further colony purified by serial passage on solid medium. Mutants were confirmed by sequence analysis of the *mbhL* region and qPCR analyses (see below).

RNA ISOLATION AND qPCR ANALYSES

Total RNA from various *P. furiosus* strains was obtained, and cDNA was synthesized as previously described (Lipscomb et al., 2011). Quantitative PCR was carried out using an Mx3000P instrument (Agilent) and the Brilliant SYBR green qPCR master mix (Agilent) with primers specific to the genes encoding SHI and SHII beta subunits (PF0891 and PF1329), *mbhK* (PF1433), *mbhL* (PF1434), *mbhM* (PF1435), and *pyrF* (PF1114), along with the constitutively expressed genes encoding the pyruvate ferredoxin oxidoreductase (POR) gamma subunit (PF0971) and DNA Polymerase (PF0983) as internal controls. Successful gene deletions were verified by the absence of a specific qPCR signal.

CELL PROTEIN, H_2S, AND H_2 ANALYSES

To monitor cell growth, the Bradford method (Bradford, 1976) was used to estimate total cell protein concentration from 1 mL culture samples, with bovine serum albumin as the standard. For H_2S and

Table 1 | *Pyrococcus furiosus* strains used in this study.

Strain	Genotype	Deleted ORF(s)	Reference or source
COM1c (MW0004)	Δ*pyrF::pyrF*	None	Lipscomb et al. (2011)
ΔSHI (MW0022)	Δ*pyrF::pyrF* Δ*shI*βγδα	PF0891–PF0894	This work
ΔSHII (MW0023)	Δ*pyrF::pyrF* Δ*shII*βγδα	PF1329–PF1332	This work
ΔSHIΔSHII (MW0016)	Δ*pyrF::pyrF* Δ*shI*βγδα Δ*shII*βγδα	PF0891–PF0894, PF1329–PF1332	This work
ΔmbhL (MW0024)	Δ*pyrF* Δ*mbhL::*P_{pep}*pyrF*	PF1114, PF1434	This work
ΔSHIΔSHIIΔmbhL (MW0025)	Δ*pyrF* Δ*shI*βγδα Δ*shII*βγδα Δ*mbhL::*P_{pep}*pyrF*	PF1114, PF0891–PF0894, PF1329–PF1332, PF1434	This work

H_2 analyses, headspace, and medium samples (500 μL each) were taken at 6 and 9 h during growth and transferred anaerobically into the double-vial system as previously reported (Schut et al., 2007). H_2S production was assayed by the methylene blue method (Chen and Mortenson, 1977), and abiotic sulfide production was subtracted from the experimental samples using control bottles containing uninoculated medium. H_2 production was measured using a gas chromatograph (GC-8A, Shimadzu, Columbia, MD, USA). Hydrogenase activity in cell-free extracts was determined by H_2 production using sodium dithionite (5 mM) as the electron donor with methyl viologen (1 mM) as the electron carrier as described previously (Ma and Adams, 2001b).

RESULTS

CHARACTERIZATION OF SHI AND SHII MUTANTS

We showed previously that strains containing deletions of either of the two cytosolic NADP-linked hydrogenases (SHI and SHII) alone or together did not produce any growth phenotype under the conditions tested (Lipscomb et al., 2011). In order to verify the effect of disruption of SHI and SHII on hydrogenase activity in the cytoplasm, we prepared cell-free extracts from ΔSHI, ΔSHII and ΔSHIΔSHII. The amount of hydrogenase activity (using the artificial electron carrier methyl viologen) in cytoplasmic fractions was not significantly affected in the ΔSHII strain but was much lower in the ΔSHI strain (<10% of that produced in the parental strains), while in the ΔSHIΔSHII strain, no hydrogenase activity could be detected (**Figure 1**). These data indicate that SHI is responsible for the majority of hydrogenase activity in the cytoplasm and confirms that the activity of the MBH is strictly associated with the membrane (Sapra et al., 2000; Silva et al., 2000).

CHARACTERIZATION OF *mbhL* MUTANTS

We have constructed strains containing a deletion of the catalytic subunit (mbhL) of the membrane-bound ferredoxin-linked H_2-producing hydrogenase (MBH), either alone or in combination with deletions of SHI and SHII. Growth of these strains was compared on maltose-based medium containing minimal yeast extract with either no S^0, limiting S^0 (0.5 g/L), or sufficient S^0 (2 g/L). Both mutants containing a deletion of *mbhL* displayed no detectable growth in the absence of S^0, but had no growth defect in the presence of sufficient S^0 (**Figure 2**). In the presence of limiting S^0 (0.5 g/L), the MBH disruption strains exhibited ∼40% less final protein at the end of log phase, although growth rate was similar to the parental strains initially. The strain devoid of all three hydrogenases (ΔSHIΔSHIIΔmbhL) did not produce any detectable H_2 under any of the growth conditions (**Figure 3**). With sufficient S^0 (2 g/L) only a very small amount of H_2 was produced in the parental strains and in ΔmbhL (<5% of that produced in the

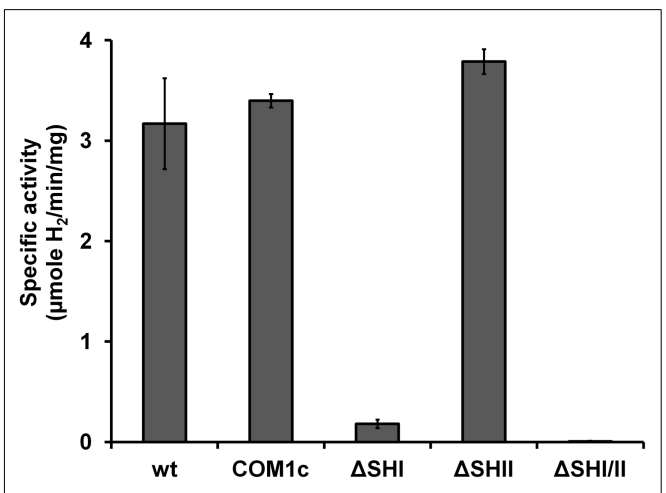


FIGURE 1 | Hydrogenase activity (using methyl viologen as electron carrier) in cytoplasmic fractions obtained from *P. furiosus* cultures of hydrogenase disruption mutants and parental strains. See **Table 1** for strain definitions.

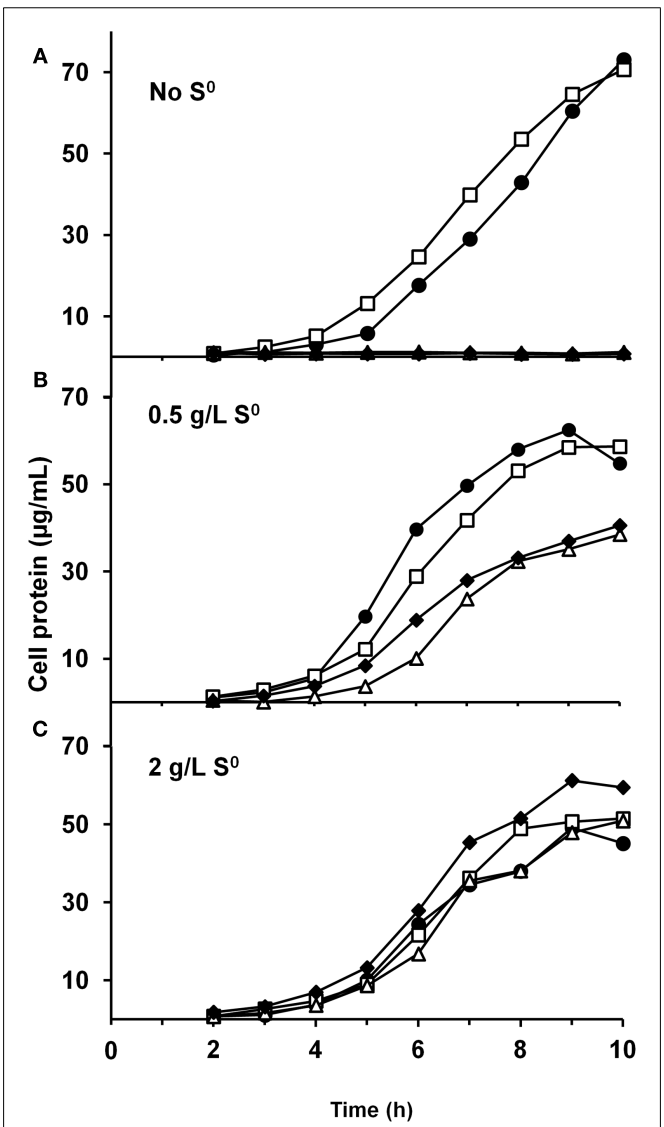


FIGURE 2 | Growth characteristics of *P. furiosus* strains grown in maltose-based medium (A), supplemented with 0.5 g/L S^0 (B), and supplemented with 2 g/L S^0 (C). The symbols represent: closed circles, COM1c; open squares, ΔSHIΔSHII; open triangles, ΔmbhL; closed diamonds, ΔSHIΔSHIIΔmbhL.

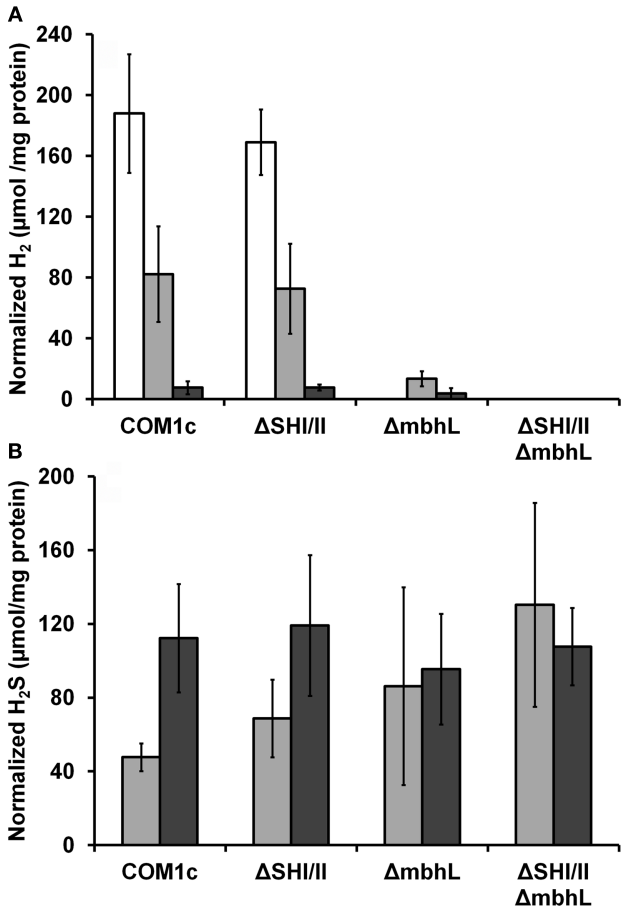


FIGURE 3 | Comparison of H_2 (A) and H_2S (B) production by *P. furiosus* strains. White bars, cells grown without S^0 (no H_2S is produced); light gray bars, cells grown with 0.5 g/L S^0; dark gray bars, cells grown with 2 g/L S^0. The amounts of H_2 and H_2S produced are given with respect to the total cell protein in the culture.

absence of S^0, **Figure 3**; Schut et al., 2007). With limiting S^0, both S^0 is reduced and H_2 is produced (c.a. 50% compared to no S^0) in the parental strains (**Figures 2** and **3**). Interestingly, when ΔmbhL was grown with limiting S^0, a small but significant amount of H_2 was produced (c.a. 20% compared to the parental strains). Therefore, the H_2 produced in the ΔmbhL strain must be catalyzed by SHI, showing that this cytosolic "uptake" hydrogenase can also produce H_2 from NADPH *in vivo*. A concentration of 0.5 g/L S^0 (equivalent to ca. 15 mM) appears to be limiting for the cell's metabolism even though the cultures are continuously shaken during growth. We observed significant amounts of S^0 left in suspension in the culture medium after growth. Since only up to 3 mM rather than 15 mM sulfide is produced in these cultures, it appears that not all S^0 in the medium is accessible to the cells.

DISCUSSION

The glycolytic pathway of *P. furiosus* only uses a low potential ferredoxin that is linked to MBH, for the disposal of all reducing equivalents as H_2, with simultaneous production of an ion gradient for energy generation (Sapra et al., 2003; Verhees et al., 2004). From this study involving all hydrogenases in *P. furiosus*, it is clear that MBH is the only enzyme that produces H_2 in wild-type cells and that no alternative electron pathway is available to *P. furiosus* that can allow for growth in the absence of S^0. In addition, no other enzyme (for example MBX) is capable of producing H_2 *in vivo*, as shown by the lack of H_2 formation in the strain lacking all three hydrogenases. Similar observations were made for the related archaeon *Thermococcus kodakaraensis*, in which the disruption of its MBH also did not allow growth under H_2 evolving conditions (Kanai et al., 2005, 2011; Santangelo et al., 2011). We propose that the *in vivo* function of SHI is to recycle H_2 for the formation of NADPH needed for biosynthesis. Although in the double deletion mutant lacking both SHI and SHII no hydrogenase activity was observed in the cytoplasmic fraction of cellular extracts, no growth phenotype was observed when either or both of these hydrogenases are absent (Lipscomb et al., 2011). In this case other enzymes must provide the pool of NADPH, and a potential candidate is the ferredoxin:NADPH oxidoreductase (FNOR) described previously (Ma and Adams, 2001a). In a previous study, an SHI overexpression strain was constructed, and this strain also did not display any obvious phenotype, although it contained almost an order of magnitude more SHI activity (Chandrayan et al., 2011). The relative amount of H_2 produced in the strain lacking both SHI and SHII was not significantly different than that in the parental strains (**Figure 3**). However, conflicting results have been reported with a *T. kodakaraensis* deletion strain lacking its SHI. One study found only a small increase in H_2 production (c.a. 10%; Kanai et al., 2011), which is more or less in agreement with our results, while another reported over a fivefold increase in relative H_2 production (Santangelo et al., 2011). Both studies used similar growth media and H_2 measurement methods, and it is not clear why these studies give such different results. In general, *T. kodakaraensis* displays growth yields and H_2 production rates similar to what has been observed in *P. furiosus* and other Thermococcales (Kanai et al., 2005; Verhaart et al., 2010).

In the natural environment, the use of H_2 recycling could have a distinct energetic advantage because the cytosolic hydrogenases could provide reductant in the form of NADPH without interfering with the energy balance through electron transport phosphorylation. However, we predict that there is a low level of H_2 recycling in *P. furiosus* when grown with maltose as the carbon source, especially considering the low growth yields on this substrate (25 g cdw/mol glucose utilized; Kengen and Stams, 1994). Assuming all major cellular components (protein, nucleic acids, and lipids) are synthesized *de novo*, we estimate about 6% H_2 recycling (Kanehisa et al., 2004, 2008). In the laboratory setting when these organisms are grown in nutrient rich conditions, H_2 recycling would not be important to the overall growth of the organism, and this may be the reason why there is a lack of phenotype for the SHI and SHII deletion strains. When SHI was first described, it was proposed to be responsible for the production of H_2, but the subsequent discovery of MBH called this into question (Ma et al., 1993; Sapra et al., 2000; Silva et al., 2000). The generation of H_2 from NADPH is thermodynamically unfavorable; however, *in vitro* this reaction can be easily demonstrated (Ma and Adams, 2001b; Verhaart et al., 2010). In this study, we have now shown that

this reaction can actually take place *in vivo*, although at a low level since SHI cannot compensate for the absence of MBH to allow growth of the ΔmbhL mutant in the absence of S^0.

Members of the order Thermococcales are characterized by the ability to use S^0 as an electron acceptor (Kelly and Adams, 1994). We have previously shown that peptides can only be utilized by *P. furiosus* (and likely most Thermococcales) in the presence of S^0, and we have concluded that S^0 is the preferred electron acceptor (Adams et al., 2001; Schut et al., 2007). When S^0 is made available to the cell, a rapid switch from H_2 production to S^0 metabolism occurs, and this is orchestrated at least in part by the redox sensitive SurR regulator (Schut et al., 2007; Lipscomb et al., 2009; Yang et al., 2010). However, *P. furiosus* does not appear to possess a high affinity S^0 binding system such as that described for *Wolinella succinogenes* (Sud; Klimmek et al., 1998). From the results presented herein it appears that the addition of sufficient S^0 to a maltose-based medium seems to consistently reduce the overall cell yield (by c.a. 10–20%). At 0.5 g/L, S^0 appears to be limiting to the cells, but the overall concentration (15 mM "S" atoms) should be sufficient to provide the sole electron sink. In this case, the cells are able to utilize both H_2 and S^0 metabolism simultaneously and produce both H_2 and H_2S. This type of mixed H_2 and S^0 metabolism has also been observed for *Staphylothermus marinus* which also contains orthologous MBH and MBX gene clusters (Hao and Ma, 2003; Anderson et al., 2009). Altogether, this suggests that *P. furiosus* has a relatively low affinity for S^0, and that, when growing on carbohydrates, it might actually prefer to generate H_2 rather than utilizing the poorly soluble S^0 as an electron acceptor.

ACKNOWLEDGMENTS

We acknowledge the Division of Chemical Sciences, Geosciences and Biosciences, Office of Basic Energy Sciences of the U.S. Department of Energy for funding in part the strain construction through grant DE-FG05-95ER20175 (to Michael W. W. Adams) and the Office of Biological and Environmental Research of the Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy for funding strain analyses through grant FG02-08ER64690 (to Robert A. Scott).

REFERENCES

Adams, M. W. W., Holden, J. F., Menon, A. L., Schut, G. J., Grunden, A. M., Hou, C., Hutchins, A. M., Jenney, F. E., Kim, C., Ma, K. S., Pan, G. L., Roy, R., Sapra, R., Story, S. V., and Verhagen, M. F. J. M. (2001). Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* 183, 716–724.

Anderson, I. J., Dharmarajan, L., Rodriguez, J., Hooper, S., Porat, I., Ulrich, L. E., Elkins, J. G., Mavromatis, K., Sun, H., Land, M., Lapidus, A., Lucas, S., Barry, K., Huber, H., Zhulin, I. B., Whitman, W. B., Mukhopadhyay, B., Woese, C., Bristow, J., and Kyrpides, N. (2009). The complete genome sequence of *Staphylothermus marinus* reveals differences in sulfur metabolism among heterotrophic Crenarchaeota. *BMC Genomics* 10, 145. doi:10.1186/1471-2164-10-145

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.

Bridger, S. L., Clarkson, S. M., Stirrett, K., Debarry, M. B., Lipscomb, G. L., Schut, G. J., Westpheling, J., Scott, R. A., and Adams, M. W. (2011). Deletion strains reveal metabolic roles for key elemental sulfur-responsive proteins in *Pyrococcus furiosus*. *J. Bacteriol.* 193, 6498–6504.

Chandrayan, S. K., Mcternan, P. M., Hopkins, R. C., Sun, J., Jenney, F. E. Jr., and Adams, M. W. (2011). Engineering hyperthermophilic archaeon *Pyrococcus furiosus* to overproduce its cytoplasmic [NiFe]-hydrogenase. *J. Biol. Chem.* 287, 3257–3264.

Chen, J. S., and Mortenson, L. E. (1977). Inhibition of methylene blue formation during determination of the acid-labile sulfide of iron-sulfur protein samples containing dithionite. *Anal. Biochem.* 79, 157–165.

de Vos, W. M., Kengen, S. W. M., Voorhorst, W. G. B., and Van Der Oost, J. (1998). Sugar utilization and its control in hyperthermophiles. *Extremophiles* 2, 201–205.

Driskill, L. E., Kusy, K., Bauer, M. W., and Kelly, R. M. (1999). Relationship between glycosyl hydrolase inventory and growth physiology of the hyperthermophile *Pyrococcus furiosus* on carbohydrate-based media. *Appl. Environ. Microbiol.* 65, 893–897.

Fiala, G., and Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov., represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* 145, 56–61.

Hagedoorn, P. L., Driessen, M. C., Van Den Bosch, M., Landa, I., and Hagen, W. R. (1998). Hyperthermophilic redox chemistry: a re-evaluation. *FEBS Lett.* 440, 311–314.

Hao, X., and Ma, K. (2003). Minimal sulfur requirement for growth and sulfur-dependent metabolism of the hyperthermophilic archaeon *Staphylothermus marinus*. *Archaea* 1, 191–197.

Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K., and Pease, L. R. (1989). Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 77, 61–68.

Kanai, T., Imanaka, H., Nakajima, A., Uwamori, K., Omori, Y., Fukui, T., Atomi, H., and Imanaka, T. (2005). Continuous hydrogen production by the hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1. *J. Biotechnol.* 116, 271–282.

Kanai, T., Matsuoka, R., Beppu, H., Nakajima, A., Okada, Y., Atomi, H., and Imanaka, T. (2011). Distinct physiological roles of the three [NiFe]-hydrogenase orthologs in the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 193, 3109–3116.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280.

Kelly, R. M., and Adams, M. W. (1994). Metabolism in hyperthermophilic microorganisms. *Antonie Van Leeuwenhoek* 66, 247–270.

Kengen, S. W., De Bok, F. A., Van Loo, N. D., Dijkema, C., Stams, A. J., and De Vos, W. M. (1994). Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus*. *J. Biol. Chem.* 269, 17537–17541.

Kengen, S. W. M., and Stams, A. J. M. (1994). Growth and energy-conservation in batch cultures of *Pyrococcus furiosus*. *FEMS Microbiol. Lett.* 117, 305–309.

Klimmek, O., Kreis, V., Klein, C., Simon, J., Wittershagen, A., and Kroger, A. (1998). The function of the periplasmic Sud protein in polysulfide respiration of *Wolinella succinogenes*. *Eur. J. Biochem.* 253, 263–269.

Lipscomb, G. L., Keese, A. M., Cowart, D. M., Schut, G. J., Thomm, M., Adams, M. W., and Scott, R. A. (2009). SurR: a transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in *Pyrococcus furiosus*. *Mol. Microbiol.* 71, 332–349.

Lipscomb, G. L., Stirrett, K., Schut, G. J., Yang, F., Jenney, F. E. Jr., Scott, R. A., Adams, M. W., and Westpheling, J. (2011). Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* 77, 2232–2238.

Ma, K., and Adams, M. W. (2001a). Ferredoxin:NADP oxidoreductase from *Pyrococcus furiosus*. *Meth. Enzymol.* 334, 40–45.

Ma, K., and Adams, M. W. (2001b). Hydrogenases I and II from *Pyrococcus furiosus*. *Meth. Enzymol.* 331, 208–216.

Ma, K., Schicho, R. N., Kelly, R. M., and Adams, M. W. (1993). Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc. Natl. Acad. Sci. U.S.A.* 90, 5341–5344.

Park, J. B., Fan, C. L., Hoffman, B. M., and Adams, M. W. (1991).

Potentiometric and electron nuclear double resonance properties of the two spin forms of the $[4Fe\text{-}4S]^{+}$ cluster in the novel ferredoxin from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *J. Biol. Chem.* 266, 19351–19356.

Santangelo, T. J., Cubonova, L., and Reeve, J. N. (2011). Deletion of alternative pathways for reductant recycling in *Thermococcus kodakarensis* increases hydrogen production. *Mol. Microbiol.* 81, 897–911.

Sapra, R., Bagramyan, K., and Adams, M. W. (2003). A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7545–7550.

Sapra, R., Verhagen, M. F. J. M., and Adams, M. W. W. (2000). Purification and characterization of a membrane-bound hydrogenase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* 182, 3423–3428.

Schut, G. J., Bridger, S. L., and Adams, M. W. (2007). Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon *Pyrococcus furiosus*: characterization of a coenzyme A- dependent NAD(P)H sulfur oxidoreductase. *J. Bacteriol.* 189, 4431–4441.

Shima, S., and Thauer, R. K. (2007). A third type of hydrogenase catalyzing H_2 activation. *Chem. Rec.* 7, 37–46.

Silva, P. J., Van Den Ban, E. C., Wassink, H., Haaker, H., De Castro, B., Robb, F. T., and Hagen, W. R. (2000). Enzymes of hydrogen metabolism in *Pyrococcus furiosus*. *Eur. J. Biochem.* 267, 6541–6551.

Smith, E. T., Blamey, J. M., Zhou, Z. H., and Adams, M. W. (1995). A variable-temperature direct electrochemical study of metalloproteins from hyperthermophilic microorganisms involved in hydrogen production from pyruvate. *Biochemistry* 34, 7161–7169.

van Haaster, D. J., Silva, P. J., Hagedoorn, P. L., Jongejan, J. A., and Hagen, W. R. (2008). Reinvestigation of the steady-state kinetics and physiological function of the soluble NiFe-hydrogenase I of *Pyrococcus furiosus*. *J. Bacteriol.* 190, 1584–1587.

Verhaart, M. R., Bielen, A. A., Van Der Oost, J., Stams, A. J., and Kengen, S. W. (2010). Hydrogen production by hyperthermophilic and extremely thermophilic bacteria and archaea: mechanisms for reductant disposal. *Environ. Technol.* 31, 993–1003.

Verhees, C. H., Kengen, S. W. M., Tuininga, J. E., Schut, G. J., Adams, M. W. W., De Vos, W. M., and Van Der Oost, J. (2004). The unique features of glycolytic pathways in Archaea. *Biochem. J.* 377, 819–822.

Vignais, P. M., and Billoud, B. (2007). Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* 107, 4206–4272.

Vignais, P. M., Billoud, B., and Meyer, J. (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* 25, 455–501.

Yang, H., Lipscomb, G. L., Keese, A. M., Schut, G. J., Thomm, M., Adams, M. W., Wang, B. C., and Scott, R. A. (2010). SurR regulates hydrogen production in *Pyrococcus furiosus* by a sulfur-dependent redox switch. *Mol. Microbiol.* 77, 1111–1122.

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

Received: 20 March 2012; accepted: 12 April 2012; published online: 01 May 2012.

Citation: Schut GJ, Nixon WJ, Lipscomb GL, Scott RA and Adams MWW (2012) Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon Pyrococcus furiosus. Front. Microbio. ***3****:163. doi: 10.3389/fmicb.2012.00163*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 05 June 2012
doi: 10.3389/fmicb.2012.00192

Heteroduplex formation, mismatch resolution, and genetic sectoring during homologous recombination in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*

Dominic Mao and Dennis W. Grogan*

Department of Biological Sciences, University of Cincinnati, Cincinnati, OH, USA

Edited by:
Frank T. Robb, University of California at Riverside, USA

Reviewed by:
Barbara Methé, The J. Craig Venter Institute, USA
Imke Schroeder, University of California Los Angeles, USA
Marla Tuffin, University of the Western Cape, South Africa

****Correspondence:***
Dennis W. Grogan, Department of Biological Sciences, University of Cincinnati, 614 Rieveschl Hall, ML0006, Clifton Court, Cincinnati, OH 45221-0006, USA.
e-mail: grogandw@ucmail.uc.edu

Hyperthermophilic archaea exhibit certain molecular-genetic features not seen in bacteria or eukaryotes, and their systems of homologous recombination (HR) remain largely unexplored *in vivo*. We transformed a *Sulfolobus acidocaldarius pyrE* mutant with short DNAs that contained multiple non-selected genetic markers within the *pyrE* gene. From 20 to 40% of the resulting colonies were found to contain two Pyr^+ clones with distinct sets of the non-selected markers. The dual-genotype colonies could not be attributed to multiple DNAs entering the cells, or to conjugation between transformed and non-transformed cells. These colonies thus appear to represent genetic sectoring in which regions of heteroduplex DNA formed and then segregated after partial resolution of inter-strand differences. Surprisingly, sectoring was also frequent in cells transformed with single-stranded DNAs. Oligonucleotides produced more sectored transformants when electroporated as single strands than as a duplex, although all forms of donor DNA (positive-strand, negative-strand, and duplex) produced a diversity of genotypes, despite the limited number of markers. The marker patterns in the recombinants indicate that *S. acidocaldarius* resolves individual mismatches through un-coordinated short-patch excision followed by re-filling of the resulting gap. The conversion events that occur during transformation by single-stranded DNA do not show the strand bias necessary for a system that corrects replication errors effectively; similar events also occur in pre-formed heteroduplex electroporated into the cells. Although numerous mechanistic details remain obscure, the results demonstrate that the HR system of *S. acidocaldarius* can generate remarkable genetic diversity from short intervals of moderately diverged DNAs.

Keywords: linear DNA, genetic transformation, mismatch repair, gene conversion

INTRODUCTION

All cellular organisms can transfer strands between two DNA molecules of nearly identical sequences in an active process termed homologous (or generalized) recombination. In addition to promoting genome diversification, homologous recombination (HR) plays various roles that support genome replication and partitioning, and these functions seem to account for its ubiquity (Michel et al., 2001; Burgess, 2004; Kreuzer, 2005). Multiple pathways of HR have been identified in model organisms, and HR proteins have diverged across major phylogenetic groups. Thus, bacteria and eukaryotic cells use distinct sub-families of ssDNA-binding proteins (the RecA and Rad51 families, respectively) to promote strand exchange, and differ further with respect to the helicases, nucleases, and accessory proteins that interact to catalyze and regulate the steps before and after this exchange (Kowalczykowski et al., 1994; Aylon and Kupiec, 2004; Krogh and Symington, 2004). Consistent with their extremely early divergence from the bacterial and eukaryotic lineages, archaea encode a third, structurally distinct sub-family of strand-exchange proteins called the RadA family (Sandler et al., 1999). Although archaea resemble bacteria with respect to certain features of cell and genome structure, archaeal RadA proteins most closely resemble the Rad51 and Dmc1 proteins of eukaryotes (Sandler et al., 1999).

Homologous recombination affects the structure and diversification of all microbial genomes, yet its functional properties in archaea have been investigated for only a few species, primarily extreme halophiles and methanogens. For hyperthermophilic archaea, HR has become increasingly important as the means of targeted gene mutation and other forms of genetic engineering, but it has not been analyzed extensively as a genetic process. Most quantitative analyses of HR in hyperthermophilic archaea have been done in *Sulfolobus acidocaldarius*, due in part to the availability of natural genetic exchange and various mutants that allow it to be quantified. In *Sulfolobus* spp., 5-fluoro-orotic acid selects spontaneous *pyrE* mutants that are uracil auxotrophs, and these mutants provide a convenient assay of recombination events that replace the defective sequence in the recipient genome with the homologous functional sequence.

Detailed analyses of recombinants generated by conjugation and transformation suggest that HR in *S. acidocaldarius* transfers multiple, short intervals of input DNA unidirectionally to the recipient genome (Hansen et al., 2005; Grogan and Rockwood,

2010). This short-patch, gene-conversion mode of HR mimics bacterial and eukaryotic HR in the absence of functional DNA mismatch repair proteins (Coic et al., 2000; Barnes and McCulloch, 2007; Lin et al., 2009), and thus remains consistent with the natural lack of MutSL homologs in hyperthermophilic archaea (White and Grogan, 2008). A defining feature of this form of HR is that it occurs within a region of heteroduplex formed between complementary strands of the two parental DNAs (Aylon and Kupiec, 2004). However, the genetic analyses performed to date in *S. acidocaldarius* do not distinguish among three possible alternatives for a heteroduplex formed between a strand of donor DNA and the opposite strand of the recipient chromosome (**Figure 1**).

For example, if the structure shown in **Figure 1** were stabilized by ligation, subsequent DNA replication and cell division would yield one recombinant cell acquiring donor markers and an unaltered recipient cell which, under selective conditions, would not be recovered (**Figure 1A**). Alternatively, after ligation, a large region of the recipient-strand opposite the donor markers may be removed and the resulting gap re-filled by DNA polymerase, converting all the recipient-strand markers to the donor allele (**Figure 1B**). In bacteria and eukaryotes, this long-patch excision is a characteristic of the DNA mismatch repair system and promotes the co-repair of markers in heteroduplex DNA (Coic et al., 2000). Replication of the resulting duplex would generate two recombinant daughter cells with the same set of donor markers (**Figure 1B**). The third alternative is intermediate between these two extremes and would generate a unique, detectable result. In this case, the selected marker, and possibly other markers, are copied to the opposite strand by conversion, as in **Figure 1B**, but one or more additional markers escape this transfer (**Figure 1C**). Replication of this modified heteroduplex would segregate two different genotypes, and because both retain the selected marker, the transformant colony retains both genotypes. Such genetically sectored colonies thus imply (i) that a heteroduplex formed during HR and led to transfer of the selected marker by conversion, and (ii) that the conversion did not include one or more additional markers.

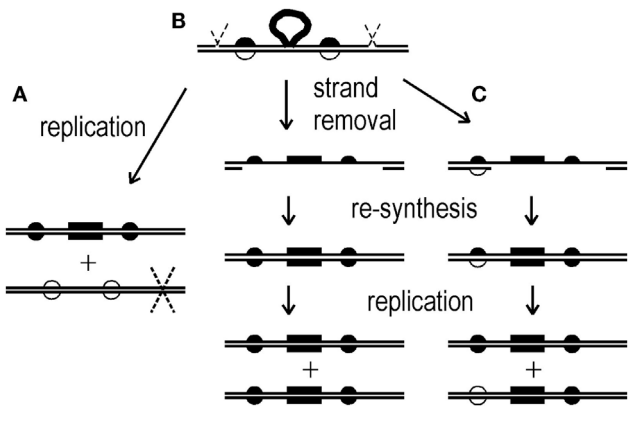


FIGURE 1 | Routes of genetic transformation by non-reciprocal events. All schemes begin with a region of heteroduplex that has formed between a donor (input) DNA and the recipient genome. The donor DNA bears a selectable marker (bar in center) and multiple non-selected markers (dark semi-circles). **(A)** Incorporation of the donor strand by trimming, gap-filling, and ligation leads to one transformed cell and a non-transformed daughter cell which is lost. **(B)** Similar incorporation of the donor strand is followed by a large gap on the opposite strand and re-filling, thereby copying all donor markers to the opposite strand (complete conversion) yielding two identical daughter cells. **(C)** A smaller gap opposite the donor strand results in partial conversion and two distinct daughter cells (genetic sectoring).

In the present study, we used transformation by multiply marked DNA to identify and analyze genetically sectored transformants of *S. acidocaldarius*. The observed patterns of marker transfer in these transformants indicate sporadic, localized resolution of individual mismatches in a heteroduplex, thereby generating extremely diverse recombinants from short intervals of moderately diverged DNAs.

MATERIALS AND METHODS

STRAINS AND GROWTH CONDITIONS

The *Sulfolobus* strain used for most experiments was *S. acidocaldarius* strain MR31, which was derived from *his-2* (histidine-requiring) mutant DG55 and has an internal deletion of the *pyrE* gene removing nt 154–171 (Reilly and Grogan, 2001). The *pyrE* gene encodes orotate phosphoribosyltransferase, which is essential for *de novo* pyrimidine biosynthesis; a loss-of-function mutation in *pyrE* thus generates uracil auxotrophs which are also resistant to 5-F orotic acid (Grogan and Gunsalus, 1993; Grogan et al., 2001). For experiments testing the effects of blocking conjugation, mutant SA1 was used; this strain is isogenic to MR31, and contains, in addition to the 18-bp internal *pyrE* deletion, a deletion of most of the *upsE* gene encoding an ATP-dependent pilus-assembly protein (Ajon et al., 2011). Cells were cultured at 76–78°C; liquid growth and plating conditions were those described previously (Grogan and Rockwood, 2010).

DONOR DNAs

The multicopy plasmid pSaPyrEv3 carries the wild-type *pyrE* sequence (Saci_1597; Chen et al., 2005) modified with synonymous base-pair substitutions (BPS) regularly spaced along the entire length of the coding sequence (Grogan and Rockwood, 2010). The plasmid does not have a *Sulfolobus* replication origin and thus cannot replicate in *Sulfolobus*. Unless otherwise noted, full-length marked *pyrE* DNA for transformation was amplified from the pyrEv3 cassette by PCR using primers sspyrEfwd and sspyrErev (**Table 1**). The thermocycler program comprised of an initial denaturation at 95°C for 3 min followed by 39 cycles of denaturation at 95°C for 30 s, annealing at 50°C, and extension at 72°C. This was followed by a final extension at 72°C for 3 min. After PCR and digestions (if any, see below), the amplicons were purified using SureClean (Bioline).

The PCR primers used to amplify the pyrEv3 cassette also provided for the specific digestion of one strand to generate the single-stranded form. The forward primer (sspyrEfwd) was designed to include a 5′-tail with *Kpn*I site (highlighted in red in **Table 1**) and a 3′-end that anneals to the start of the *pyrE* (start codon underlined in **Table 1**). The reverse primer (sspyrErev) has a corresponding 5′-tail with *Hin*P1I site (highlighted in blue in **Table 1**) and a 3′-end that anneals to end of the *pyrE* coding sequence (stop codon underlined). Digesting the resulting amplicon with *Kpn*I and *Hin*P1I creates 3′- and 5′-overhangs respectively. The amplicon was then treated with ExoIII, which specifically degrades

Table 1 | Oligonucleotides used in this study.

Designation	Sequence (5′–3′)
PCR PRIMERS	
sspyrE fwd	TGATGCGGTACCATGGATTTCGTGAAAGCTTTAC
sspyrE rev	ATACGAAGCGCTATTATTCTAGCTTTTTCCAATG TTTTTC
JRset1 fwd	ACGCCCTTAAATAAGGTTAG
JRset1 rev	GGGACATTGAAAGAACTAGA
TRANSFORMING	
PE *Ase*I+	AACTCCCAATATCATATCAAAATTAATACCTTTTAC TATGTCGACTGCAGAACTAACGAC
PE *Ase*I−	AACTCCCAATATCATATCAAAATTTATACCTTTTAC TATGTCGACTGCAGAACTAACGAC
LCR PROBES	
LCRSaPE90dntop	GATTTAAGGAAACTCCC
LCRSaPE90dnbot	AGGTAATAGGGGCTAAC
LCRSaPE90Atop	GTTAGCCCCTATTACCT**A**
LCRSaPE90Tbot	GGGAGTTTCCTTAAATC**T**
LCRSaPE90Ttop	GTTAGCCCCTATTACC**T**
LCRSaPE90Abot	GGGAGTTTCCTTAAATC**A**
LCRSaPE117dntop	GACATATTTTCATTCGTCG
LCRSaPE117dnbot	GGATGGTTGGGGAG
LCRSaPE117Ttop	CTCCCCAACCATCC**T**
LCRSaPE117Abot	CGACGAATGAAAATATGTC**A**
LCRSaPE117Atop	CTCCCCAACCATCC**A**
LCRSaPE117Tbot	CGACGAATGAAAATATGTC**T**
LCRSaPE192dntop	GTTACTGGGGGCG
LCRSaPE192dnbot	ACTCCCAATATCATATCA
LCRSaPE192Gtop	GATATGATATTGGGAGT**G**
LCRSaPE192Cbot	GCCCCCAGTAAC**C**
LCRSaPE192Ttop	GATATGATATTGGGAGT**T**
LCRSaPE192Abot	GCCCCCAGTAAC**A**

The bases in bold are the query bases for LCR probes.

the 5′-overhang strand, because the strand with 3′-overhang is not a substrate for the exonuclease (Henikoff, 1984). Any residual double-stranded amplicons were eliminated by digestion with *Sal*I, which cuts 5 nt away from the 5′-end of the selectable marker on the sense strand and 1 nt away from the 3′-end of the selectable marker on the anti-sense strand. The double-stranded amplicon cut with *Sal*I did not yield transformants, confirming that this treatment was effective in eliminating ds DNA.

Oligonucleotides were synthesized and de-salted by Integrated DNA technologies (IDT), IA, USA and used without further purification, as described for our previous studies (Grogan and Stengel, 2008; Grogan and Rockwood, 2010). Synthetic donor DNAs "markedSaPE43-227top" and "markedSaPE43-227bot" (**Table 2**) were designed with regularly spaced non-selected markers as for the pyrEv3 cassette (Grogan and Rockwood, 2010), except that only a 185-bp interval (*pyrE* nt 43 to 227) was represented. These two ssDNAs were synthesized as high-quality "Ultramers", and mass spectra provided by the manufacturer indicate that full-length product represented the bulk of both preparations. As in previous studies, no extraneous mutations (i.e., not designed in the oligonucleotides) were detected in any recombinant.

TRANSFORMATION

S. acidocaldarius was electroporated in a 1-mm cuvette with a single pulse of 1.25 kV at 2 μF and 1 kΩ. Pre- and post-electroporation treatments were those described previously (Grogan and Stengel, 2008). The amount of *pyrE*v3 DNA was 60–100 ng, and the amount of 185-nt oligo was 10 and 20 pmol. Electroporations were considered saturating with respect to input DNA, as indicated by a slight decrease in transformation efficiency at 20 pmol compared to 10 pmol.

ANALYSIS OF TRANSFORMANTS

Up to 24 transformant colonies were picked from independent electroporations of the DNA preparation, drawn mostly from 10-pmol transformations. (A few colonies from 20-pmol DNA transformations were included to determine whether frequency of sectoring varied with amount of DNA, and no such effect was seen.) In initial studies, isolated *S. acidocaldarius* Pyr^+ transformant colonies were suspended in sterile buffer and serially diluted. Dilutions (10^{-4}, 10^{-5}, and 10^{-6}) were plated on selective plates and incubated for 5 days. Eight isolated colonies of sibling cells from each transformant colony were picked and grown in selective liquid medium for about 2 days. Genomic DNA was extracted and a 995-bp fragment containing the *pyrE* loci was amplified using primers JRset1fwd and JRset1rev (**Table 1**). Most of the BPSs introduced into the donor DNAs either create or destroy restriction enzyme sites relative to the wild-type sequence, enabling scoring by restriction enzyme digestions and fragment analysis on agarose gels (Grogan and Rockwood, 2010). Restriction-digestion screening using four to five markers equally spaced within the *pyrE* gene were carried out to identify sibling strains that differed in digestion patterns. Potential cases of sectoring identified by restriction screening were then scored for all polymorphic sites by Sanger sequencing of the PCR amplicons.

In a later streamlined procedure, the original (non-purified) transformant colonies were picked and grown in 3 ml selective medium. Genomic DNA was extracted from 2 ml of culture and *pyrE* was amplified as described above, while the remaining cells were stored at room temperature until needed for plating (see below). A restriction-digestion screen was carried out on the amplicons to identify samples that showed only partial cleavage, indicating both endonuclease-sensitive and -resistant amplicons in the sample. Apparently dual-genotype cultures were then serially diluted and plated to obtain isolated colonies as described above. The markers in these "sibling colonies" were then scored as described above to confirm distinct clones, and representatives of these sibling clones were then scored at all remaining sites by Sanger sequencing. The two approaches of identifying sectored colonies (i) isolating and analyzing sibling colonies initially from each transformant, vs. (ii) screening non-purified transformant colonies followed by confirmation of purified sibling colonies, yielded comparable efficiencies in detecting sectored colonies from cells transformed with the pyrEv3 cassette. Method (ii) was therefore employed for most of the study, owing to its greater efficiency, to simplify and increase our capacity to screen more samples; only potentially sectored colonies were subjected to confirmatory tests in purified sibling colonies.

Table 2 | Markers in 185-nt transforming DNAs.

WT *pyrE* nt position	Recipient (top strand)	MarkedSaPE43-227top	MarkedSaPE43-227bot	Restriction scoring[a] (top)	Restriction scoring[a] (bottom)
63	A	C	C	*Bsu*36I	*Hpy*CH4IV
90	A	G	A	*Bst*NI	LCR
117	T	A	C	LCR	*Bsp*EI
144	T	C	C	*Ssi*I	*Pst*I
163, 165, 168	G, T, T	T, C, A	A, C, T	*Nsi*I (selectable)	*Bsp*DI (selectable)
192	T	G	G	LCR	*Hin*fI
207	T	A	G	*Rsa*I	*Bsa*HI

[a]*The indicated restriction site is created by the corresponding top or bottom-strand marker. LCR indicates markers scored by ligase chain reaction.*

SCORING NON-SELECTED MARKERS

Most of the BPSs introduced into the native *pyrE* sequence in the various donor DNAs either create or destroy restriction sites, and were scored in screening procedures by restriction enzyme digestions of PCR amplicons and fragment analysis on agarose gels, as described above. For markers derived from pSaPyrEv3, final scoring was by Sanger sequencing of the amplicons. For oligonucleotide transformants, most markers were scored by restriction. The endonuclease *TsoI* proved to be unreliable in scoring the marker incorporated at *pyrE* nt 192, however, so ligase chain reaction (LCR; Wiedmann et al., 1994) was used to score this marker and two other donor alleles that did not alter restriction sites (**Table 1**). The LCR mixture consisted of approximately 10 ng *pyrE* DNA, 2 μM of the four probes (**Table 1**), 1× SYBR Green, 1 × 9°N™ligase buffer and 5U of 9°N™(*Thermococcus*) DNA ligase (New England Biolabs), in a total volume of 20 μl. After an initial denaturation at 95°C for 3 min, the following two-step cycle was used: denaturation at 95°C for 30 s and annealing and ligation at 59.3°C, for 39 cycles. This was followed by a final annealing and ligation at 59.3°C for 3 min. A melting curve of the reaction mixture was obtained from 60 to 95°C by plotting decrease in fluorescence per increment of 0.5°C, and ligation product was identified as a peak at intermediate T_m in the first-derivative curve.

RESULTS

GENETIC EVIDENCE OF HETERODUPLEX SEGREGATION

To investigate whether heteroduplex DNA formed by HR can be detected in *Sulfolobus* spp., we electroporated strain MR31 with a *pyrE* DNA marked by a set of phenotypically silent mutations (Grogan and Rockwood, 2010). Each of the resulting Pyr^+ transformant colonies was then re-suspended in liquid medium, diluted, and re-plated on selective medium to obtain multiple isolated colonies. Four to eight such sub-clones were then analyzed with respect to non-selected (silent) *pyrE* markers, and apparent cases of two genotypes, indicated by differences found at one or more sites, were confirmed by analysis of all the markers in each of the two distinct daughter lineages (see Materials and Methods). About one-fourth (11/42) of the transformant colonies examined revealed different alleles (base pairs) at one or more marked sites. Similar to previous observations on clonally purified transformants (Grogan and Rockwood, 2010), replacement tracts ranged in length from the entire donor segment down to four markers (**Figure 2**). Discontinuity was also observed, as typified by two replacement tracts separated by a block of recipient markers (**Figure 2**).

The recombinants defined several categories of transfer patterns. For example, 9 out of 12 dual-genotype colonies analyzed were pairs in which each sibling incorporated one tract of donor markers, and the resulting tracts shared one endpoint (**Figure 2**). In four of the colonies, one of the siblings had two separate tracts of donor markers, and in three of these, the second donor tract consisted of a single marker (dsPE1, 4, 6). Other configurations included a single tract in each sibling that differed at both ends (dsPE12).

To confirm that the two-genotype colonies derived from heteroduplex DNA, we first considered whether they could result from two independent recombination events within the same cell. In principle, two DNAs could enter the same recipient cell and recombine separately with two completed copies of the chromosome, for example (Bernander and Poplawski, 1997), or with different arms of a replicating chromosome. In one set of experiments, two distinctly marked 60-nt DNAs which represent the same strand and interval of the *pyrE* gene (PE *Ase*I+ and PE *Ase*I−; **Table 1**), were combined to yield an equimolar mixture which was electroporated into recipient cells. A total of 55 Pyr^+ transformants were analyzed as unpurified colonies. The selected marker containing an *Ase*I site was recovered in 25 transformants and the alternative allele, without the *Ase*I site, was found in 30. No transformant colony was observed to retain both of these alleles, indicating that most transformants under our conditions reflect incorporation of a single DNA molecule. This implied either that only one DNA enters most cells under these conditions, or that most cells have only one site competent for incorporating exogenous DNA at any given time. In addition, we noted that most of the dual-genotype transformant colonies exhibited a common donor-tract endpoint in *pyrE* (**Figure 2**), despite the eight possible left-hand endpoints and 21 possible right-hand endpoints available in this assay. The over-representation of shared endpoints thus argued that the two genotypes originated from a common recombination intermediate.

We also considered whether two-genotype colonies could have resulted from diversification of an initial recombinant cell through a second round of recombination by conjugation with a non-transformed cell on the selective plate. The efficiency of the DNA transfer step in *S. acidocaldarius* has not been measured

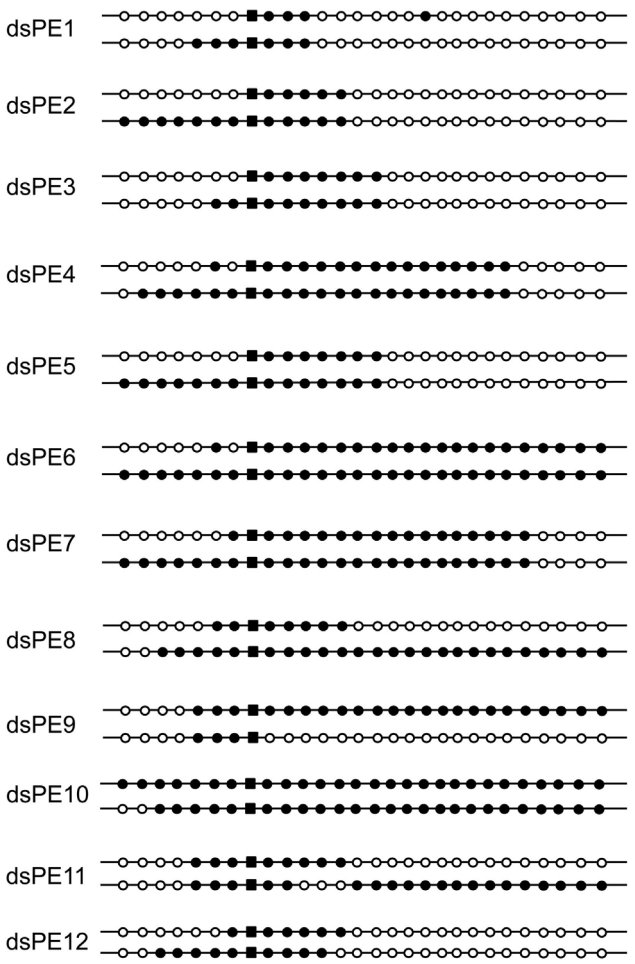


FIGURE 2 | Dual-genotype transformants. Diagrams show the marker (donor vs. recipient allele) at each of 28 sites distinguishing the input (donor) *pyrE* sequence from the chromosomal gene (marker positions are not drawn to scale). For details on the marked sites, see Grogan and Rockwood (2010). The donor allele is represented by a filled dot, the recipient allele is an open circle, and the selected marker is represented by a filled square. Each pair of genotypes was recovered from a colony of cells on selective medium that were transformed to the Pyr^+ phenotype by double-stranded pyrEv3 DNA (see Materials and Methods).

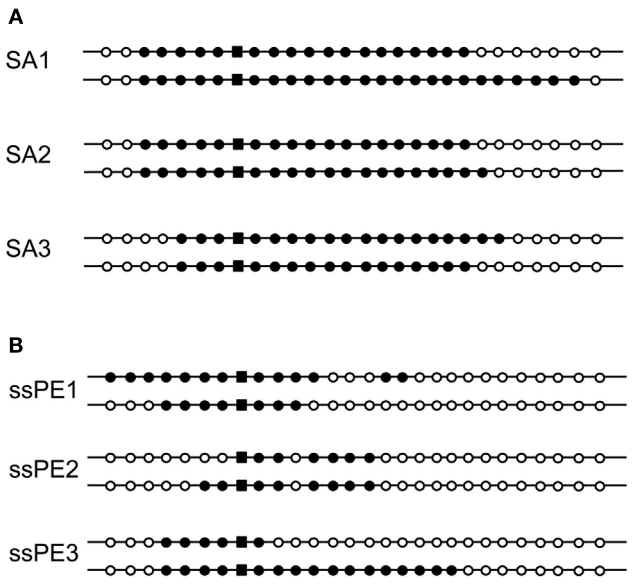


FIGURE 3 | Effects of blocking conjugation or transforming with single-stranded DNA. The genotypes of dual-genotype transformants are shown as for **Figure 2**. **(A)** Strain SA1 (conjugation-deficient) transformed with the double-stranded pyrEv3 cassette. **(B)** Strain MR31 transformed with single-stranded pyrEv3 DNA (see Materials and Methods).

experimentally and may not be especially high, since recombinants rarely form at frequencies above 10^{-4} (Grogan, 1996). However, in contrast to the conjugation system of *S. solfataricus*, the *S. acidocaldarius* system functions constitutively (Ajon et al., 2011) and thus cannot be discounted on the basis that the recipient cells were not treated with DNA-damaging agents under our conditions. We therefore tested the dependence of genetic sectoring on conjugation, by performing the same experiments with the conjugation-defective Pil^- strain SA1 (Ajon et al., 2011). Dual-genotype colonies represented three of 19 colonies examined, and their patterns of markers resembled those found in the isogenic Pil^+ recipient MR31, i.e., uninterrupted tracts of donor markers differing at one end but not the other (**Figure 3A**). This result demonstrated that the process leading to the dual-genotype colonies was fully functional in the Δ*upsE* strain, and, therefore, that conjugation is not necessary to produce this class of transformants.

In addition, the hypothesis of conjugation after recombination predicts that, on average, the genotype with the fewer donor markers should be the one generated by conjugation. This reflects the fact that the second recombination event (the one proposed to be initiated by conjugation) would sometimes occur after an additional round of genome replication and thus affect one-fourth or fewer of the clones representing the original HR event. We therefore examined the relative abundance of the genotypes in sectored colonies. Out of six colonies identified using method (i) (see Materials and Methods), five showed greater abundance of genotypes with fewer donor markers and one colony had equal numbers of both, contrary to the prediction of diversification through conjugation. Thus, in all the various control experiments, properties of the dual-genotype colonies remained consistent with segregation of two distinct products of a single HR event.

GENETIC SECTORING FROM ssDNA

In the absence of mutants lacking specific, or even putative, HR functions, we probed mechanistic features of the HR process by altering the donor DNA. We first tested the ability of ssDNA to generate genetic sectoring. Multiply marked *pyrE* amplicons were digested with exonuclease to remove the top strand; any residual duplex DNA in the preparation was then inactivated by restriction endonuclease before the DNA was electroporated into cells (see Materials and Methods). At least 7% of the resulting transformants were sectored (3/43), and the differences defined by the sibling pairs appeared to be more extensive than those produced by dsDNA (**Figure 3B**, compare to **Figure 2**).

Based on this result from enzymatically produced ssDNA, we designed a new assay that used long synthetic oligonucleotides as multiply marked but chemically defined donor. In this scheme, two oligonucleotides represent the top and bottom strands, respectively, of the same 185-bp interval of the functional *S. acidocaldarius pyrE* gene and incorporated silent BPSs within the selected region and at regular intervals on both sides (**Table 2**). When electroporated separately into MR31 cells, the two oligonucleotides showed no obvious strand bias (averages of 2.6 vs. 1.8 Pyr^+ transformants per pmol for top and bottom strand, respectively). In addition, both top- (T-) and bottom- (B-) strand transformants included sectored colonies, indicating that these short ssDNAs form extensive heteroduplex structures with the *S. acidocaldarius* chromosome which can segregate two distinct genotypes (**Figure 4**). The six non-selected, phenotypically silent, markers in each oligonucleotide (**Table 2**), define $2^6 = 64$ possible genotypes for the transformants in each case. We found 13 of these genotypes among 48 T-strand transformants, and 15 among 48 B-strand transformants. **Figure 5** compares the two or three most common of the genotypes recovered under these and other conditions.

In the *pyrE* gene of the transformants, the frequency of the donor allele was 100% at the site of selected marker, and decreased with increasing distance in both directions (**Figure 6A**). The resulting gradients differed somewhat between T-strand and B-strand

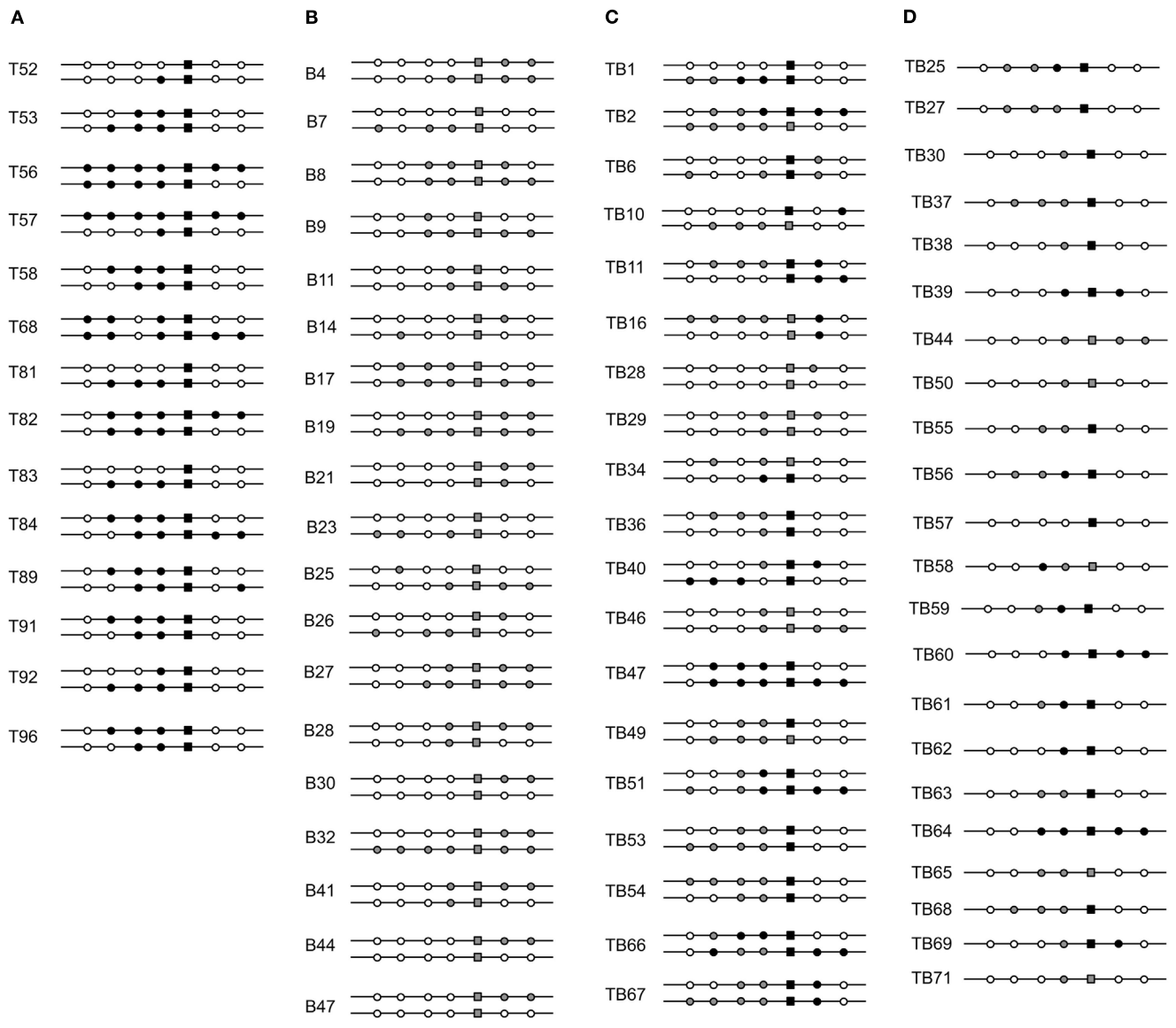


FIGURE 4 | Transformation by differentially marked oligonucleotides. Two synthetic DNAs (**Table 2**), representing the top strand (T) and bottom strand (B) of base pairs 43–227 of *pyrE*, were electroporated individually **(A,B)** or as a duplex **(C,D)**. Markers are given in **Table 2**; the transformants were selected and analyzed as for **Figure 2**. **(C)** Shows pairs of genotypes from sectored colonies, whereas **(D)** shows the single genotypes of non-sectored colonies recovered in the same experiment. Symbols: top-strand donor allele, black; bottom-strand donor allele, gray; recipient allele, white.

DNA Type	Genotype	Percent of total recombinants
Top strand		35.7
		14.3
Bottom strand		18.4
		13.2
		13.2
Duplex (sectored)		5.3
		5.3
Duplex (nonsectored)		13.6
		9.1
		9.1

FIGURE 5 | Common genotypes formed by differentially marked oligonucleotides. The two or three most frequent genotypes (data of **Figure 4**) are indicated with their frequency among all recombinants. It should be noted that the number of potential genotypes is much higher for the duplex than for the individual strands of donor DNA, resulting in lower absolute frequencies of the most frequent genotypes. Symbols follow the scheme of **Figure 4**.

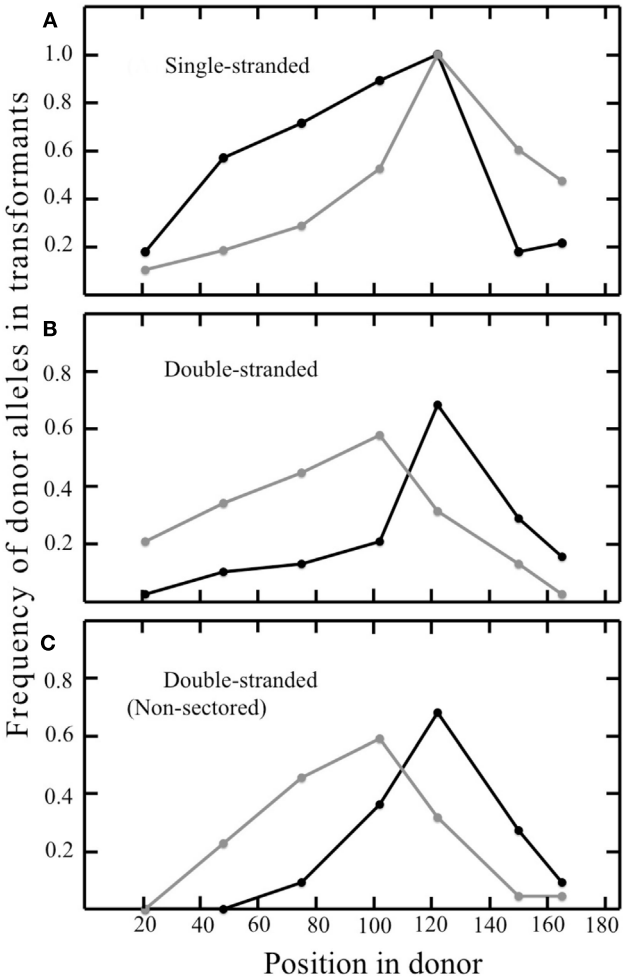


FIGURE 6 | Frequency of strand-specific donor marker incorporation. Graphs show the proportion of recombinants retaining the indicated donor marker at the corresponding position in the synthetic *pyrE* DNA described in **Table 2**. Panels and symbols correspond to **Figure 4**. **(A)** Sectored transformants from ssDNAs; **(B)** sectored transformants from dsDNA; **(C)** non-sectored transformants from dsDNA.

transformants however, and the pattern suggested preferential loss of the 3′ portion of both oligonucleotides (**Figure 6A**). Each transformant contained a continuous block of donor alleles that includes, at a minimum, the selected marker, and often additional, silent markers from the donor; we term this feature the selected donor tract (Grogan and Rockwood, 2010). In most cases, the selected tracts of the two siblings shared one end in common; they differed at both ends in only 8% (1/12) of sectored T-strand colonies, and 16% (3/19) of sectored B-strand colonies.

PROCESSING OF PRE-FORMED HETERODUPLEX

Both the frequent discontinuity of the recombination tracts and the genetic sectoring implied a form of HR that often separates markers that are very close together, i.e., within 20–30 nt of each other. This pattern would seem difficult to achieve by reciprocal strand exchanges but relatively straightforward by nucleolytic removal of short intervals around mismatches, in a manner resembling excision of the damaged strand during nucleotide excision repair (NER). In this latter scheme, the resulting small gaps would be re-filled by DNA synthesis, thereby copying any allele in the gap to the opposite strand.

To investigate removal of mismatches from a defined heteroduplex, we took advantage of the fact that the two ssDNAs used in the previous study were differentially marked, so that the marker contributed by the top strand could be distinguished from its bottom-strand counterpart at all positions, including the selected region. Equimolar amounts of the two 185-nt strands were therefore annealed *in vitro* to produce the corresponding duplex with seven regularly spaced mismatches. MR31 cells were electroporated with the resulting duplex, and the *pyr*$^+$ transformants were analyzed by PCR of the *pyrE* locus and scoring of the seven markers. The efficiency of transformation by the linear duplex (average of 1 Pyr$^+$ per pmol DNA) was slightly below the efficiencies for each strand separately (see above), while the proportion of sectored transformants exhibiting sectoring was similar [26% (19 of 72 screened) vs. 29 and 31% for T and B strands, respectively].

The sectored transformants were unexpectedly complex. Four of the 19 had acquired markers from only one donor, whereas another four differed with respect to the selected marker, i.e., one sibling had the T-strand allele the other had the B-strand allele (see **Figure 4**, colonies TB2, TB10, TB34, TB49). These latter recombinants preserve the original duplex configuration at the selected site, but their frequency was not high: 36% (4/11) among all sectored colonies retaining markers from both donor strands, compared to 22% predicted by random association. This suggests that most preformed heteroduplex DNAs experienced considerable processing

during HR that discarded and resynthesized much of the original duplex.

The sibling pairs that retained both T- and B-strand markers showed a bias in positioning, with B-strand tracts generally extending farther in the left-hand portion, and T-strand tracts extending farther in the right-hand portion. Examining the frequency of T- vs. B-strand markers across all the sectored recombinants confirmed a similar pattern, i.e., lower frequencies of T-strand markers at the left end of the 185-bp interval and lower frequencies of B-strand markers at the right end (**Figures 6B,C**). This suggested preferential loss of 5′ ends, which is the reverse of the donor vs. recipient bias seen when the same oligonucleotides were electroporated individually (**Figure 6A**). Thus, although the HR system of transfers markers from both single-stranded and double-stranded DNAs to the *S. acidocaldarius* genome, processing of the two forms differs in basic respects. We also observed a number of individual clones that incorporated both T- and B-strand donor markers; in each of these cases the transition from one donor source to the other was always immediate, i.e., never interrupted by a recipient marker (**Figures 4C,D**).

Finally, we investigated whether these patterns of marker transfer from the dually marked duplex to the *S. acidocaldarius* chromosome were a general feature of HR under these conditions, or alternatively, associated specifically with genetic sectoring. To test this, we analyzed non-sectored Pyr^+ transformants from the experiments described above with respect to the T- and B-strand markers. The recombinants appeared qualitatively similar to those recovered from sectored colonies (**Figures 4C,D**), which was also supported by quantitative comparisons. For example, sectored and non-sectored transformants incorporated statistically indistinguishable numbers of top- and bottom-strand markers ($P = 0.778$ for sectored and 0.299 for non-sectored, by two-tailed T-test). They also had similar proportions of genotypes incorporating both donor markers (55.3 vs. 54.6%), and similar positions where the transition between the two occurred (*pyrE* co-ordinate 149 vs. 146). The T-strand allele provided the selected marker in 68.4% of sectored transformants and 68.2% of non-sectored ones, and the two sets of recombinants showed similar profiles of marker incorporation as a function of position within the marked interval (**Figures 6B,C**). Thus, we found no quantitative evidence that genetic sectoring involves processes distinct from those that generate a single recombinant genotype in cells transformed under these conditions.

DISCUSSION

In the present study, genetic assays applied to a thermoacidophilic crenarchaeote demonstrate for the first time that an archaeon can assimilate three forms of linear DNA (positive-strand, negative-strand, and duplex) into pairs of related recombinant genotypes that represent alternative products of recombination events. The specific features of these sibling pairs, and of the other, non-sectored, recombinants, further indicate that sporadic strand-removal events, many of them highly localized, resolve mismatches in heteroduplex DNA. The apparently stochastic, combinatorial nature of these events distinguish HR in *S. acidocaldarius* from the dominant modes of HR in eukaryotes and bacteria, and has significant implications for genome stability in *Sulfolobus* species and other hyperthermophilic archaea.

INCORPORATION OF LINEAR dsDNA INTO THE *S. ACIDOCALDARIUS* CHROMOSOME

Transformation by linear DNA involves recombination in an "ends-out" orientation, and at least two pathways have been proposed for "ends-out" HR in eukaryotes and bacteria, which differ with respect to the predicted segregation of markers flanking the selected marker or gene. The first pathway, sometimes termed "double-crossover," involves strand exchange, and represents the initial steps of classical DSB/gap repair oriented outward instead of inward: both 5′ ends of the linear duplex are resected to expose 3′ tails, which find their complements in the intact recipient DNA (**Figure 7A**), and invade the duplex at those sites to generate two displacement (D-) loops (Langston and Symington, 2004). The D-loop intermediate has hDNA flanking the selected marker; sequence divergence of these end regions creates a distinctive configuration of mismatches, defined by the original 3′ tails. This configuration puts input (donor) DNA markers to the right-hand side in one strand and to the left in the other (**Figure 7A**). If any of these mismatches segregate without being repaired, one daughter cell inherits the donor markers only on the right side of the selected central segment, while the other daughter cell inherits donor markers only on the left side; this arrangement is called the "*trans*" configuration. The alternative, "strand-assimilation," pathway (**Figure 7B**) has been proposed for *E. coli* expressing the λ *red* genes (Maresca et al., 2010) and for yeast (Leung et al., 1997), and involves one of the two input strands forming uninterrupted hDNA across the entire region (i.e., the selected region and both flanks), which keeps the flanking markers in a "*cis*" configuration (i.e., on both sides of the selected gene or marker; **Figure 7B**).

By genotyping sectored Pyr^+ transformant colonies, we could determine the configuration of flanking markers in hDNA that had segregated to yield two distinct selectable *Sulfolobus* recombinants. The most common configuration lacked the basic, bilateral distribution predicted for ends-out (gene-replacement) HR by both the double-crossover and strand-assimilation models, in that markers segregated on only one side of the central selected region.

Another test of conformity to existing models of "ends-out" HR was provided by the distinct genetic markers in the top and bottom strands of the 185-bp duplex. This short dsDNA documented the original source of the donor markers in all recombinants, including those that did not exhibit sectoring. Both published models of "ends-out" HR predict recombinants that, when considered individually, retain donor markers from either the top or bottom strands of the input duplex, not both (**Figures 7A,B**). As a test of symmetry, we calculated the net excess of T over B alleles at each marked site for each of the 38 recombinants from sectored colonies and 22 from non-sectored. In both groups (**Figure 7C**), this measure of marker bias did not give the clearly bimodal distribution predicted by early strand separation (**Figure 7A**) or complete discarding of one strand before annealing (**Figure 7B**), even if the choice of strand varied among transformants.

RESOLUTION OF HETERODUPLEX FORMED BY ssDNA

The genetic sectoring we observed in transformation by ssDNA implies that (i) in the course of HR, the selected marker was

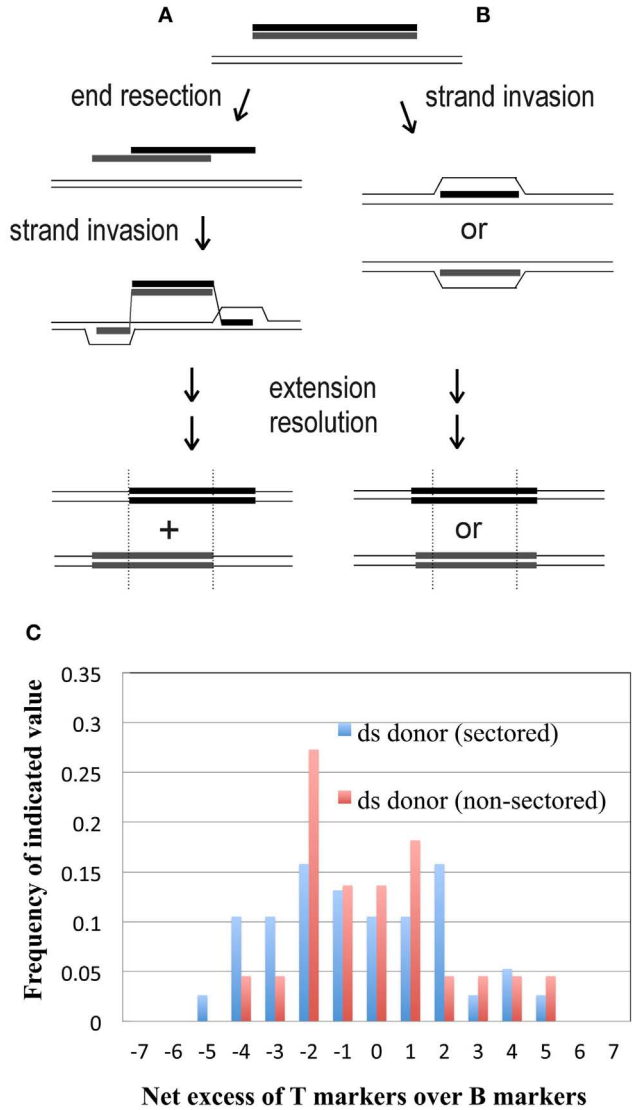


FIGURE 7 | Alternative models of "ends-out" recombination. Essential features of two alternatives proposed to explain integration of a linear DNA sequence into a cellular genome by HR. **(A)** "Double-crossover" (strand exchange) model (see Langston and Symington, 2004); **(B)** "strand-assimilation" model (see Leung et al., 1997); **(C)** histogram showing the distribution of the net excess of T over B markers among sectored and non-sectored (simple) recombinants from the differentially marked dsDNA.

retained in the donor strand and also copied to the recipient strand, and (ii) at least one non-selected marker escaped the copying process (depicted schematically in **Figure 1C**). In our assay, all the non-selected markers are BPS; thus, each sectored transformant generated by ssDNA represents at least one mismatch that was segregated, presumably by replication, without being repaired. However, in only seven of 33 sectored ssDNA transformants was this unrepaired mismatch not accompanied by conversion of at least one other non-selected marker. It seems likely, therefore, that many of the non-sectored ssDNA transformants represent cases in which blocks of markers were copied from donor to chromosomal strand of a heteroduplex before it was replicated, corresponding in **Figure 1** to outcome B rather than outcome A.

Conversely, other features of ssDNA transformants argue that mismatches of a heteroduplex intermediate can be repaired individually, not as part of a large block. Specifically, isolated recipient markers (donor-tract interruptions) were generated by both full-length and 185 nt single-stranded forms of *pyrE* (**Figures 2** and **4**), which cannot be explained without invoking resolution of mismatches. Our data do not address whether the mismatches triggered the postulated excision events or otherwise influenced their position, but they do provide a test for directionality of the resolution, i.e., whether the donor or the recipient marker was preferentially copied to the opposite strand. This test is provided by the relative abundance of isolated donor markers vs. isolated recipient markers in the transformants, representing resolution of a single mismatch that preserves the donor allele vs. the recipient allele. Among the events in T-strand transformants, one retained the donor allele and three retained the recipient allele. Among B-strand transformants, five retained the donor allele and six retained the recipient allele (the selected marker was excluded from this analysis). Since these ratios are not statistically different from 1, they indicate no directional bias with respect to the discontinuous (donor) strand.

The donor strand in these experiments was an ssDNA 185 nt or shorter annealed to the chromosome; it can thus be considered a plausible mimic of a *Sulfolobus* Okazaki fragment (Beattie and Bell, 2012), and accordingly, a potential substrate of any hypothetical MutSL-independent post-replication error-correction system that *S. acidocaldarius* may have. Such a system operating on this type of hDNA would be expected to transfer the recipient allele to the annealed oligonucleotide nearly exclusively. The fact that we observed frequent mismatch resolution but no strand discrimination thus fails to support the hypothesis that *S. acidocaldarius* has a generalized system of post-replicational error correction. Alternatively, if the mispair resolution we observed does involve components of such a system, the strand discrimination function of that system must require a specific context that was not provided by our experimental conditions.

We also note that patterns of marker transfer similar to those we observe in *S. acidocaldarius* have been reported for MMR-deficient strains of eukaryotic micro-organisms. Numerous short recombination tracts, including those transferring a single base-pair change, were observed following intragenic HR between diverged sequences in yeast (Coic et al., 2000) and trypanosomes (Barnes and McCulloch, 2007). The behavior appeared only in *msh2* mutants, however, indicating that it did not involve MutSL homologs, and in fact was normally masked by their activity. In the yeast assay, the short-patch events were frequently accompanied by segregation of hDNA (Coic et al., 2000), whereas no such segregation was seen in the trypanosomes, for unknown reasons (Barnes and McCulloch, 2007). The similarities between these recombination results and our own thus reinforce the hypothesis that the observed short-patch mode of recombination observed in *S. acidocaldarius* involves localized resolution of mispairs in heteroduplex DNA without the influence of a post-replicational error-correcting system.

In addition to revealing apparently erratic resolution of mismatches in *S. acidocaldarius* during HR, both the top- and bottom-strand oligonucleotides exhibited preferential loss of the 3′-end (i.e., right-side markers for T-strand transformants and left-side markers for B-strand transformants; **Figure 6A**). However, when the same two strands were introduced as a duplex, the 5′-end markers of each strand were lost preferentially (**Figures 6B,C**), demonstrating that dsDNAs and ssDNAs are processed differently during HR. Resection of 5′ ends of duplex DNA is a critical feature of HR in eukaryotes and bacteria, and may play an important role *S. acidocaldarius* HR, as well. However, the extent of resection implied by donor markers in the sectored *S. acidocaldarius* recombinants varied erratically at the two ends of the marked interval. In none of the 19 sectored colonies analyzed did apparent 5′-end resection remove the same number of markers from each end, and in only three cases did it remove at least one marker from each end.

IMPLICATIONS FOR HR IN *S. ACIDOCALDARIUS*

These various functional properties are not typical for HR in model micro-organisms, and may reflect a set of recombination mechanisms peculiar to *S. acidocaldarius*, *Sulfolobus* species, or hyperthermophilic Archaea as a group. As a basis for designing future studies that could clarify these questions, we offer four functional interpretations that seem mechanistically consistent with our results.

1. *S. acidocaldarius processes heteroduplex intermediates of HR by removing single-stranded intervals encompassing individual mispairs or short tracts of mispairs and re-filling the resulting gaps.* The intervals are neither especially uniform in size, nor directed with respect to the discontinuous strand of the heteroduplex.
2. *S. acidocaldarius incorporates homologous ssDNA into the chromosome to form an extensive heteroduplex.* The region of hDNA can be hundreds of bp long, as indicated by the number of markers that can be transferred by one ssDNA. The process that incorporates the ssDNA involves preferential loss of the 3′ end, but remains otherwise undefined.
3. *Short-patch conversion can resolve a mispair in any hDNA present in S. acidocaldarius cells at any stage of the HR process.* When conversion copies the recipient allele onto the donor side of the heteroduplex, it will often interrupt a string of donor markers, and when it copies the donor allele to the other strand, it will often generate a lone donor marker. Such isolated recipient and donor markers represent the most common form of non-selected marker tract in *S. acidocaldarius* recombinants (Grogan and Rockwood, 2010). When the transforming DNA has two differentially marked strands, we observe that most of the recombinants inherit donor markers from both strands, and these disparate donor alleles are never separated by recipient alleles. Among sectored transformants, only a minority indicates segregation of the two original donor alleles (positive- vs. negative-strand, or in our terminology, T vs. B) at any position.

   Taken together, these properties suggest that localized conversion events copy most of the surviving donor markers from one donor strand to the other before the original strands segregate. These conversion events in pre-formed hDNA are analogous to those proposed to generate discontinuous recombination tracts from ssDNA, but occur within the context of a T:B donor heteroduplex, rather than a donor:recipient heteroduplex.
4. *Un-coordinated excision events near the ends of a linear duplex contribute to net resection of 5′ ends.* Although the pattern of end loss in duplex DNA was biased in favor of 5′ ends, very few sectored transformants indicate the bilateral 5′-end resection of classical HR pathways. In most cases, recombination tracts differed between the two siblings only at one end, and these segregating intervals varied considerably in length. When the two donor strands were distinguished from each other by markers, sectored colonies similarly showed little bilateral resection; a few indicated 5′ resection at each end, but one was mixed (5′ and 3′), one suggested bilateral 3′-end resection, and none had equal numbers of markers removed from the two ends, despite the similar spacing of the markers (**Figure 4C**).

This lack of uniformity or correlation between the two ends could reflect erratic processing by the HerA helicase and NurA nuclease of *S. acidocaldarius*, which have been implicated in double-strand break end resection (Constantinesco et al., 2004; Hopkins and Paull, 2008; Quaiser et al., 2008). We note that it also could result, in principle, from the same intrinsically stochastic short-patch excision events proposed to resolve mispairs in hDNA. These events should form extensions (overhangs) of both polarities when they occur very close to one end of the duplex. The 5′ overhangs would be prone to strand re-synthesis primed by the recessed 3′ ends, but no such priming is available for a 3′ extension created by loss of the complementary 5′ end. As a result, 3′ extensions would tend to persist and accumulate. Regardless of how they are formed, 3′-extensions would be expected to promote integration of the donor duplex into the recipient chromosome. However, the "random-excision" pathway we hypothesize predicts that most dsDNAs will have experienced conversion events by the time they integrate into the recipient chromosome. This seems consistent with the high density of T/B conversion events evident in transformants from the dually marked donor duplex (**Figure 4C**).

Although the four features proposed above leave a number of important steps undefined, including those that form and stabilize heteroduplex DNA, they provide a context for formulating and testing models of HR involving linear DNA in *S. acidocaldarius* and other hyperthermophilic archaea. The schemes of **Figure 8** illustrate a few of the possibilities, but various alternatives also cannot be excluded. Short-patch excision events play central roles in these schemes, and the molecular details of those events also remain unclear, although we favor the idea of a process analogous to NER, in which each mispair promotes formation of a small bubble, one strand of which is excised by structure-specific endonucleases cutting at the boundaries.

Figure 8A depicts annealing of a ssDNA (with exclusion of its 3′ end) to a pre-existing gap, incorporation of the base-paired segment by ligation, conversion of one donor marker by a short-patch event, and transfer of a block of markers by a larger conversion event before replication. **Figure 8B** illustrates an alternative process in which ssDNA anneals to an exposed nascent-strand terminus at a regressed replication fork, followed

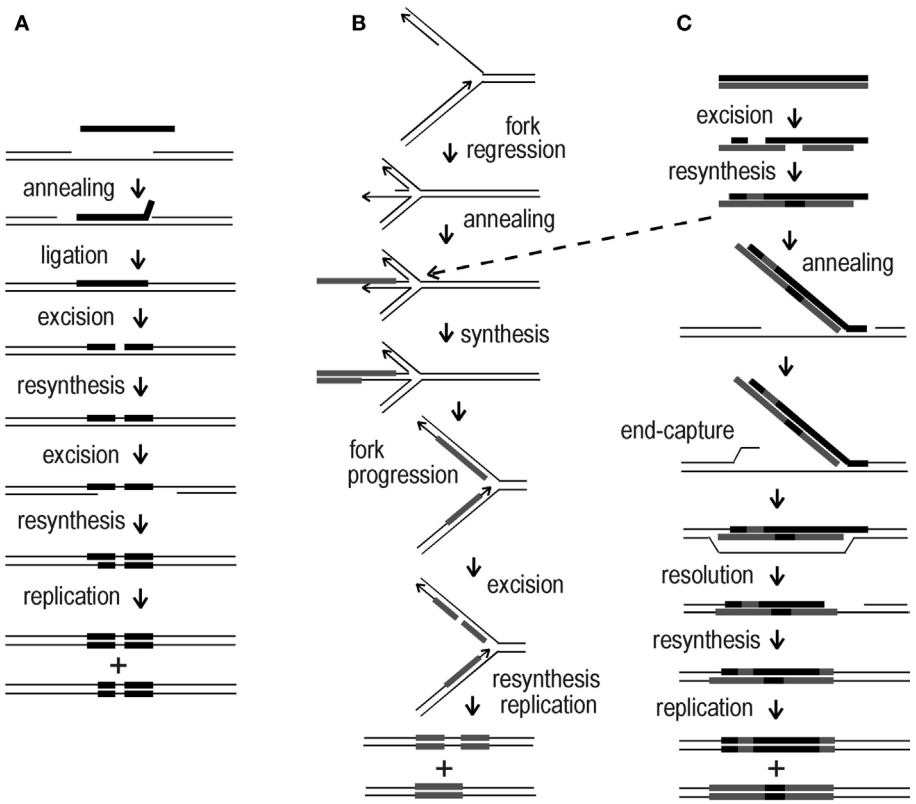


FIGURE 8 | Mechanistic alternatives for "ends-out" HR in *Sulfolobus*. (A) ssDNA annealing at a transient gap in the chromosome; **(B)** ssDNA annealing to a transiently regressed replication fork; **(C)** dsDNA incorporated at a gap in the chromosome, or alternatively, at a regressed replication fork (dotted arrow). Arrowheads indicate 3′ ends, black bars correspond to the top (sense) strand of the *pyrE* gene, gray bars indicate the bottom (anti-sense) strand.

by second-strand synthesis primed by the terminus, and resumption of replication; this creates two heteroduplex structures of differing length, which can diverge further through short-patch conversions. **Figure 8C** depicts dsDNA initially undergoing multiple short-patch excision events, which rearrange the markers and expose a 3′ extension. This single-strand extension then anneals to a chromosomal gap and anchors the remaining double-stranded portion until the opposite end can be captured by the recipient genome *via* some form of strand exchange. Another route to initial capture of the duplex is offered by a regressed replication fork, as indicated by the dotted arrow from scheme **Figures 8B,C**; in this case, subsequent formation and divergence of two heteroduplexes would follow, as shown in **Figure 8B**. The schemes, which have deliberately emphasized relatively simple transactions, seem to accommodate nearly all of the marker arrangements observed in recombinants, although we note that ssDNA transformants T11, B11, B25, and B26, and dsDNA transformants TB66 (all shown in **Figure 4**) imply additional, reciprocal forms of strand exchange not represented in **Figure 8**.

ACKNOWLEDGMENTS

We thank J. Rockwood for technical assistance and unpublished data. This work was supported by grant MCB0543910 from the National Science Foundation.

REFERENCES

Ajon, M., Frols, S., Van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A. J., Grogan, D. W., Albers, S. V., and Schleper, C. (2011). UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817.

Aylon, Y., and Kupiec, M. (2004). DSB repair: the yeast paradigm. *DNA Repair (Amst.)* 3, 797–815.

Barnes, R. L., and McCulloch, R. (2007). *Trypanosoma brucei* homologous recombination is dependent on substrate length and homology, though displays a differential dependence on mismatch repair as substrate length decreases. *Nucleic Acids Res.* 35, 3478–3493.

Beattie, T. R., and Bell, S. D. (2012). Coordination of multiple enzyme activities by a single PCNA in archaeal Okazaki fragment maturation. *EMBO J.* 31, 1556–1567.

Bernander, R., and Poplawski, A. (1997). Cell cycle characteristics of thermophilic archaea. *J. Bacteriol.* 179, 4963–4969.

Burgess, S. M. (2004). Homolog pairing in *S. pombe*: the ends are the means. *Mol. cell* 13, 766–768.

Chen, L., Brugger, K., Skovgaard, M., Redder, P., She, Q., Torarinsson, E., Greve, B., Awayez, M., Zibat, A., Klenk, H. P., and Garrett, R. A. (2005). The genome of *Sulfolobus acidocaldarius*, a model organism of the Crenarchaeota. *J. Bacteriol.* 187, 4992–4999.

Coic, E., Gluck, L., and Fabre, F. (2000). Evidence for short-patch mismatch repair in *Saccharomyces cerevisiae*. *EMBO J.* 19, 3408–3417.

Constantinesco, F., Forterre, P., Koonin, E. V., Aravind, L., and Elie, C. (2004). A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea. *Nucleic Acids Res.* 32, 1439–1447.

Grogan, D. W. (1996). Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 178, 3207–3211.

Grogan, D. W., Carver, G. T., and Drake, J. W. (2001). Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7928–7933.

Grogan, D. W., and Gunsalus, R. P. (1993). *Sulfolobus acidocaldarius* synthesizes UMP via a standard *de novo* pathway: results of a biochemical-genetic study. *J. Bacteriol.* 175, 1500–1507.

Grogan, D. W., and Rockwood, J. (2010). Discontinuity and limited linkage in the homologous recombination system of a hyperthermophilic archaeon. *J. Bacteriol.* 192, 4660–4668.

Grogan, D. W., and Stengel, K. R. (2008). Recombination of synthetic oligonucleotides with prokaryotic chromosomes: substrate requirements of the *Escherichia coli*/lambdaRed and *Sulfolobus acidocaldarius* recombination systems. *Mol. Microbiol.* 69, 1255–1265.

Hansen, J. E., Dill, A. C., and Grogan, D. W. (2005). conjugational genetic exchange in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*: intragenic recombination with minimal dependence on marker separation. *J. Bacteriol.* 187, 805–809.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28, 351–359.

Hopkins, B. B., and Paull, T. T. (2008). The *P. furiosus* mre11/rad50 complex promotes 5′ strand resection at a DNA double-strand break. *Cell* 135, 250–260.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* 58, 401–465.

Kreuzer, K. N. (2005). Interplay between DNA replication and recombination in prokaryotes. *Annu. Rev. Microbiol.* 59, 43–67.

Krogh, B. O., and Symington, L. S. (2004). Recombination proteins in yeast. *Annu. Rev. Genet.* 38, 233–271.

Langston, L. D., and Symington, L. S. (2004). Gene targeting in yeast is initiated by two independent strand invasions. *Proc. Natl. Acad. Sci. U.S.A.* 101, 15392–15397.

Leung, W., Malkova, A., and Haber, J. E. (1997). Gene targeting by linear duplex DNA frequently occurs by assimilation of a single strand that is subject to preferential mismatch correction. *Proc. Natl. Acad. Sci. U.S.A.* 94, 6851–6856.

Lin, E. A., Zhang, X. S., Levine, S. M., Gill, S. R., Falush, D., and Blaser, M. J. (2009). Natural transformation of helicobacter pylori involves the integration of short DNA fragments interrupted by gaps of variable size. *PLoS Pathog.* 5, e1000337. doi:10.1371/journal.ppat.1000337

Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A. F. (2010). Single-stranded heteroduplex intermediates in lambda Red homologous recombination. *BMC Mol. Biol.* 11, 54. doi:10.1186/1471-2199-11-54

Michel, B., Flores, M. J., Viguera, E., Grompone, G., Seigneur, M., and Bidnenko, V. (2001). Rescue of arrested replication forks by homologous recombination. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8181–8188.

Quaiser, A., Constantinesco, F., White, M. F., Forterre, P., and Elie, C. (2008). The Mre11 protein interacts with both Rad50 and the HerA bipolar helicase and is recruited to DNA following gamma irradiation in the archaeon *Sulfolobus acidocaldarius*. *BMC Mol. Biol.* 9, 25. doi:10.1186/1471-2199-9-25

Reilly, M. S., and Grogan, D. W. (2001). Characterization of intragenic recombination in a hyperthermophilic archaeon via conjugational DNA exchange. *J. Bacteriol.* 183, 2943–2946.

Sandler, S. J., Hugenholtz, P., Schleper, C., Delong, E. F., Pace, N. R., and Clark, A. J. (1999). Diversity of radA genes from cultured and uncultured archaea: comparative analysis of putative RadA proteins and their use as a phylogenetic marker. *J. Bacteriol.* 181, 907–915.

White, M. F., and Grogan, D. W. (2008). "DNA stability and repair," in *Thermophiles: Biology and Technology at High Temperatures*, eds F. T. Robb, G. Antranikian, D. W. Grogan, and A. J. Driessen (Boca Raton, FL: CRC Press), 179–188.

Wiedmann, M., Wilson, W. J., Czajka, J., Luo, J., Barany, F., and Batt, C. A. (1994). Ligase chain reaction (LCR) – overview and applications. *PCR Methods Appl.* 3, S51–S64.

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

Received: 06 March 2012; accepted: 11 May 2012; published online: 05 June 2012.

Citation: Mao D and Grogan DW (2012) Heteroduplex formation, mismatch resolution, and genetic sectoring during homologous recombination in the hyperthermophilic archaeon Sulfolobus acidocaldarius. Front. Microbio. ***3****:192. doi: 10.3389/fmicb.2012.00192*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 05 June 2012
doi: 10.3389/fmicb.2012.00200

A genetic study of SSV1, the prototypical fusellovirus

Eric Iverson and Kenneth Stedman*

Biology Department, Center for Life in Extreme Environments, Portland State University, Portland, OR, USA

Edited by:
Frank T. Robb, University of Maryland, USA

Reviewed by:
Brian P. Hedlund, University of Nevada Las Vegas, USA
Matthias Hess, Washington State University, USA

****Correspondence:***
Kenneth Stedman, Biology Department, Center for Life in Extreme Environments, Portland State University, P.O. Box 751, Portland, OR 97207-0751, USA.
e-mail: kstedman@pdx.edu

Viruses of thermophilic Archaea are unique in both their structures and genomic sequences. The most widespread and arguably best studied are the lemon-shaped fuselloviruses. The spindle-shaped virus morphology is unique to Archaea but widespread therein. The best studied fusellovirus is SSV1 from Beppu, Japan, which infects *Sulfolobus solfataricus*. Very little is known about the function of the genes in the SSV1 genome. Recently we have developed genetic tools to analyze these genes. In this study, we have deleted three SSV1 open reading frames (ORFs) ranging from completely conserved to poorly conserved: VP2, d244, and b129. Deletion of the universally conserved ORF b129, which encodes a predicted transcriptional regulator, results in loss of infectivity. Deletion of the poorly conserved predicted DNA-binding protein gene VP2 yields viable virus that is indistinguishable from wild-type. Deletion of the well-conserved ORF d244 that encodes a predicted nuclease yields viable virus. However, infection of *S. solfataricus* with virus lacking ORF d244 dramatically retards host growth, compared to the wild-type virus.

Keywords: DNA binding, nuclease, transcription factor

INTRODUCTION

Viruses of Archaea are very poorly understood with only about 50 known archaeal viruses relative to the ca. 5000 characterized viruses of bacteria, plants, and animals (Pina et al., 2011). The best studied of archaeal viruses are those infecting the thermoacidophiles, with an unprecedented new seven virus families introduced in the last few years to accommodate the astonishing morphological and sequence diversity present in these viruses (Pina et al., 2011).

The *Sulfolobus* spindle-shaped viruses (SSVs) of the family Fuselloviridae were the first discovered and probably the best studied family of archaeal viruses. SSVs are found throughout the world in high temperature (>70°C) and acidic (pH < 4) environments where their hosts, *Sulfolobus solfataricus* and its close relatives thrive (Wiedenheft et al., 2004; Held and Whitaker, 2009). The type virus, SSV1, encodes a positively supercoiled, 15.5 kbp circular dsDNA genome (NC_001338.1) that is enclosed within a lemon or spindle-shaped capsid (Yeats et al., 1982; Martin et al., 1984; Nadal et al., 1986). The genome encodes 34 open reading frames (ORFs; Palm et al., 1991), most of which have no recognizable homologs apart from other Fuselloviridae. The only SSV1 gene with clear homology to proteins outside the Fuselloviridae is the viral integrase, encoded by ORF d355. The main structural proteins purified from virus particles are the major and minor capsid proteins VP1 and VP3 and the putative DNA packaging protein VP2 (Reiter et al., 1987a). More recently, mass spectrometric analysis of SSV1 virions revealed two additional proteins, the products of ORFs c792 and d244 (Menon et al., 2008; **Figure 1**).

In the absence of homologous sequences, three complementary approaches have been used to try and determine the function of the proteins encoded in the SSV1 genome; structural genomics, comparative genomics, and genetics. Atomic resolution structures have been obtained by C. Martin Lawrence and his group for proteins encoded by SSV1 ORFs b129, f112, d63, e96, f93, and d244 or their homologs from other fuselloviruses. The products of ORFs b129, f112, and f93 resemble transcriptional regulators and d244 a novel nuclease (Lawrence et al., 2009; Menon et al., 2010). However, the function of these proteins in virus replication remains to be determined. Two of these ORFs, b129 and d244, are the targets of the current study.

In parallel, we and others have undertaken comparative genomic studies. Fifteen ORFs are completely conserved in 12 canonical SSV genomes (Stedman et al., 2003; Wiedenheft et al., 2004; Held and Whitaker, 2009; Redder et al., 2009; Stedman, unpublished; **Figure 1**). Most of the universally conserved genes are clustered in half of the genome with the notable exception of the VP2 gene, a target of this study. Conservation in the rest of the genome is lower. Nonetheless, there are very few completely unique genes in the SSV1 genome (**Figure 1**). It is highly probable that the conserved genes are required for virus function, but again this has not been confirmed.

We developed methods for gene disruption in order to determine the requirements for genes in the virus genome directly. About 10 years ago, we showed that four SSV1 ORFs did not tolerate insertion of the 3.2 kbp pBluescript plasmid and allow virus function. Twelve other SSV1 ORFs appeared, indirectly, to not tolerate insertion. However, two ORFs, e178 and e51, were able to tolerate insertion of the entire pBluescript plasmid (Stedman et al., 1999). This result allowed the development of viral shuttle vectors and the beginnings of *Sulfolobus* genetics (Jonuscheit et al., 2003). Insertion of the pBluescript plasmid and up to ca. 5 kbp of exogenous DNA in these ORFs does not appear to have a noticeable effect on virus function (Stedman et al., 1999; Jonuscheit et al., 2003; Clore and Stedman, 2006; Albers et al., 2006).

However, insertion of large DNA fragments into the SSV1 genome is not straightforward and the possible insertion locations

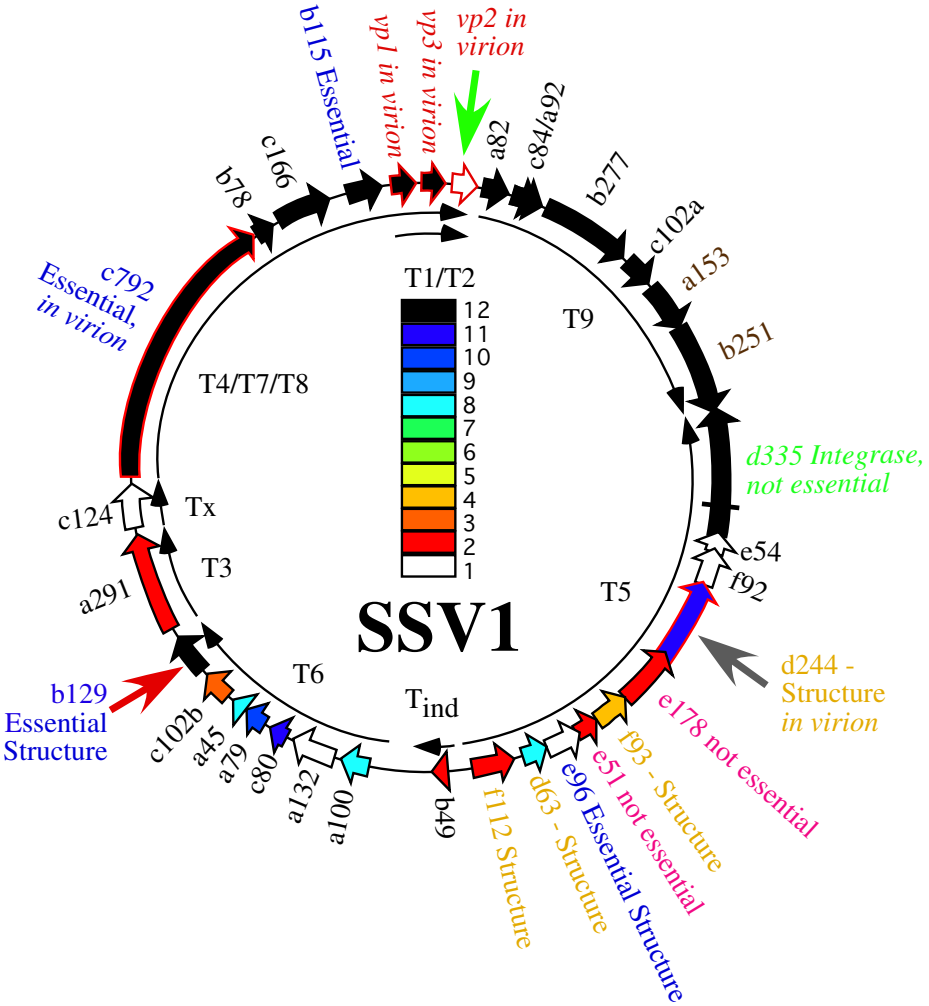


FIGURE 1 | Genome map of SSV1. Open reading frames are shown as block arrows and labeled as in Palm et al. (1991). Virus structural protein genes (Reiter et al., 1987a) and other proteins found in the virion (Menon et al., 2008) are outlined in red and labeled as "in virion." Conservation of open reading frames in 12 canonical SSV genomes (SSV1, SSV2, SSV3, SSV4, SSV5, SSVRH, SSVK1, SSVL, SSVKM1, SSVKU1, SSVL2, and SSVGV1; Redder et al., 2009; Held and Whitaker, 2009; Stedman, unpublished) is listed with the color code in the middle of the genome with ORFs conserved in 12 genomes in black, ORFs conserved in 11 genomes in dark blue, etc. ORFs which did not tolerate insertion of the pBluescript plasmid are labeled as "Essential" in blue type. ORFs allowing insertion of the pBluescript plasmid without loss of virus function are labeled as "not essential" (Stedman et al., 1999). All ORFs whose products have been crystallized and structure determined are labeled as "Structure" (Lawrence et al., 2009; Menon et al., 2010). The gene for the SSV1-integrase is labeled in green and was shown to be not essential by deletion (Clore and Stedman, 2006). Transcripts are labeled as curved thin arrows (Reiter et al., 1987b; Fröls et al., 2007). ORFs targeted in this study are indicated with large arrows outside the genome map.

are limited. Therefore, Long Inverse PCR (LIPCR) using high-fidelity highly processive DNA polymerases (e.g., Phusion®) was developed to specifically change the SSV1 genome at single nucleotide resolution. LIPCR was used to delete precisely the SSV1 viral integrase gene. Surprisingly, this "integrase-less" SSV1 was functional (Clore and Stedman, 2006). However, consistent with its conservation, the virus lacking the integrase gene is at a competitive disadvantage relative to integrase-containing viruses (Clore and Stedman, 2006). All of the SSV1 ORFs that can be deleted or tolerate insertion without abrogating virus function are in the "early" transcript, T5, that is induced soon after UV-irradiation of SSV-infected cultures (Reiter et al., 1987b; Fröls et al., 2007).

Three ORFs in the SSV1 genome were targeted for gene disruption in this study. The VP2 gene (NP_039802.1) was chosen for disruption because it is only present in SSV1 and the very distantly related SSV6 (Held and Whitaker, 2009; Redder et al., 2009), and is in the middle of the most highly conserved part of fusellovirus genomes (**Figure 1**). VP2 has DNA-binding activity (Reiter et al., 1987a; Iverson and Stedman, unpublished) that is presumably required for DNA packaging. ORF b129 (NP_039795.1) was chosen because it is intolerant of insertional mutagenesis (Stedman et al., 1999), a high resolution structure is known (Lawrence et al., 2009) and the gene is completely conserved in all SSVs (**Figure 1**). Finally, ORF d244 (NP_039781.1) was chosen for gene disruption because a high-resolution structure of its homolog from SSVRH

is known (Menon et al., 2008) and it is conserved in most SSV genomes with the exception of SSVK1.

MATERIALS AND METHODS

CULTURE CONDITIONS

Sulfolobus solfataricus strains, **Table 1**, were grown aerobically at 76°C on plates or in liquid media containing yeast extract and sucrose as carbon and energy sources (YS Media), both as in Jonuscheit et al. (2003). *Escherichia coli* strains were grown in LB medium at 37°C as suggested by the manufacturer (Novagen).

PURIFICATION OF DNA

Plasmid DNA used for LIPCR was purified from *E. coli* using the alkaline lysis method of Birnboim and Doly (1979). Plasmid DNA used to transform *Sulfolobus* was purified using the GeneJet Plasmid Purification Kit (Fermentas) following the manufacturer's protocols. Total genomic DNA was isolated from *S. solfataricus* in late log phase growth (OD600 ∼0.6) as in Stedman et al. (1999). Plasmid DNA was purified from a 50 mL culture of *S. solfataricus* transformed with SSV-Δd244 (late log, OD600 ∼0.6) using the GeneJet plasmid purification kit (Fermentas) following the manufacturer's protocols. This DNA was retransformed into *E. coli* (Novagen), purified therefrom and analyzed by restriction endonuclease digestion with *Eco*RI (Fermentas).

CONSTRUCTION OF SSV1 DELETION MUTANTS

Deletion mutants were constructed from the pAJC97 shuttle vector using LIPCR (Clore and Stedman, 2006). Primers were designed to overlap with the start and stop codon of the ORF to be deleted to keep the deletion in frame. Initially primers were designed using the archaea genome browser[1]. Primer melting temperatures were matched and then checked for potential primer dimer and secondary structure formation using online tools from IDT[2]. **Table 2** contains a list of oligonucleotide sequences used. LIPCR was performed using Phusion® High-Fidelity DNA Polymerase (NEB/Finnzymes) at a final concentration of 0.005 U/μL. LIPCR cycling conditions as follows: initial denaturation at 98°C for 3 min; 35 cycles of 98°C for 15 s, annealing for 15 s, 72°C for 6 min, and a final extension at 72°C for 6 min. The annealing temperatures for deletion of VP2, ORF d244, and ORF b129 were 59, 53, and 66°C, respectively. DNA was precipitated directly from LIPCR reactions using sodium acetate at a final concentration of 0.3 M and 95% EtOH. This DNA was phosphorylated using T4 polynucleotide kinase according to the manufacturer's protocols (Fermentas). DNA was ligated overnight (∼20 h) at 16°C using 5 Weiss units of T4 DNA ligase (Fermentas). Ligated DNA was transformed into NovaBlue Singles chemically competent *E. coli* following the manufacturer's protocol (Novagen). Plasmids were purified from single colonies and deletion constructs were identified by restriction endonuclease digestions. The deletion borders were confirmed by sequencing of the plasmids.

ELECTROPORATION OF *SULFOLOBUS*

Purified plasmid DNA was electroportated into *Sulfolobus* strain GΘ as in Schleper et al. (1992). Following electroporation (400Ω, 1.5 kV, 25 μF), cells were immediately resuspended in 1 mL of YS media at 75°C and incubated for 1 h at 75°C. The cells were then added to 50 mL of prewarmed YS media (75°C) and grown in liquid media as outlined below.

SCREEN FOR FUNCTIONAL INFECTIOUS VIRUS/HALO ASSAY

To confirm the presence of infectious virus, halo assays were performed in duplicate 48 and 72 h post-electroporation (Stedman et al., 2003). Uninfected *Sulfolobus* GΘ cells were diluted to an OD600 nm = ∼0.3 and allowed to grow until the OD600 nm reached ∼0.35 (about 2.5 h). Half of a milliliter of this uninfected culture was added to 5 mL YS media containing 0.2% wt/vol Gelrite® as a softlayer and poured onto prewarmed YS plates. Two microliters of supernatant from electroporated cultures was spotted onto the lawns and plates were incubated at 75°C for up to 3 days. A halo of host growth inhibition, typically observed 48–72 h after incubation, indicated the presence of an infectious virus (**Figure 2**).

GROWTH CURVES

Portions of halos of growth inhibition from infected *S. solfataricus* GΘ cells were removed from plates with a sterile pipette tip and inoculated into liquid YS media. The culture was grown to an OD600 nm of ∼0.6. One milliliter of this culture was diluted in 100 mL YS media to an OD600 ∼0.050. Cultures were placed in a shaking incubator at 75°C and the OD600 nm was measured

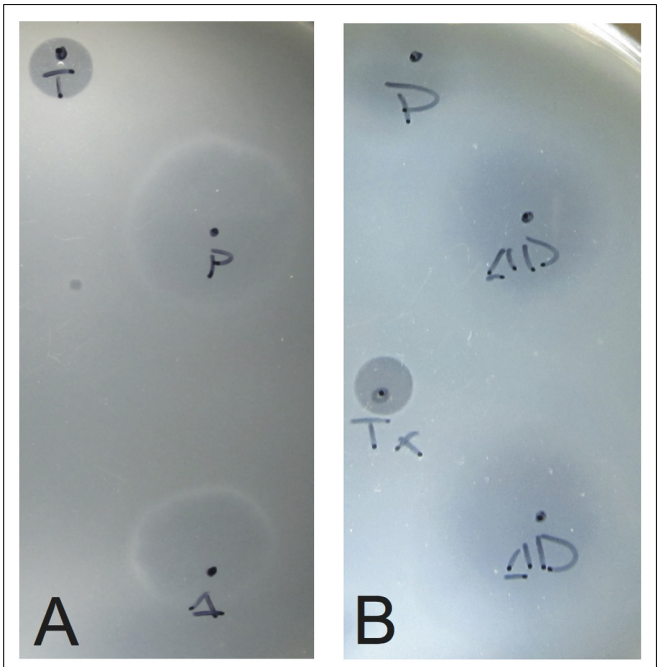


FIGURE 2 | Typical growth inhibition of *S. solfataricus* on plates due to infectious virus. Lawns of *S. solfataricus* strain GΘ were prepared as in Stedman et al. (2003). Two microliters of supernatant from cultures transformed with either **(A)** SSV-ΔVP2 or **(B)** SSV-Δd244 were placed on the lawns where indicated. Δ indicates where SSV-ΔVP2 was spotted, ΔD where SSV-Δd244 was spotted. P indicates SSV-WT spotted as a positive control. T or Tx indicates 2 μL of 0.01% Triton X-100 spotted as a control for lawn growth.

[1]http://archaea.ucsc.edu

[2]http://www.idtdna.com

every 24 h. After 96 h, 1 mL of culture was diluted into 100 mL fresh YS media and returned to 75°C. One milliliter of culture was removed 72 h after each dilution, cells removed by centrifugation (14000 rpm for 5 min in a microcentrifuge) and the supernatant was screened for virus using the halo assay above.

TRANSMISSION ELECTRON MICROSCOPY

Supernatant from infected cultures was collected by centrifugation at 14,000 rpm for 5 min in a microcentrifuge. Five microliters of supernatant was absorbed onto a 400 mesh carbon/formvar grid (Ted Pella) for 2 min and negatively stained with 2% uranyl acetate for 20 sec. Grids were viewed on a JEOL 100CX TEM operated at 100 keV and images captured with a Gatan imager.

RESULTS

SSV1 IS INFECTIOUS WITHOUT THE VP2 GENE

The VP2 protein was purified from SSV1 virus particles and reported to be a DNA-binding protein (Reiter et al., 1987a). Surprisingly, a gene for VP2 was not found in SSV2 (Stedman et al., 2003) or SSVRH or SSVK1 (Wiedenheft et al., 2004). Moreover, a homolog is not present in the *S. solfataricus* or *S. islandicus* genomes (She et al., 2001; Reno et al., 2009; Guo et al., 2011). However, a very distant relative of SSV1, SSV6, which also contains an atypical putative tail fiber protein, has a VP2 gene (Redder et al., 2009). Thus, it is not clear whether SSV1 can function without a VP2 gene.

Therefore, we made an in-frame deletion of the majority of the VP2 gene by LIPCR in the context of the pAJC97 SSV1 shuttle vector (Clore and Stedman, 2006), leaving the first four codons and the last four codons (including the stop codon) of the ORF intact (see **Table 1**). The putative promoter for the T9 "early" transcript was also left intact. The construct containing the deletion, pAJC97-ΔVP2, is hereafter referred to as SSV-ΔVP2.

To determine if the SSV-ΔVP2 was able to make infectious virus, the shuttle vector was electroporated into *S. solfataricus* strain GΘ. Two days after electroporation, the supernatant from the transformed strains caused inhibition of growth of uninfected *S. solfataricus* strain GΘ on plates (**Figure 2**) that was indistinguishable from growth inhibition caused by the virus containing the VP2 gene. Similar growth inhibition was also observed on lawns of uninfected *S. solfataricus* strain S443, a new *S. solfataricus* isolate from Lassen Volcanic National Park that is a host for all tested SSVs (Ceballos et al., in preparation). Moreover, the supernatant contained SSV-like particles when observed by transmission electron microscopy (**Figure 3**).

Infection by wild-type SSV1 and shuttle vectors does not drastically slow growth of cells in liquid culture for unknown reasons (Martin et al., 1984; Schleper et al., 1992; Stedman et al., 1999). The same is true of SSV-ΔVP2 (**Figure 4**). Infection with SSV-ΔVP2 was confirmed via PCR amplification (data not shown)

Table 1 | Strains and plasmid vectors used in this work.

Strain/vector	Description	Reference
S. solfataricus GΘ	MT4 Derivative	Cannio et al. (1998)
S. solfataricus S443	Novel *Sulfolobus* isolate	Unpublished data
E. coli NovaBlue®	Expression strain	Novagen, Inc.
pAJC97	SSV1 with TOPO PCR Blunt II	Clore and Stedman (2006)
pAJC97-ΔVP2	pAJC97 lacking VP2 gene	This Work
pAJC97-Δd244	pAJC97 lacking ORF d244	This Work
pAJC97-Δb129	pAJC97 lacking ORF b129	This Work

SSV1 CONSTRUCTS LACKING THE CONSERVED ORF b129 DO NOT APPEAR TO MAKE INFECTIVE VIRUSES

The b129 ORF in SSV1 is universally conserved in all fuselloviruses (Redder et al., 2009). Moreover shuttle vectors with pBluescript inserted into ORF b129 did not produce infective virus when electroporated into *Sulfolobus* (Stedman et al., 1999). However, a similar insertion mutant in the equally conserved SSV1 viral integrase appears to be non-functional (Stedman et al., 1999), but an in-frame deletion was functional (Clore and Stedman, 2006). A structure for the b129 ORF is also known (Lawrence et al., 2009) and it contains two Zn-finger putative DNA-binding motifs.

The b129 ORF was deleted with LIPCR. The deletion of the b129 ORF left the first four and last two codons of the ORF intact and maintained the predicted T3 promoter (Reiter et al., 1987b). This construct is referred to as SSV-Δb129. Unlike the SSV-ΔVP2 construct, supernatants from *Sulfolobus* cells electroporated with SSV1-Δb129 did not cause zones of growth inhibition when spotted on lawns of uninfected *S. solfataricus* strain GΘ. A total of nine independent transformations were performed in which the wild-type virus consistently caused growth inhibition but SSV-Δb129 did not. Moreover, no halos of growth inhibition were formed on lawns of *S. solfataricus* strain S443. It is not currently known at which step of virus replication the SSV-Δb129 is deficient.

SSV1 LACKING ORF d244 IS INFECTIOUS BUT HAS A NOVEL PHENOTYPE

SSV1 ORF d244 is in the UV-inducible transcript T5, upstream of the viral integrase gene (**Figure 1**). The entire pBluescript plasmid can be inserted into the ORF directly upstream of ORF d244 without abrogating SSV1 function (Stedman et al., 1999). ORF d244 is well conserved in other Fusellovirus genomes with the exception of SSVK1 (Wiedenheft et al., 2004; Redder et al., 2009). The X-ray crystal structure of the homolog of SSV1 ORF d244, SSVRH ORF d212 has been solved and it is predicted to be a nuclease (Menon et al., 2010). Moreover, the product of ORF d244 has been reported to be in purified SSV1 particles (Menon et al., 2008).

The SSV1 d244 ORF was deleted with LIPCR. The deletion of the d244 ORF left the first two and last three codons of the ORF intact as well as maintained the ORF to avoid polar effects. This construct is referred to as SSV-Δd244.

To determine if SSV-Δd244 was able to make infectious virus, the shuttle vector was electroporated into *S. solfataricus* strain GΘ. Two days after electroporation, the supernatant from the transformed strains caused inhibition of growth of uninfected *S. solfataricus* strain GΘ on plates (**Figure 2**) and also inhibited growth of *S. solfataricus* strain S443 (data not shown).

Table 2 | Oligonucleotides used in this work.

Name	Sequence	Description
VP2 LIPCR F	5′-CAC CGC AAG TAG GCC-3′	Flanks VP2 gene for deletion
VP2 LIPCR R	5′-CAC CCA CTT CAT ATC ACT CC-3′	Flanks VP2 gene for deletion
d244 LIPCR F	5′-ATC CAT TTA CCA TAA TCC ACC-3′	Flanks ORF d244 for deletion
d244 LIPCR R	5′-GGA AAA TGA TAT TCA ACT CAG AGG-3′	Flanks ORF d244 for deletion
b129 LIPCR F	5′-AGT TAG GCT CTT TTT AAA GTC TAC C-3′	Flanks ORF b129 for deletion
b129 LIPCR R	5′-TGA CTC CGT CAT CCT CTA AC-3′	Flanks ORF b129 for deletion
VP2 Check F	5′-ATT CAG ATT CTG WAT WCA GAA C-3′	Amplifies VP2 gene and flanking sequences
VP2 Check R	5′-TCS CCT AAC GCA CTC ATC-3′	Amplifies VP2 gene and flanking sequences
d244 Check F	5′-GGA ACT CCT CTC ATT AAC C-3′	Amplifies ORF d244 and flanking sequences
d244 Check R	5′-GAT CAT CAA CGA GTA TAT TGA CC-3′	Amplifies ORF d244 and flanking sequences
b129 Check F	5′-ATG AAG GCT GAG GAA ACA ATC GTG-3′	Amplifies ORF b129 and flanking sequences
b129 Check R	5′-TTA ATA TAG CTG CGA TGC AGT ATA GTT TAT TTG TGC-3′	Amplifies ORF b129 and flanking sequences

**Underlined sequence indicates ORF.*

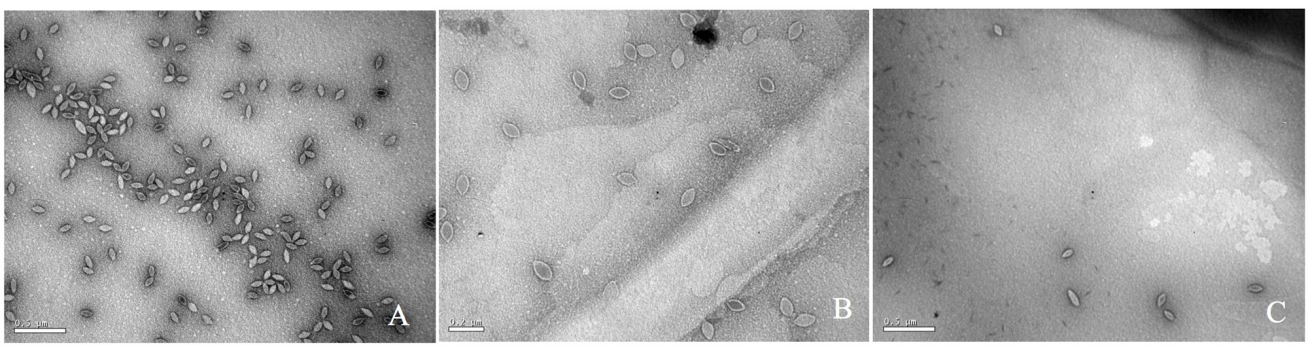


FIGURE 3 | Transmission electron micrographs of SSV particles. Supernatants from cultures of *S. solfataricus* strain GΘ transformed with **(A)** pAJC97, **(B)** SSV-ΔVP2, **(C)** SSV-Δd244, were negatively stained with uranyl acetate and observed with a JEOL 100CX transmission electron microscope. Bar represents 0.2 μm **(B)** or 0.5 μm **(A,C)**.

The supernatant contained SSV-like particles when observed by transmission electron microscopy (**Figure 3**).

Infection by wild-type SSV1, shuttle vectors and SSV-ΔVP2 does not slow growth of cells in liquid culture (Martin et al., 1984; Schleper et al., 1992; Stedman et al., 1999; see above). However, infection by SSV-Δd244 drastically slows growth of *S. solfataricus* strains GΘ and S443 in liquid culture (**Figure 4**). Infection with SSV-Δd244 was confirmed via PCR. Moreover, restriction endonuclease digestion of viral DNA recovered from transformed *S. solfataricus* cells and retransformed into *E. coli* revealed no obvious alterations of the SSV-Δd244 construct (data not shown).

DISCUSSION

THE PUTATIVE DNA PACKAGING PROTEIN VP2 IS NOT REQUIRED FOR SSV1 FUNCTION

The deletion of VP2 from SSV1 results in a functional virus that is indistinguishable from the wild-type virus (**Figures 2–4).** Based on the lack of conservation of VP2 this result is not completely unexpected. However, almost all viruses contain a genome packaging protein. There is no clear sequence homolog of VP2 in the host genome, but there are a number of small DNA-binding proteins, such as Sso7d or Cren7 that may be able to functionally substitute for VP2 in SSV1 genome packaging (Choli et al., 1988; Guo et al., 2008). This will be tested with mass spectrometry of SSV-ΔVP2 particles. Alternatively, the VP2 protein may be involved in maintenance of the positive supercoiling of the SSV1 viral genome (Nadal et al., 1986). It would be interesting to know if the topology of the viral DNA is affected by the absence of VP2. It is predicted that positive supercoiling should increase the thermal stability of the DNA, so SSV-ΔVP2 may be less thermally stable than the wild-type virus.

The VP2 gene may be more prevalent than previously thought. VP2-like sequences have been reported from metagenomic studies, one in an acid mine drainage metagenome (Andersson and Banfield, 2008) and the other from Boiling Springs Lake in California (Diemer and Stedman, unpublished). These VP2 genes may be in the context of a SSV6 or ASV-like genome (Redder et al., 2009).

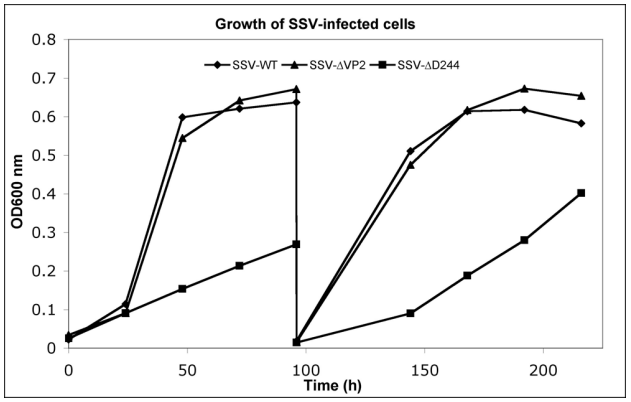


FIGURE 4 | Typical growth inhibition in liquid culture of virus constructs. Cultures of *S. solfataricus* GΘ infected with wild-type SSV1, diamonds, SSV-ΔVP2, triangles and SSV-Δd244, squares, were diluted in YS media to equal starting OD600 nm and incubated at 75°C. At the indicated times, samples were removed and the OD600 nm was determined and the presence of virus was confirmed in each culture via halo assay. After 96 h, 1 mL of cells were diluted 1:100 in fresh YS media and returned to 75°C.

THE PRODUCT OF ORF b129 APPEARS TO BE ESSENTIAL FOR SSV1 INFECTIVITY

Homologs of SSV1 ORF b129 are present in all known SSVs (Redder et al., 2009). The b129 ORF also does not tolerate insertion of the pBluescript plasmid (Stedman et al., 1999). Thus, it is not surprising that deletion of ORF b129 leads to an incompletely replicating virus. However, the SSV1 integrase, a gene also conserved in all fuselloviruses, did not appear to tolerate insertion of pBluescript (Stedman et al., 1999), but could be deleted with LIPCR without abrogating virus function (Clore and Stedman, 2006). This indicates that either polar effects are important, which seems unlikely since the SSV1 integrase is at the end of the T5 transcript, or that insufficient replicate transformations were performed in the earlier study.

Nine replicate transformations of *S. solfataricus* with SSV-Δb129 did not generate functional virus. However, we cannot absolutely determine that SSV1 ORF b129 is essential for virus function without complementation experiments, which are underway. The reasons for the apparent necessity of SSV1 ORF b129 are unclear, but the structure of the b129 ORF product, a predicted transcriptional regulator (Lawrence et al., 2009) and induction of the T6 transcript containing ORF b129 after UV-irradiation (Reiter et al., 1987b; Fröls et al., 2007) provides clues to its function.

The assay used herein for virus infection, ability to cause a zone of growth inhibition on a lawn of uninfected cells, is for virus spread and infectivity. There are many other aspects of virus replication that could be affected by disruption of ORF b129. An attractive hypothesis is that the b129 protein activates transcription of virus structural genes encoded by the "late" transcripts T7/8/9, T1, and T2 (Reiter et al., 1987b; Fröls et al., 2007; see **Figure 1**). This would be one of very few archaeal transcriptional activators characterized to date and the only the second archaeal viral transcriptional activator (Kessler et al., 2006). Thus, the SSV-Δb129 construct may be able to replicate its genome, integrate into the host, and have genome replication induced by UV-irradiation or some subset of these activities. Experiments to test these hypotheses are underway.

TRANSFECTION WITH SSV-Δd244 PRODUCES VIRUS AND RETARDS HOST CELL GROWTH

The SSV1 d244 ORF is well-conserved in fuselloviruses with the exception of SSVK1 (Wiedenheft et al., 2004; Redder et al., 2009). However, SSV1 lacking ORF d244 clearly makes infectious virus particles (**Figures 2 and 3**). Moreover, the zones of clearing produced by supernatants of cells transfected with SSV-Δd244 are clearer than those produced by either the wild-type or SSV-ΔVP2 viruses (**Figure 3**; unpublished data). They are reminiscent of zones of clearing produced by SSVK1 (data not shown). Unlike wild-type virus and SSV-ΔVP2, transfection by SSV-Δd244 leads to drastically reduced host growth (**Figure 4**). The reasons for this growth inhibition are unclear. Similar growth phenotypes have been observed in SSVK1 infections (Stedman et al., in preparation). SSVK1 consistently produces more virus than similar cultures of the wild-type virus, so this may account for the growth defect (unpublished data). Whether SSV-Δd244 consistently produces more virus than the wild-type or SSV-ΔVP2 is currently unknown.

The structure of the product of SSV1 ORF d244 is a predicted nuclease (Menon et al., 2010), similar to Holiday junction resolvase enzymes. Why the lack of a resolvase leads to slower host growth is unclear. Possibly SSV1 ORF d244 is involved in the specificity of SSV1 integration. SSV-K1 is known to integrate into multiple positions in the host genome (Wiedenheft et al., 2004), which may contribute to its higher copy number. Whether SSV-Δd244 integrates into multiple positions in the host genome is under investigation. On the other hand, there may be a defect in SSV-Δd244 replication or resolution of SSV replication intermediates that leads to accumulation of aberrant DNA, which, in turn, leads to slower host growth.

After multiple transfers of *Sulfolobus* cultures transfected with SSV-Δd244 into fresh media, growth rates recover to near wild-type rates (unpublished data). The virus is still present in these cultures by PCR and is able to inhibit *Sulfolobus* growth on plates (unpublished data) so the virus is not lost or apparently rearranged (see Results). Whether there are other genetic changes in the virus or host under these conditions remains to be determined. One attractive possibility is changes to the CRISPR repeat structures that are proposed to be important for acquired immunity in *Sulfolobus* (Held and Whitaker, 2009).

SUMMARY AND OUTLOOK

Comparative and structural genomics has identified a number of targets for gene disruption in the SSV1 genome. Here precise gene disruptions of the poorly conserved VP2 gene, and the well-conserved ORFs b129 and d244 are described. Deletions in VP2 may allow insights into DNA packaging in the SSV1 genome. Deletion of ORF b129 may allow the identification of the second archaeal virus transcriptional activator. Deletion of ORF d244 may allow insight into copy number regulation in SSVs, previously thought to be regulated by ORF d63 (Lawrence et al., 2009). Clearly, there are many more genes to be analyzed in the SSV1

genome and more insights that can be gained by combining comparative genomics, structural biology, and genetics. In the future, biochemical work will be added to this suite of techniques to gain fundamental understanding of this fascinating, unique, and ubiquitous archaeal virus family.

ACKNOWLEDGMENTS

This research was supported by Portland State University, NASA award NNX07AT63A. Subaward G258-08-W1951 and NSF grant MCB-0702020. The authors would like to thank Adam Clore for his advice on Long Inverse PCR and SSV comparative genomics.

REFERENCES

Albers, S. V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampé, R., Driessen, A. J., and Schleper, C. (2006). Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl. Environ. Microbiol.* 72, 102–111.

Anderson, A. F., and Banfield, J. F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* 320, 1047–1050.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7, 1513–1523.

Cannio, R., Contursi, P., Rossi, M., and Bartolucci, S. (1998). An autonomously replicating transforming vector for *Sulfolobus solfataricus*. *J. Bacteriol.* 180, 3237–3240.

Choli, T., Henning, P., Wittmann-Liebold, B., and Reinhardt, R. (1988). Isolation, characterization and microsequence analysis of a small basic methylated DNA-binding protein from the Archaebacterium, *Sulfolobus solfataricus*. *Biochim. Biophys. Acta* 950, 193–203.

Clore, A. J., and Stedman, K. M. (2006). The SSV1 viral integrase is not essential. *Virology* 361, 103–111.

Fröls, S., Gordon, P. M., Panlilio, M. A., Schleper, C., and Sensen, C. W. (2007). Elucidating the transcription cycle of the UV-inducible hyperthermophilic archaeal virus SSV1 by DNA microarrays. *Virology* 365, 48–59.

Guo, L., Brügger, K., Liu, C., Shah, S. A., Zheng, H., Zhu, Y., Wang, S., Lillestøl, R. K., Chen, L., Frank, J., Prangishvili, D., Paulin, L., She, Q., Huang, L., and Garrett, R. A. (2011). Genome analyses of Icelandic strains of *Sulfolobus islandicus* model organisms for genetic and virus-host interaction studies. *J. Bacteriol.* 193, 1672–1680.

Guo, L., Feng, Y., Zhang, Z., Yao, H., Luo, Y., Wang, J., and Huang, L. (2008). Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among crenarchaea. *Nucleic Acids Res.* 36, 1129–1137.

Held, N. L., and Whitaker, R. J. (2009). Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. *Environ. Microbiol.* 11, 457–466.

Jonuscheit, M., Martusewitsch, E., Stedman, K. M., and Schleper, C. (2003). A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol.* 48, 1241–1252.

Kessler, A., Sezonov, G., Guijarro, J. I., Desnoues, N., Rose, T., Delepierre, M., Bell, S. D., and Prangishvili, D. (2006). A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters. *Nucleic Acids Res.* 34, 4837–4845.

Lawrence, C. M., Menon, S., Eilers, B. J., Bothner, B., Khayat, R., Douglas, T., and Young, M. J. (2009). Structural and functional studies of archaeal viruses. *J. Biol. Chem.* 284, 12599–12603.

Martin, A., Yeats, S., Janekovic, D., Reiter, W. D., Aicher, W., and Zillig, W. (1984). SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. *EMBO J.* 3, 2165–2168.

Menon, S. K., Eilers, B. J., Young, M. J., and Lawrence, C. M. (2010). The crystal structure of D212 from *Sulfolobus* spindle-shaped virus ragged hills reveals a new member of the PD-(D/E)XK nuclease superfamily. *J. Virol.* 84, 5890–5897.

Menon, S. K., Maaty, W. S., Corn, G. J., Kwok, S. C., Eilers, B. J., Kraft, P., Gillitzer, E., Young, M. J., Bothner, B., and Lawrence, C. M. (2008). Cysteine usage in *Sulfolobus* spindle-shaped virus 1 and extension to hyperthermophilic viruses in general. *Virology* 376, 270–278.

Nadal, M., Mirambeau, G., Forterre, P., Reiter, W. D., and Duguet, M. (1986). Positively supercoiled DNA in a virus-like particle of an archaebacterium. *Nature* 321, 256–258.

Palm, P., Schleper, C., Grampp, B., Yeats, S., McWilliam, P., Reiter, W. D., and Zillig, W. (1991). Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. *Virology* 185, 242–250.

Pina, M., Bize, A., Forterre, P., and Prangishvili, D. (2011). The archeoviruses. *FEMS Microbiol. Rev.* 35, 1035–1054.

Redder, P., Peng, X., Brügger, K., Shah, S. A., Roesch, F., Greve, B., She, Q., Schleper, C., Forterre, P., Garrett, R. A., and Prangishvili, D. (2009). Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. *Environ. Microbiol.* 11, 2849–2862.

Reiter, W. D., Palm, P., Henschen, A., Lottspeich, F., and Zillig, W. D. (1987a). Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet.* 206, 144–153.

Reiter, W. D., Palm, P., Yeats, S., and Zillig, W. (1987b). Gene-expression in Archaebacteria - Physical mapping of constitutive and UV-inducible transcripts from the *Sulfolobus* virus-like particle SSV1. *Mol. Gen. Genet.* 209, 270–275.

Reno, M. L., Held, N. L., Fields, C. J., Burke, P. V., and Whitaker, R. J. (2009). Biogeography of the *Sulfolobus* pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8605–8610.

Schleper, C., Kubo, K., and Zillig, W. (1992). The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: Demonstration of infectivity and of transfection with viral DNA. *Proc. Natl. Acad. Sci. U.S.A.* 89, 7645–7649.

She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W., and Van der Oost, J. (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* 98, 7835–7840.

Stedman, K. M., Schleper, C., Rumpf, E., and Zillig, W. (1999). Genetic requirements for the function of the archaeal virus SSV1 in *Sulfolobus solfataricus*: construction and testing of viral shuttle vectors. *Genetics* 152, 1397–1405.

Stedman, K. M., She, Q., Phan, H., Arnold, H. P., Holz, I., Garrett, R. A., and Zillig, W. (2003). Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. *Res. Microbiol.* 154, 295–302.

Wiedenheft, B., Stedman, K. M., Roberto, F., Willits, D., Gleske, A. K., Zoeller, L., Snyder, J., Douglas, T., and Young, M. (2004). Comparative genomic analysis of hyperthermophilic archaeal Fuselloviridae genomes. *J. Virol.* 78, 7438–7442.

Yeats, S., McWilliam, P., and Zillig, W. (1982). A plasmid in the archaebacterium *Sulfolobus acidocaldarius*. *EMBO J.* 1, 1035–1038.

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

Received: 15 March 2012; paper pending published: 16 April 2012; accepted: 15 May 2012; published online: 05 June 2012.

*Citation: Iverson E and Stedman K (2012) A genetic study of SSV1, the prototypical fusellovirus. Front. Microbio. **3**:200. doi: 10.3389/fmicb.2012.00200*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 13 June 2012
doi: 10.3389/fmicb.2012.00214

Versatile genetic tool box for the crenarchaeote *Sulfolobus acidocaldarius*

Michaela Wagner, Marleen van Wolferen, Alexander Wagner, Kerstin Lassak, Benjamin H. Meyer, Julia Reimann and Sonja-Verena Albers*

Molecular Biology of Archaea, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

Edited by:
Zvi Kelman, University of Maryland, USA

Reviewed by:
Paul Beare, Rocky Mountain Laboratories, USA
Miao Pan, University of Maryland, USA

****Correspondence:***
Sonja-Verena Albers, Molecular Biology of Archaea, Max Planck Institute for terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany.
e-mail: albers@mpi-marburg.mpg.de

For reverse genetic approaches inactivation or selective modification of genes are required to elucidate their putative function. *Sulfolobus acidocaldarius* is a thermoacidophilic Crenarchaeon which grows optimally at 76°C and pH 3. As many antibiotics do not withstand these conditions the development of a genetic system in this organism is dependent on auxotrophies. Therefore we constructed a *pyrE* deletion mutant of *S. acidocaldarius* wild type strain DSM639 missing 322 bp called MW001. Using this strain as the starting point, we describe here different methods using single as well as double crossover events to obtain markerless deletion mutants, tag genes genomically and ectopically integrate foreign DNA into MW001. These methods enable us to construct single, double, and triple deletions strains that can still be complemented with the pRN1 based expression vector. Taken together we have developed a versatile and robust genetic tool box for the crenarchaeote *S. acidocaldarius* that will promote the study of unknown gene functions in this organism and makes it a suitable host for synthetic biology approaches.

Keywords: archaea, *Sulfolobus*, genetics, deletion mutant, expression system, in-frame deletion

INTRODUCTION

In contrast to the bacterial domain, research on Archaea has long been hampered by the absence of genetic tools to study gene functions *in vivo*. One major reason for this was that archaea are resistant to most commonly used antibiotics that are used in microbial genetics as selectable markers (Leigh et al., 2011). These compounds very often target the peptidoglycan synthesis which is absent in most archaea (Albers and Meyer, 2011) or the antibiotics are readily degraded at the conditions at which various archaea live.

Genetic systems were described for haloarchaea and methanogens only in the late 1990s, respectively, and have constantly improved since then (Leigh et al., 2011). For another euryarchaeon *Thermococcus kodakariaensis* an effective gene deletion system was established by Sato et al. (2003, 2005) for which also a complementary expression vector has been optimized (Santangelo et al., 2008). The most recent system was developed for *Pyrococcus furiosus*. Hence, nowadays genetic toolboxes exist for a variety of euryarchaeota (Waege et al., 2010).

In the kingdom crenarchaeota, only for organisms of the order Sulfolobales genetic systems have been developed. Sulfolobales are thermoacidophilic microorganisms that grow optimally at temperatures around 80°C and pH values between 2.5 and 3.5. Since the first description of *Sulfolobus acidocaldarius* by Brock et al. (1972), members of the Sulfolobales have developed into model organisms for studying DNA transcription, replication, translation, DNA repair, RNA processing, and cell division. Moreover, most Sulfolobales possess the non-phosphorylated Entner–Douderoff pathway and its regulation upon temperature shifts was the topic of a systems biology approach (Albers et al., 2009). This initiative has led to the development of standard operating procedures for omics approaches in Sulfolobales and should facilitate the exchange and comparability of obtained data (Zaparty et al., 2009).

The first shuttle vectors and a gene deletion method in *S. solfataricus* were based on the β-galactosidase, LacS, as a selection marker (Worthington et al., 2003; Aucelli et al., 2006; Berkner et al., 2007) and this method was successfully employed in a number of studies (Schelert et al., 2004, 2006; Szabo et al., 2007; Zolghadr et al., 2007; Frols et al., 2008; Maaty et al., 2009). However, as the selection is based on minimal media containing only lactose as a carbon source, this method is quite tedious as *Sulfolobus* species do not grow well on sugar minimal media. In some archaeal genetic systems the complementation of uracil auxotrophic mutants by the *pyrEF* genes is used on uracil free media to obtain gene deletion mutants. This marker also enables for the use of 5-fluoroorotic acid (5-FOA), which can be employed as counter selection marker for constructing markerless deletion mutants. For *S. islandicus* a system for obtaining unmarked deletion mutants relying on the *pyrEF* selection including a shuttle vector has been established (Deng et al., 2009; Peng et al., 2009) and successfully used in different studies (Zhang et al., 2010; Gudbergsdottir et al., 2011). Other *Sulfolobus* shuttle vectors that rely on *pyrEF* selection were constructed based on the virus SSV1 and the plasmid pRN1 (Jonuscheit et al., 2003; Berkner et al., 2007). For *S. islandicus* it has furthermore been shown recently that the drug simvastatin can be used to select for the presence of plasmids in a host that overexpresses the *hmg* gene, encoding the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA; Zheng et al., 2012).

Recently, we employed *pyrEF* as a selectable marker to obtain insertional deletion mutants in *S. acidocaldarius* and moreover for persistence of a shuttle vector with an inducible promoter

for expression (Wagner et al., 2009; Berkner et al., 2010). It was earlier demonstrated that *S. acidocaldarius* can efficiently recombine primers down to a length of 14 bp into its genome (Grogan and Stengel, 2008). Taking advantage of this ability Grogan and colleagues showed very recently that short flanking regions of only 40–50 bp incorporated into the 5' end of primers led to site-specific integration of the obtained PCR products containing *pyrE* as a selectable marker into the *S. acidocaldarius* genome (Sakofsky et al., 2011). This method has been used in high-through put screens especially in *Saccharomyces cerevisiae* and opens these possibilities now for *S. acidocaldarius*. However, these methods lead to the consumption of the only available marker and therefore abolish the possibility of complementation in trans and also the construction of double or triple gene mutants.

Therefore we established different efficient methods to obtain unmarked deletion mutants in *S. acidocaldarius* relying on uracil auxotrophy as a selectable marker using the pop in/pop out method. These methods were employed for genomically tagging of genes enabling pull down experiments for studies on protein complexes. In addition, an ABC transporter of *S. solfataricus* was ectopically inserted into the *upsE* locus of *S. acidocaldarius*. Moreover, we present the use of a copper inducible promoter for homologous and heterologous production of proteins. Summarizing, we present here a versatile and complete genetic toolbox for *S. acidocaldarius*.

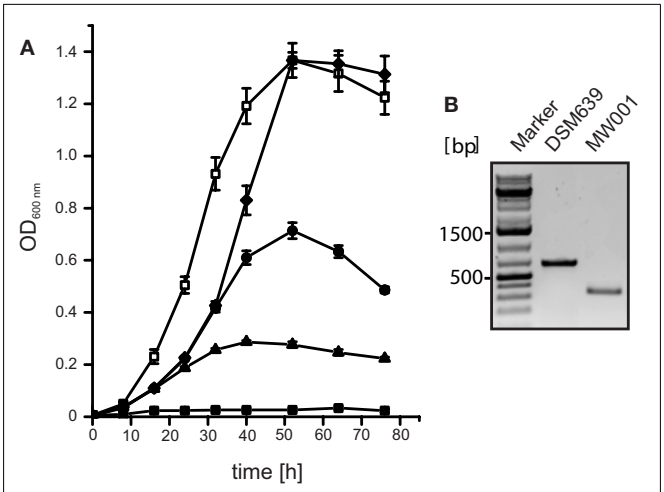


FIGURE 1 | Construction of the uracil auxotrophic strain MW001. (A) Growth curves of MW001 in medium containing 0.2% NZ-amine and 0 (closed rectangles), 1 (closed triangles), 2.5 (closed circles), 5 (closed diamonds), 10 μg/ml (open rectangles) uracil. **(B)** PCR analysis of the wild type *S. acidocaldarius* strain DSM 639 and MW001 using primers 914/915 that amplify 663 bp of the full length *pyrE* and 340 bp from the deletion mutant *pyrE*.

RESULTS

CONSTRUCTION OF MW001

Several different auxotrophic mutants of *S. acidocaldarius* were published before. However, quite a few of these were obtained by UV mutagenesis (Grogan, 1991). Therefore the possibility existed that these mutants would also contain secondary mutations in their genomes. For that reason a new directed *S. acidocaldarius pyrE* deletion mutant was constructed. This mutant, dubbed MW001, was obtained by transforming the *S. acidocaldarius* wild type strain DSM639 with a PCR product that contained 930–1500 bp of the up- and downstream flanking regions of the *pyrE* gene (*saci1597*), which would delete the full length *pyrE* gene by homologous recombination. After transformation the cells were streaked on gelrite plates containing uracil and 5-FOA. Surprisingly, none of the obtained colonies exhibited the expected complete deletion of the *pyrE* gene, but only a deletion of 322 bp (91–412 bp). This phenomenon happened repeatedly and this particular deletion has also been isolated before (Grogan and Hansen, 2003). The deletion of the 322 bp in *pyrE* in MW001 was confirmed by PCR (**Figure 1B**) and sequencing. Growth of MW001 in medium containing NZ-amine medium was only possible upon the addition of uracil and was completely restored at a concentration of 10 μg/ml uracil (**Figure 1A**).

IN-FRAME UNMARKED DELETION MUTANTS

We employed three different methods to obtain in-frame unmarked deletion mutants in *S. acidocaldarius* MW001. The first method is based on the classical "pop in/pop out" scheme using a single crossover recombination step (see **Figure 2A**). We designed pSVA406 for this purpose, which contained the *pyrEF* cassette of *S. solfataricus* ($pyrEF_{SSO}$) and a multiple cloning site upstream of it, which was used to insert the approximately 500 bp long up- and downstream flanking regions of the gene of interest. The obtained deletion mutant plasmid was methylated and 100 ng plasmid DNA was electroporated into MW001. Cells were streaked on first selection plates that contained no uracil to select for the cells that would integrate the plasmid into their genome via a single crossover (**Figure 2A**). To enforce the "pop out" by a second single crossover recombination event, cells were then streaked on second selection plates that contained 5-FOA as only $pyrEF^-$ cells are resistant to 5-FOA. This second single crossover recombination step will either produce the wild type situation or the expected deletion mutant (**Figure 2A**). The method is here illustrated by deleting *ups*E (*saci1494*). In the last screening step we obtained in 50% of the colonies the wild type situation and 50% of analysed colonies we got the expected in-frame deletion mutants (**Figure 3A**). As *ups*E is part of the UV inducible pili operon in *S. acidocaldarius* which encodes proteins that after UV induction mediate cellular aggregation (Ajon et al., 2011), we confirmed that the obtained Δ*ups*E strain could indeed not aggregate anymore upon UV treatment (**Figure 3B**). The correctness of the obtained deletion mutants was confirmed by sequencing of PCR products that were achieved by using primers that were at least 200 bp up and downstream located to the primers used to construct the flanking regions for the deletion plasmid. This deletion method has been successfully used to produce single gene deletion mutants (Ellen et al., 2011; Meyer et al., 2011), and also double and triple deletion mutants as the marker cassette can be reused as often as wanted (Henche et al., 2012; Lassak et al., 2012). These mutants have also been successfully complemented by expression vectors.

To improve the deletion mutant procedure we introduced the reporter gene *lac*S that encodes a β-galactosidase from *S.*

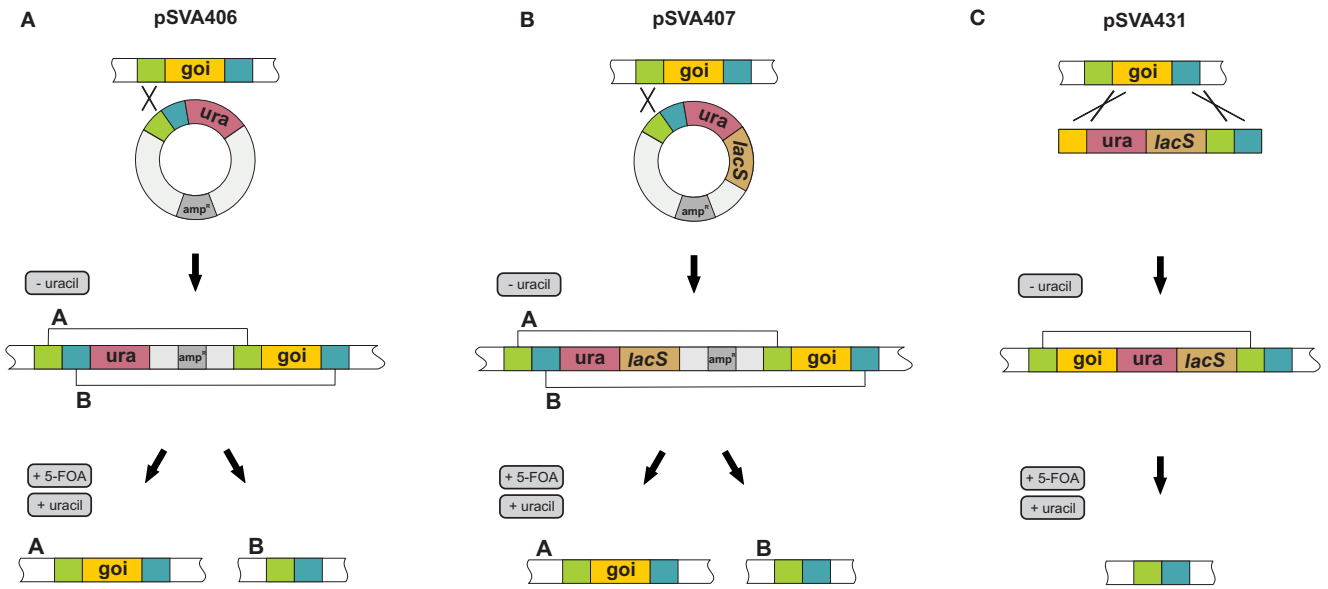


FIGURE 2 | Depiction of the different methods to obtain markerless in-frame mutants. (A) Plasmid pSVA406 integrates via single crossover into the genome and by addition of 5-FOA it can be selected for a second single crossover event that leads either to the wild type situation or the aimed add deletion mutant. **(B)** By using pSVA407 deletion mutants can be obtained via the same method as employed by pSVA406, only that pSVA407 contains the lacS gene in the selection cassette that allows for the easy identification of integrants at the first selection level. **(C)** pSVA431 allows for double crossover insertion via the sequence of the gene and a downstream region of the gene of interest of the selection cassette. Thereby positives clones should be staining blue. Upon selection on 5-FOA plates the selection cassette will be excised and positive deletion mutants should be white. Goi, gene of interest; ura, $pyrEF_{sso}$ cassette; amp, ampicillin resistance cassette for selection in *E.coli*.

solfataricus into pSVA406 yielding the deletion mutant plasmid pSVA407 (**Figure 2B**). Cells expressing *lac*S turn blue when sprayed with substrates like X-gal (5-bromo-4-chloro-indolyl-β-D-galactopyranoside), therefore integrants could be easily detected on the first selection plate. **Figure 3C** shows an example of such a plate, also demonstrating that most of the growing colonies were *lac*S positives.

Furthermore, we developed a third method which yields markerless deletion mutants via double crossover events. The deletion mutant plasmid pSVA431 consists of the $pyrEF_{SSO}$ marker cassette together with *lacS* reporter cassette that are flanked by two different multiple cloning sites. The first multiple cloning site harbors an approximately 500 bp part of the gene of interest and in the second multiple cloning site the up and downstream flanking regions were placed consecutively (**Figure 2C**). This cassette was transformed as a linear fragment into MW001. As the absence of uracil on the first selection plates selects for colonies containing the $pyrEF_{SSO}$ cassette, cells will only be able to form colonies when they integrate the linear fragment with double crossover via the matching part of the gene and the downstream flanking part of the gene (**Figure 2C**). Therefore all obtained colonies on first selection plates by means of this method turned blue after X-gal spraying (data not shown). When these strains were streaked on second selection plates containing 5-FOA and uracil, only the cells that loop the $pyrEF_{SSO}$ cassette and the *lacS* gene via a single crossover of the upstream flanking region out, can form colonies. Consequently, all white colonies that grow on second selection plates are deletion mutants (**Figure 3D**, right half of the gel). One further advantage of this method is that the genotype of integrants still harbors the functional full length gene of interest. If via this technique no colonies can be obtained on second selection plates or if the obtained colonies stay blue after X-gal spraying (which would mean that point mutations in the $pyrEF_{SSO}$ cassette occurred), the target gene can be considered essential for growth under these conditions.

ECTOPIC INTEGRATION OF FOREIGN DNA INTO THE *S. ACIDOCALDARIUS* MUTANT

To demonstrate that we can introduce ectopically DNA sequences into the *S. acidocaldarius* genome, we introduced the glucose transporter of *S. solfataricus* (*sso2847-sso2850*; *glc*S,T,U, and V, respectively, see **Figure 4A**; Albers et al., 1999) into the *ups*E locus (*saci1494*) of MW001. To this end plasmid pSVA445 was constructed containing the whole glucose transporter under the control of the maltose inducible promoter of *mal*E (*saci1665*; Berkner et al., 2010) and the flanking regions that matched the sequence of *upsE*. pSVA445 was methylated and transformed linearized into MW001. Cells were plated on first selection plates. Colonies that had integrated pSVA445 were streaked on second selection plates. Obtained colonies were screened for positive clones that would have integrated the glucose transporter cassette of 4700 bp (glc transporter with promoter; 4450 bp only transporter; 6220 bp glc transporter together with *pyrEF* marker; data not shown) under the control of the maltose promoter. Successful production of the glucose transporter from the *S. acidocaldarius* genome was tested by analyzing wild type cells and the insertion mutant by western blot analysis using antibodies against GlcV, encoded by the last gene of the glucose transporter operon. As shown in **Figure 4B**,

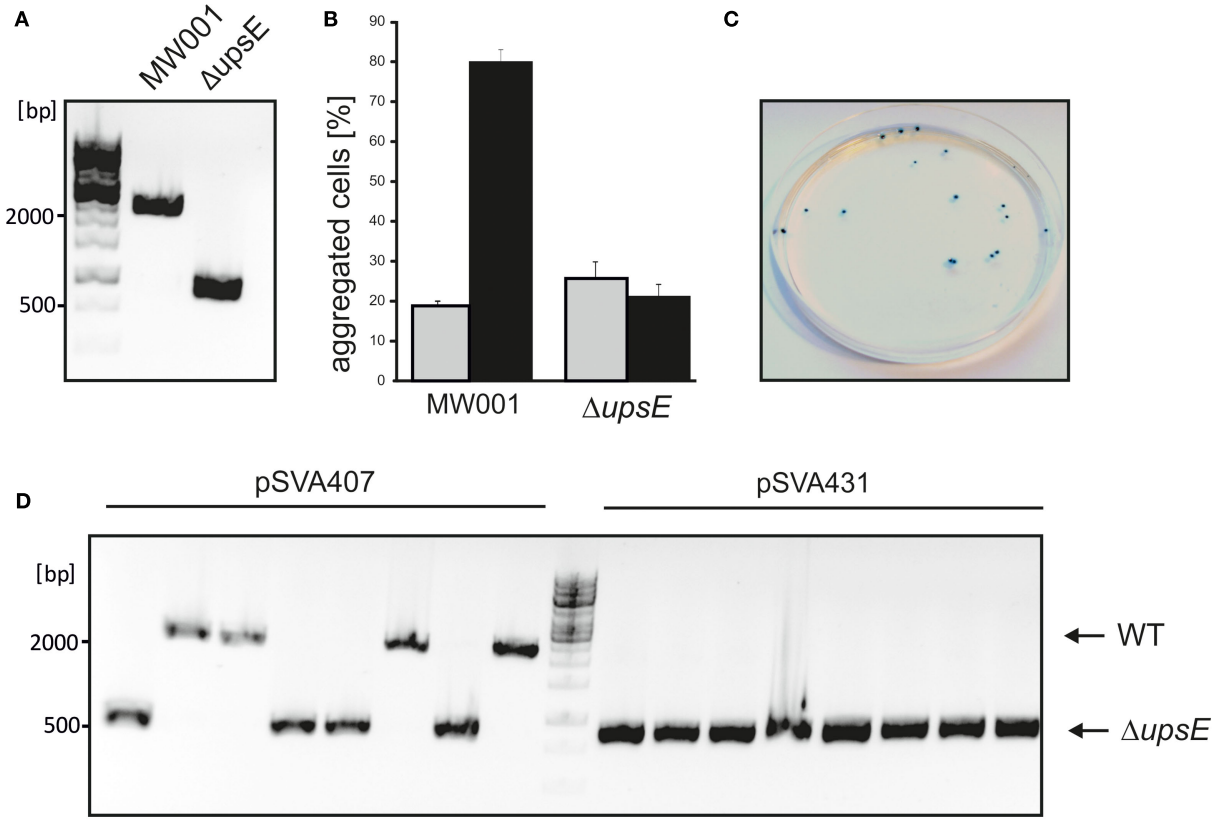


FIGURE 3 | UpsE deletion mutants obtained by three different methods. **(A)** PCR to exemplify the difference in PCR product size obtained by primers 2073/2015 on either MW001 or any of the Δ*upsE* mutants. **(B)** As *upsE* is part of the pili operon that leads to aggregation of UV treated *S. acidocaldarius* cells (Ajon et al., 2011), an aggregation assay was performed with all the Δ*upsE* deletion mutants that were either obtained using pSVA406, pSVA407, or pSVA431. A typical result of an UV induced aggregation assay is shown. Aggregation of wild type and deletion mutant were scored before (gray bar) or after UV treatment (black bar). **(C)** First selection plate colonies from a pSVA407 transformation was sprayed with X-gal. **(D)** A typical result of a mutant PCR screen with primers 2073/2015 using either plasmid pSVA406 or pSVA431 to obtain the deletion mutant.

GlcV was detected in the membrane fraction of the insertion mutant, showing that the *mal* promoter upstream of *glc*S was sufficient to drive expression of the whole ABC transporter operon.

GENOMICALLY TAGGING OF PROTEINS

To enable tag based affinity purification of homologously produced proteins and identification of native protein complexes, we used the "pSVA406" method (**Figure 2A**) to add tags to genes of interest in the genomic context that encoded for, e.g., His- or One-Strep-tags. Two examples will be discussed here, the purification of AglB (Saci_1274), a multiple spanning membrane protein, and of Saci_1210, a cytosolic protein. AglB was tagged in the genome by constructing plasmid pSVA1252 that contained the C-terminal part of *agl*B in which the stop codon was exchanged by a One-Strep-tag coding sequence and the downstream part of *agl*B. Using the pop in/pop out method the tag was inserted into the genome of MW001 and correct insertion was verified by sequencing MW001 *saci1274*::*saci1274*-One-Strep. This strain was grown and cells processed as described in Section "Materials and Methods." Solubilized membrane proteins were subjected to Strep-tag affinity chromatography and the samples were analyzed by SDS-PAGE and immunoblotted with Streptactin antibodies (**Figures 5A,B**). The purified AglB-One-Strep appeared fuzzy, most probably as it is glycosylated, but its identity was confirmed by mass spectrometry. Two other proteins were co-purified that were identified as Saci_0260 and Saci_0262, subunits of the pyruvate carboxylase. These two proteins seem to be biotinylated proteins that bind to the chromatography material. The same proteins have been observed during Strep affinity purifications from cell extracts from *S. solfataricus* (Albers et al., 2006). Binding of these two proteins can be avoided by the addition of avidin to the cytoplasmic fraction.

In contrast to AglB, Saci_1210 was tagged with a tandem tag containing a Strep and a 6xHis-tag. This enabled us to test purification by either Strep- or His affinity chromatography using the same strain. Strep affinity chromatography of Saci_1210-Strep/His resulted in the same two contamination bands seen also in the AglB purification and no detectable Saci_1210 (data not shown). His-tag affinity chromatography resulted in highly pure Saci_1210-Strep/His in one step (**Figures 5C,D**). However, in contrast to *S. solfataricus* *S. acidocaldarius* exhibits a 14.7 kDa hypothetical protein, Saci_0386, that contains a natural stretch of six histidine residues at the C-terminus and therefore this protein co-purifies during His affinity chromatography (not shown in this blot). Concluding, His-tag affinity chromatography seems to be better

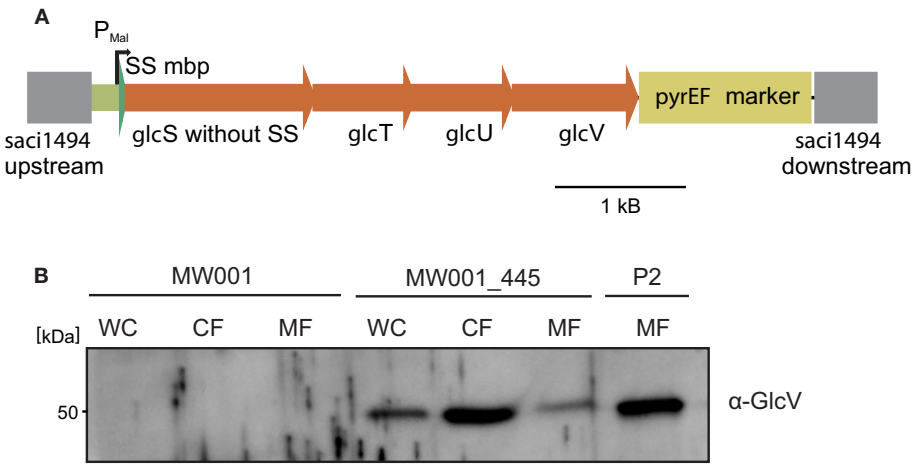


FIGURE 4 | Insertion and expression of the *S. solfataricus* ABC glucose transporter in the *S. acidocaldarius* genome. (A) Depiction of the genomic integration of the glucose ABC transport operon of *S. solfataricus* in the *upsE locus* in the *S. acidocaldarius* genome. The signal peptide of GlcS, the glusoce binding protein, was exchanged by the MalE signal peptide from *S. acidocaldarius*. GlcTU are the predicted permase, whereas GlcV is the ATP binding protein in the glucose ABC transporter. **(B)** Western blot with specific GlcV antibodies on MW001, MW001_445, and *S. solfataricus* P2 cells. WC, whole cells; CF, cytoplasmic fraction; MF, membrane fraction.

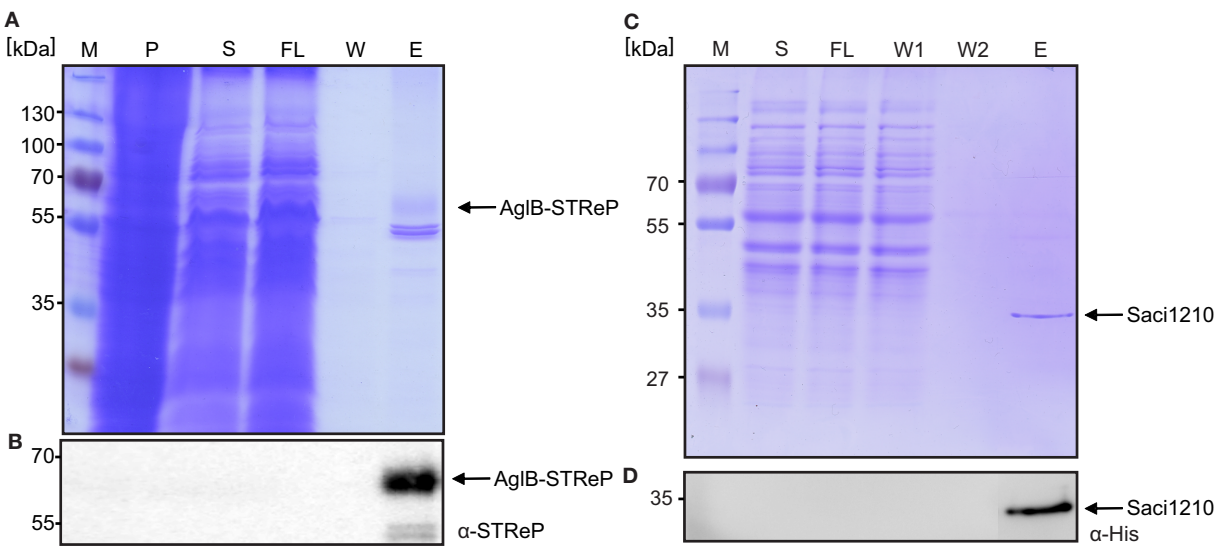


FIGURE 5 | Purification of the genomic c-terminal Strep-Saci1274 fusion protein. (A) Commassie stained SDS-PAGE and **(B)** immuno blot with STRep-tag antibodies of the purification of Strep-Saci1274 fusion protein from *S. acidocaldarius* cells using Streptactin affinity material. **(C)** Commassie SDS-PAGE and an immune blot with a-His-tag antibodies **(D)** of the purification of Saci1210 which was genomically tagged using His_SELECT material. M, marker; P, insoluble membrane pellet; S, solubilized membrane proteins; FL, flow through; W, wash fraction; E, elution fraction.

suited for isolation of proteins from *S. acidocaldarius* as Strep-tag isolation leads to the co purification of two very prominent contaminating proteins of around 50 kD.

INDUCIBLE PROMOTERS

Several promoters have already been tested for efficient production of proteins in *S. acidocaldarius* of which the maltose inducible promoter of the maltose binding protein (*saci1165*) turned out to be the most reliable one (Berkner et al., 2010; Meyer et al., 2011; Henche et al., 2012; Lassak et al., 2012). Still this promoter exhibits quite some basic expression even under non-induced conditions. As a possible alternative we examined the copper inducible promoter of the copper inducible copper resistance cassette (*saci0874–saci0872*) that encodes CopT, the regulator of the operon, CopM, a metallochaperon, and CopA, a P-type copper ATPase (Ettema et al., 2006; **Figure 6A**). On the one hand CopT was described to be a negative regulator (Ettema et al., 2006) and on the other hand in a recent study (Villafane et al., 2011) CopT (called CopR in that publication) was described to be an activator of the *cop* operon. Therefore, the *mal* promoter of pCmalLacS was replaced

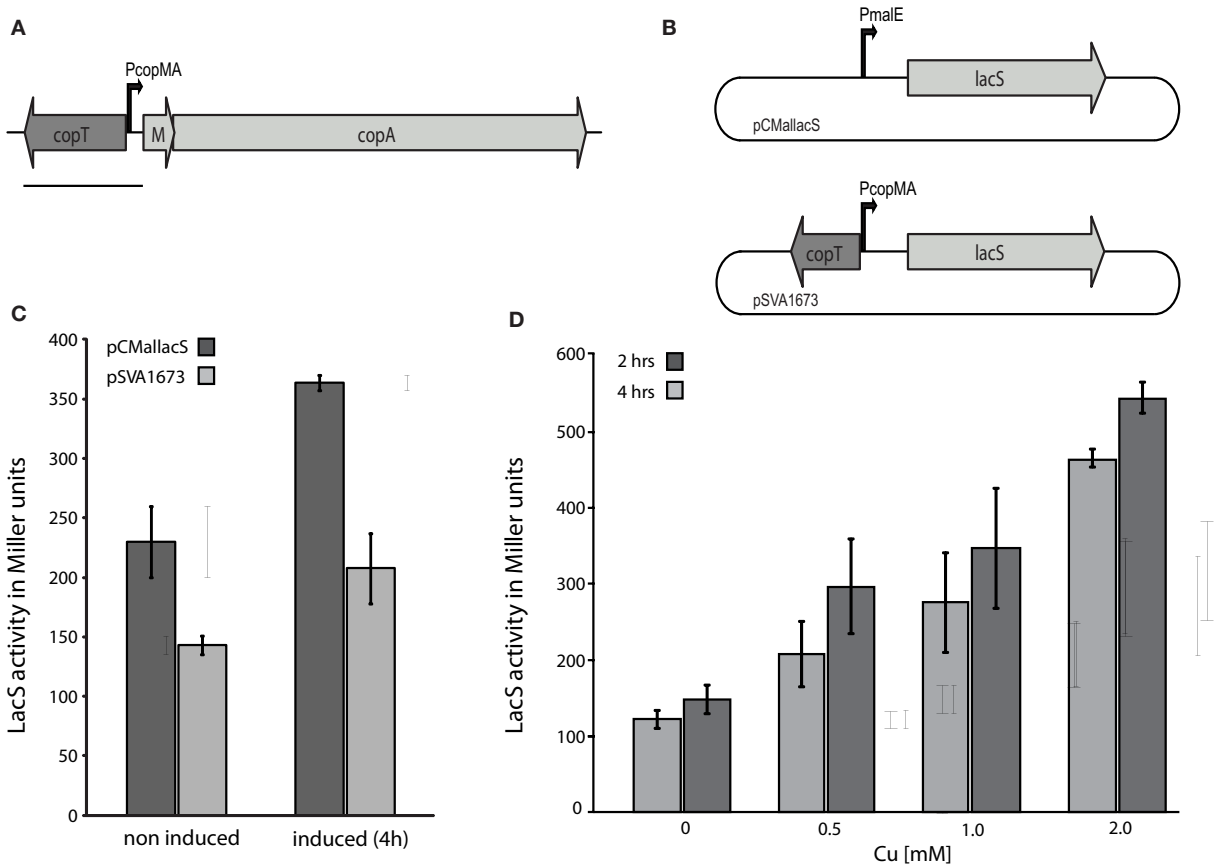


FIGURE 6 | Comparison of copper inducible promoter and maltose inducible promoter for expression in MW001. (A) Operon structure of the copper resistance cluster of *S. acidocaldarius*. Black line indicates the part of the operon that was included into pSVA1673. **(B)** Schematic representation of the plasmids pCMallacS and pSVA1673 used in **(C,D)**. **(C)** Direct comparison of LacS expression under control of Pmal (pCMallacS; dark gray) or PcopMA (pSVA1673; light gray) in MW001 cells under non-induced and induced conditions (0.4% maltose or 1 mM Cu, respectively). **(D)** Specific LacS activity by induction of PcopMA with increasing amounts of Cu in the medium. For each concentration the LacS activity was measured after 2 (light gray) and 4 h (dark gray) of induction.

with the c*opMA* promoter region together with *copT* (highlighted in **Figure 6A** by a gray line, plasmids are schematically shown in **Figure 6B**). Hence, LacS expression would be driven by the *copMA* promoter (**Figure 6A**). The regulatory gene *copT* was included to ensure correct repression/activation of the *copMA* promoter.

The pSVA1673 was transformed into MW001 which was grown in medium supplemented with a minimal trace element solution to ensure very low levels of copper. The basic levels of expression were half of that obtained by the *mal* promoter under comparable conditions (**Figure 6C**). However, the induction of the promoter was only 1.4 fold after 4 h at a copper concentration of 1 mM whereas the maltose promoter was induced up to 1.8 fold. The copper promoter could be induced gradually by varying the amount of copper present in the medium (**Figure 6D**). Copper concentrations higher than 2 mM were not suitable because of cell lysis.

DISCUSSION

Detailed *in vivo* studies were long hampered in archaea as the development of genetic systems was very often obstructed by the absence of plasmids or selection markers. However, in the last few years enormous progress has been made and for most of the archaeal model systems genetic tools now exist (Leigh et al., 2011). Especially in the Sulfolobales genetic systems have been developed. For *S. solfataricus* and *S. islandicus*, respectively (Worthington et al., 2003; Albers and Driessen, 2008; She et al., 2009) these are proven to be successful as work based on them has been published.

Here we present a complete and versatile genetic system for *S. acidocaldarius*. Our system advances the system developed by Grogan that is based on marker insertion into the target gene (Sakofsky et al., 2011), to a system in which the selection marker can be reused. This therefore enables us to construct multiple allele mutants (single, double, and triple), specifically mutate genes in the genomic background or tag proteins genomically and complement obtained mutants by using an expression vector. In this manuscript we showed exemplary three different methods that can be employed to obtain markerless deletion mutants. We have used the same method for different projects in which we deleted target genes, constructed single, double, and triple mutants and complemented these using the pCmal based vector system (Berkner et al.,

2007; Ajon et al., 2011; Meyer et al., 2011; Henche et al., 2012; Lassak et al., 2012).

We have also experienced that some genes could not be deleted by the pop in/pop out method, but we were able to obtain a deletion mutant by the PCR based insertion method (Sakofsky et al., 2011). This might be due to regulatory elements that might be disturbed by the insertion method. Therefore both methods can and should be used complementary.

We demonstrated that we can genomically tag genes *in cis* which will facilitate protein complex isolation and testing protein truncates or point mutations in their original genomic neighborhood. For affinity purification of such tagged proteins, His-tag affinity chromatography in our experience yielded better purification results, which could of course be due to a better accessible tag as the His-tag was located at the outmost C-terminus of the protein whereas the Strep-tag might have been occluded. However, His-tag affinity purification only yields one co purifying impurity. In the future other tags as, e.g., FLAG and HA tag will be tested to expand the genetic tool box.

Moreover, we introduced a ∼5 kB ABC transporter operon from *S. solfataricus* into the *S. acidocaldarius* genome and could demonstrate that the last gene was transcribed and translated correctly. This opens the option of integrating large sequences into the *S. acidocaldarius* genome and therefore streamlines its application in synthetic biology approaches.

For the expression in *S. acidocaldarius* we also tested the copper inducible *copMA* promoter, which showed decreased basic expression levels in comparison to the maltose inducible promoter. Therefore the *copMA* promoter might be useful for production of toxic proteins in *S. acidocaldarius*, especially regarding that the expression levels can be tightly titrated by the amount of copper added to the medium.

Concluding, we present here a very robust and versatile genetic tool box for *S. acidocaldarius* that has already proven to be widely applicable and very useful in delineating gene functions in this organism. We are currently working on the development of additional selective markers and the optimization of the expression vector.

MATERIALS AND METHODS

STRAINS AND GROWTH CONDITIONS

Sulfolobus acidocaldarius DSM 639, MW001, and all constructed deletion mutants were aerobically grown in Brock media (Brock et al., 1972) with a pH of 3 at 76°C. The media were supplemented with 0.1% (w/v) tryptone or with 0.1% (w/v) N-Z-Amine and 0.2% dextrine. The growth of the cells was monitored by measurement of the optical density at 600 nm.

For pouring *Sulfolobus* plates a two times concentrated Brock media was supplemented with 6 mM $CaCl_2$ and 20 mM $MgCl_2$. For first selection plates 0.2% NZ-Amine (Fluka) and 0.4% dextrin, for second selection plates 0.2% tryptone, 0.4% dextrin, 200 µg/ml 5-FOA, and 20 µg/ml uracil was added to the two times concentrated solution and prewarmed to 75°C. This solution was mixed with an equal volume of fresh boiling 1.4% Gelrite solution (Carl Roth, Karlsruhe, Germany) and poured in 40 ml portions into petri dishes (150 × 20 mm, Sarstedt, Nümbrecht, Germany).

CONSTRUCTION OF DELETION MUTANT PLASMIDS

To obtain an uracil auxotrophic mutant of *S. acidocaldarius* DSM 639 deletion plasmid pSVA402 was constructed. To this end the up- and downstream regions of *pyrE* (*saci1597*) were amplified using primers 390/916 and 917/920, respectively. An overlap extension PCR was performed with both PCR products and the resulting *pyrE* deletion fragment was blunt end cloned into pGEM-T Easy (Promega, Mannheim, Germany) yielding pSVA402.

For the construction of the deletion plasmid pSVA406 the *pyrEF* (*sso0615–sso0616*) cassette of *S. solfataricus* together with its own promoter region was amplified using primers 938 and 939 (primer sequences available in the Appendix). This marker cassette was cloned into the *Mlu*I *Nsi*I site of the recirculated pGEM-T Easy vector. Primer 938 contained additional *Bam*HI and *Avr*II sites to expand the multiple cloning site of the resulting plasmid.

For construction of the deletion plasmid pSVA407 the *lac*S (*sso3019*) gene of *S. solfataricus*, encoding a β-galactosidase, was amplified with primers 933 and 940 and fused to the maltose inducible promoter of the maltose binding protein (*saci1165*) of *S. acidocaldarius*, which was amplified with primers 937 and 932, via overlap extension PCR. The resulting reporter cassette was cloned into the *Nsi*I site of pSVA407. A correctly oriented clone was identified by restriction analysis.

The deletion plasmid pSVA431 was constructed by amplifying the *pyrEF* (*sso0615–sso0616*) and *lacS* (*sso*3019) cassette of *S. solfataricus* from pSVA407 using primer 939 and 1072. The resulting marker – reporter cassette was cloned into the pGEM-T Easy vector. The pSVA431 exhibited two multiple cloning sites up and downstream to the marker – reporter cassette, respectively.

For the construction of the *upsE* (*saci1949*) deletion plasmids via single crossover event the upstream region of *upsE* was amplified with the primers 2010 and 2011 and the downstream region was amplified with the primers 2012 and 2013. With both PCR products an overlap extension PCR was performed and cloned into the *Apa*I *Bam*HI sites of pSVA406 and pSVA407 leading to the plasmids pSVA1804 and pSVA447. For the construction of the *upsE* deletion plasmid via double crossover the *upsE* gene was amplified with primers 532 and 639. The resulting PCR product was cloned into the *Bam*HI and *Pst*I site of the MCSI of pSVA431. A PCR was performed on pSVA1804 with the outer primers 2264 and 2265 to introduce different restriction sites and cloned into the *Nco*I and *Kpn*I site of MCSII.

For the construction of the ectopic integration plasmid of the *S. solfataricus* glucose transporter into *S. acidocaldarius* pSVA445 the *upsE* (*saci*1949) upstream region was amplified using primers 2256 and 2257. The promoter region together with the signal sequence of the maltose binding protein (*saci*1165) was amplified employing primers 2258 and 986. *glcS* (sso2847) without signal sequence was amplified using the primers 987 and 988. *glcT* (sso2848), *glcU* (sso2849), and *glcV* (sso2850) were amplified in a single PCR with the primers 989 and 2259. The *pyrEF* marker cassette of *S. solfataricus* was amplified with the primers 2260 and 2261 and for the amplification of the *upsE* downstream region the primers 2262 and 2263 were used. All six PCR products were mixed together to perform an overlap extension PCR. The resulting PCR product was blunt end cloned into pGEM-T Easy following the standard protocol yielding pSVA445.

To simplify the purification of *saci1274* by affinity chromatography, a One-Strep-tag was fused to this gene in the genome. For the construction of the insertion plasmid 900–1000 bp fragments of the up- and downstream regions of *saci1274* were PCR amplified. For the upstream region the forward primer introduced an *Apa*I restriction site at the 5′end, whereas the reversed primer was designed to incorporate the One-Strep-tag sequence in front of the Stop codon (primers 1818 and 1819, respectively). For the downstream fragment the forward primer 1820 was designed to incorporate the complementary strand of the One-Strep-Tag, while the reverse primer 1821 contained a *Bam*HI restriction site. Both fragments were fused via an overlapping PCR and the amplified PCR fragment was digested with *Apa*I and *Bam*HI and ligated into plasmid pSVA407 yielding plasmid pSVA1224. The correct sequence was confirmed by sequencing. To increase the accessibility of the One-Strep-tag a longer linker sequence was added leading to the One-Strep sequence that was synthesized with a flanking 5′ *Acc*I and 3′ *Pst*I restriction site (sequence present in **Figure A1** in Appendix). The product was digested with *Acc*I and *Pst*I and ligated into the plasmid pSVA1224, predigested with the same restriction enzymes, yielding the plasmid pSVA1252. Transformation and selection of the mutant strain was performed as described above.

To obtain genomically His-tagged saci1210 primers 1583 and 3106 were used to amplify the gene together with the C-terminal tag (Strep/His-tag). For the downstream region amplification with the primers 3107 and 1578 resulted in a PCR product with overlapping regions to the gene together with the tag. The overlap extension PCR product was cloned into the *Nco*I, *Bam*HI site of pSVA406 yielding plasmid pSVA1097.

For the construction of copper inducible expression plasmid the promoter region of *copM* (*saci0873*) was amplified together with the predicted regulator *copT* (*saci0874*) with the primers 3428 and 3429. The resulting PCR product was cloned into the *Sac*II, *Nco*I site of pCMalLacS which led to an exchange of the maltose inducible promoter of the maltose binding protein with the copper inducible promoter together with its regulator yielding pSVA1673.

PREPARING COMPETENT MW001 CELLS

Sulfolobus acidocaldarius MW001 were grown in 50 ml Brock media supplemented with 0.1% NZ-Amine, 0.2% sucrose, and 20 μg/ml uracil and adjusted to pH 3 with sulfuric acid. When the culture reached an $OD_{600\,nm}$ of 0.5–0.7 (exponential growth phase) an aliquot was transferred to 400 ml fresh medium and harvested at an $OD_{600\,nm}$ of 0.2–0.3. The culture was cooled down on ice, then centrifuged for 15 min at 4000 *g* and washed twice with 250 ml ice cold 20 mM sucrose. The pellet was resuspended with ice cold 20 mM sucrose to a theoretical final $OD_{600\,nm}$ of 10 and aliquotted in 50 μl portions. The aliquots were directly used for transformation or frozen at −80˚C without using liquid nitrogen for storage.

TRANSFORMATION OF PLASMIDS INTO *S. ACIDOCALDARIUS*

Prior to transformation into *S. acidocaldarius*, suicide- and shuttle plasmids were methylated to prevent restriction by the *SuaI* restriction enzyme. For that purpose *E. coli* ER1821 (New England Biolabs) bearing the additional plasmid pM.EsaBC4I (New England Biolabs, Frankfurt am Main, Germany) was transformed with plasmid DNAs.

Methylated deletion plasmids were electroporated in electrocompetent wild type cells MW001 using a Gene Pulser Xcell (BioRad, München, Germany) with a constant time protocol with input parameters 1.5 kV, 25 μF, 600 Ω in 1 mm cuvettes. Before plating on uracil lacking and NZ-Amine containing plates, cells were regenerated for 30 min at 75˚C in two-fold recovery solution (1% sucrose, 20 mM β-alanine, 1.5 mM malate buffer, pH 4.5, 10 mM $MgSO_4$). The plates were sealed in plastic bags to avoid drying-out and incubated for around seven days at 75˚C.

COLONY PCR OF *S. ACIDOCALDARIUS*

Single colonies appearing on plates were analyzed by colony PCR. To that end, single colonies were picked and lysed in 30 μl 0.2 M NaOH and the solution neutralized with 70 μl 0.2 M Tris pH 7.8. To amplify the genomic region of interest 0.5 μl lysate was used in 30 μl PCR reactions using Phusion High-Fidelity polymerase with Phusion HF buffer and monitored on an agarose gel.

BLUE–WHITE SCREENING

Integration of pSVA407 or pSVA431 constructs could be visualized by X-Gal spraying of cells using a 25 mg/ml X-Gal stock solution in DMF diluted 1:5 with water. The 5 mg/ml X-Gal solution was sprayed on plates when single colonies appeared after around one week and the plates were directly put back into the 75˚C incubator for 30 min. Transformants on first selection plates and point mutants on second selection plates turned blue while deletion mutants on second selection plates stayed white.

PURIFICATION OF SACI1274-ONE-STrEP

Ten liter culture of the strain MW001 Saci1274::Saci1274-One-Strep was grown in Brock medium until an OD of 0.8. Cells were harvested by centrifugation (3000 *g*; 4˚C; 20 min). The cell pellet was resuspended in 40 ml buffer A (100 mM NaCl, 100 mM Tris HCl, 1 mM EDTA, pH 8) and lysed by a 20 min sonification with an intensity of 60% and an interval 20 s (Bandelin Sonopuls). Unbroken cells were removed by a low spin centrifugation 3000 *g* at 4˚C for 20 min. The supernatant was centrifuged at 120,000 *g* at 4˚C for 45 min and the membrane pellet resuspended in 12 ml of buffer A. 6 ml of the membrane fraction solubilized at 42˚C under shaking condition in 30 ml Buffer S (2% n-dodecyl beta-D-maltoside (DDM), 100 mM NaCl, 100 mM Tris HCl, 1 mM EDTA, pH 8) supplemented with PMSF. Undissolved membranes were pelleted by an ultracentrifugation (120,000 *g*; 4˚C for 45 min). The supernatant was added twice to a 0.8 ml STREP-Tag column (Strep-Tactin®Superflow®, IBA, Goettingen, Germany). Before loading the column was equilibrated with 10× column volume of buffer E (0.05% DDM, 100 mM NaCl, 100 mM Tris HCl, 1 mM EDTA, pH 8) and after the loading of the supernatant washed with 15× column volumes of the same buffer. The fusion protein was eluted with the buffer E containing 1 mM desthiobiotin.

Purification of Saci-Strep/His was performed with the difference that the cytoplasm was used for purification. For His affinity chromatography His-SELECT material from Sigma was used.

WESTERN BLOT ANALYSIS

From each purification step 30 µl samples were loaded on a 11% SDS-PAGE and run at 100 V. Transfer to a PVDF membrane and blotting were performed as commonly done. The generated chemifluorescence of the Precision StrepTactin-AP Conjugate antibody (Biorad) or His-AP (Abcam, Cambridge, UK) was measured in a Fujifilm LAS-4000 Luminescent image analyzer (Fujifilm, Duesseldorf, Germany).

PROMOTER ACTIVITY ASSAY

For the promoter activity assay pSVA1673 and pCmalLacS were transformed into MW001. Single colonies containing pCmalLacs were inoculated in Brock Medium with 0.1% NZ-Amine and 0.2% sucrose whereas pSVA1673 containing cells were grown in Brock medium supplemented with 0.2% maltose and 0.2% xylose. Moreover, for these cells the trace element solution only contained $Na_2B_4O_7$ and $MnCl_2$. For induction, 0.4% maltose or 1 mM $CuSO_4$ were added to the medium, respectively.

Promoter activity was analysed by measuring the cleavage of o-nitrophenyl-β-D-galactopyranosid (ONPG) by LacS of *S. solfataricus* as described in Lassak et al. (2012). All given strains were assayed in biological as well as technical triplicates. The production of ONP was measured at 410 nm over a period of 4 h at 42°C using an Infinite 200 luminometer (Tecan, Maennedorf, Switzerland). Miller units were calculated as described previously (Lassak et al., 2012). With respect to the temperature optimum of LacS a conversion factor FT = 7 was determined and included in the final equation.

ACKNOWLEDGMENTS

Michaela Wagner, Marleen van Wolferen, and Sonja-Verena Albers were supported by intramural funds of the Max Planck Society. Alexander Wagner received funding from the BMBF in the network of ExpresSyS (0315586D). Kerstin Lassak was supported by the International Max Planck Research School (IMPRS).

REFERENCES

Ajon, M., Frols, S., van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A. J., Grogan, D. W., Albers, S. V., and Schleper, C. (2011). UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817.

Albers, S. V., Birkeland, N. K., Driessen, A. J., Gertig, S., Haferkamp, P., Klenk, H. P., Kouril, T., Manica, A., Pham, T. K., Ruoff, P., Schleper, C., Schomburg, D., Sharkey, K. J., Siebers, B., Sierocinski, P., Steuer, R., van der Oost, J., Westerhoff, H. V., Wieloch, P., Wright, P. C., and Zaparty, M. (2009). SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model for the central carbohydrate metabolism of the archaeon *Sulfolobus solfataricus* under temperature variation. *Biochem. Soc. Trans.* 37, 58–64.

Albers, S. V., and Driessen, A. J. M. (2008). Conditions for gene disruption by homologous recombination of exogenous DNA into the *Sulfolobus solfataricus* genome. *Archaea* 2, 145–149.

Albers, S. V., Elferink, M. G., Charlebois, R. L., Sensen, C. W., Driessen, A. J., and Konings, W. N. (1999). Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. *J. Bacteriol.* 181, 4285–4291.

Albers, S. V., Jonuscheit, M., Dinkelaker, S., Urich, T., Kletzin, A., Tampe, R., Driessen, A. J. M., and Schleper, C. (2006). Production of recombinant and tagged proteins in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Appl. Environ. Microbiol.* 72, 102–111.

Albers, S. V., and Meyer, B. H. (2011). The archaeal cell envelope. *Nat. Rev. Microbiol.* 9, 414–426.

Aucelli, T., Contursi, P., Girfoglio, M., Rossi, M., and Cannio, R. (2006). A spreadable, non-integrative and high copy number shuttle vector for *Sulfolobus solfataricus* based on the genetic element pSSVx from *Sulfolobus islandicus*. *Nucleic Acids Res.* 34, e114.

Berkner, S., Grogan, D., Albers, S. V., and Lipps, G. (2007). Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren-)archaea. *Nucleic Acids Res.* 35, e88.

Berkner, S., Wlodkowski, A., Albers, S. V., and Lipps, G. (2010). Inducible and constitutive promoters for genetic systems in *Sulfolobus acidocaldarius*. *Extremophiles.* 14, 249–259.

Brock, T. D., Brock, K. M., Belly, R. T., and Weiss, R. L. (1972). *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* 84, 54–68.

Deng, L., Zhu, H., Chen, Z., Liang, Y. X., and She, Q. (2009). Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus*. *Extremophiles* 13, 735–746.

Ellen, A. F., Rohulya, O. V., Fusetti, F., Wagner, M., Albers, S. V., and Driessen, A. J. (2011). The sulfolobicin genes of *Sulfolobus acidocaldarius* encode novel antimicrobial proteins. *J. Bacteriol.* 193, 4380–4387.

Ettema, T. J., Brinkman, A. B., Lamers, P. P., Kornet, N. G., de Vos, W. M., and van der Oost, J. (2006). Molecular characterization of a conserved archaeal copper resistance (cop) gene cluster and its copper-responsive regulator in *Sulfolobus solfataricus* P2. *Microbiology* 152, 1969–1979.

Frols, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., Boekema, E. J., Driessen, A. J., Schleper, C., and Albers, S. V. (2008). UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. *Mol. Microbiol.* 70, 938–952.

Grogan, D. W. (1991). Selectable mutant phenotypes of the extremely thermophilic archaebacterium *Sulfolobus acidocaldarius*. *J. Bacteriol.* 173, 7725–7727.

Grogan, D. W., and Hansen, J. E. (2003). Molecular characteristics of spontaneous deletions in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 185, 1266–1272.

Grogan, D. W., and Stengel, K. R. (2008). Recombination of synthetic oligonucleotides with prokaryotic chromosomes: substrate requirements of the *Escherichia coli*/lambdaRed and *Sulfolobus acidocaldarius* recombination systems. *Mol. Microbiol.* 69, 1255–1265.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J. V., Jensen, L. R., She, Q., and Garrett, R. A. (2011). Dynamic properties of the *Sulfolobus* CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79, 35–49.

Henche, A. L., Koerdt, A., Ghosh, A., and Albers, S. V. (2012). Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. *Environ. Microbiol.* 14, 779–793.

Jonuscheit, M., Martusewitsch, E., Stedman, K. M., and Schleper, C. (2003). A reporter gene system for the hyperthermophilic archaeon *Sulfolobus solfataricus* based on a selectable and integrative shuttle vector. *Mol. Microbiol.* 48, 1241–1252.

Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S. V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124.

Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain Archaea: Methanogens, Halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol. Rev.* 35, 577–608.

Maaty, W. S., Wiedenheft, B., Tarlykov, P., Schaff, N., Heinemann, J., Robison-Cox, J., Valenzuela, J., Dougherty, A., Blum, P., Lawrence, C. M., Douglas, T., Young, M. J., and Bothner, B. (2009). Something old, something new, something borrowed; how the thermoacidophilic archaeon *Sulfolobus solfataricus* responds to oxidative stress. *PLoS ONE* 4, e6964. doi:10.1371/journal.pone.0006964

Meyer, B. H., Zolghadr, B., Peyfoon, E., Pabst, M., Panico, M., Morris, H. R., Haslam, S. M., Messner, P., Schaffer, C., Dell, A., and Albers, S. V. (2011). Sulfoquinovose synthase – an important enzyme in the N-glycosylation pathway of *Sulfolobus acidocaldarius*. *Mol. Microbiol.* 82, 1150–1163.

Peng, N., Xia, Q., Chen, Z., Liang, Y. X., and She, Q. (2009). An upstream activation element exerting differential transcriptional activation on an archaeal promoter. *Mol. Microbiol.* 74, 928–939.

Sakofsky, C. J., Runck, L. A., and Grogan, D. W. (2011). *Sulfolobus* mutants, generated via PCR products, which lack putative enzymes of UV photoproduct repair. *Archaea* 2011, 864015.

Santangelo, T. J., Cubonova, L., and Reeve, J. N. (2008). Shuttle vector expression in *Thermococcus kodakaraensis*: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Appl. Environ. Microbiol.* 74, 3099–3104.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2003). Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.* 185, 210–220.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T. (2005). Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.* 71, 3889–3899.

Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M., and Blum, P. (2004). Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J. Bacteriol.*186, 427–437.

Schelert, J., Drozda, M., Dixit, V., Dillman, A., and Blum, P. (2006). Regulation of mercury resistance in the crenarchaeote *Sulfolobus solfataricus*. *J. Bacteriol.* 188, 7141–7150.

She, Q., Zhang, C., Deng, L., Peng, N., Chen, Z., and Liang, Y. X. (2009). Genetic analyses in the hyperthermophilic archaeon *Sulfolobus islandicus*. *Biochem. Soc. Trans.* 37, 92–96.

Szabo, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S. V., Blum, P., Boekema, E. J., and Driessen, A. J. (2007). Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 189, 4305–4309.

Villafane, A., Voskoboynik, Y., Ruhl, I., Sannino, D., Maezato, Y., Blum, P., and Bini, E. (2011). CopR of *Sulfolobus solfataricus* represents a novel class of archaeal-specific copper-responsive activators of transcription. *Microbiology* 157, 2808–2817.

Waege, I., Schmid, G., Thumann, S., Thomm, M., and Hausner, W. (2010). Shuttle vector-based transformation system for *Pyrococcus furiosus*. *Appl. Environ. Microbiol.* 76, 3308–3313.

Wagner, M., Berkner, S., Ajon, M., Driessen, A. J., Lipps, G., and Albers, S. V. (2009). Expanding and understanding the genetic toolbox of the hyperthermophilic genus *Sulfolobus*. *Biochem. Soc. Trans.* 37, 97–101.

Worthington, P., Hoang, V., Perez-Pomares, F., and Blum, P. (2003). Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 185, 482–488.

Zaparty, M., Esser, D., Gertig, S., Haferkamp, P., Kouril, T., Manica, A., Pham, T. K., Reimann, J., Schreiber, K., Sierocinski, P., Teichmann, D., van Wolferen, M., von Jan, M., Wieloch, P., Albers, S. V., Driessen, A. J., Klenk, H. P., Schleper, C., Schomburg, D., van der Oost, J., Wright, P. C., and Siebers, B. (2009). "Hot standards" for the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Extremophiles* 14, 119–142.

Zhang, C., Guo, L., Deng, L., Wu, Y., Liang, Y., Huang, L., and She, Q. (2010). Revealing the essentiality of multiple archaeal pcna genes using a mutant propagation assay based on an improved knockout method. *Microbiology* 156, 3386–3397.

Zheng, T., Huang, Q., Zhang, C., Ni, J., She, Q., and Shen, Y. (2012). Development of a simvastatin selection marker for a hyperthermophilic acidophile, *Sulfolobus islandicus*. *Appl. Environ. Microbiol.* 78, 568–574.

Zolghadr, B., Weber, S., Szabo, Z., Driessen, A. J., and Albers, S. V. (2007). Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus*. *Mol. Microbiol.* 64, 795–806.

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

Received: 06 March 2012; paper pending published: 09 May 2012; accepted: 24 May 2012; published online: 13 June 2012.

Citation: Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J and Albers S-V (2012) Versatile genetic tool box for the crenarchaeote Sulfolobus acidocaldarius. Front. Microbio. ***3****:214. doi: 10.3389/fmicb.2012.00214*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

APPENDIX

Table A1 | Primers used in this study.

Primer	Sequence (5′–3′)	Purpose
PRIMERS FOR pSVA402		
390	GGTACACCGTAAGAGGCTTGTTC	Δ*pyrE* upstr fw
916	GCTAAAATTATTCTAGCTTTTTCCAATTTAACCCTATTTAAACTGAAGGC	Δ*pyrE* upstr rev ol
917	GCCTTCAGTTTAAATAGGGTTAAATTGGAAAAAGCTAGAATAATTTTAGC	Δ*pyrE* downstr fw ol
920	GAGACCGAAAGAGAGCCC	Δ*pyrE* downstr rev
914	CGCCCTTAAATAAGGTTAGTC	Δ*pyrE* check primer fw
915	CTAGCTTTTTCCAATATTTTTCACC	Δ*pyrE* check primer rev
PRIMERS FOR pSVA406, pSVA407, and pSVA431		
938	CCCGATATGCATGACCGGCTATTTTTTCAC	*pyrEF* cassette SSO rev *Nsi*I
939	CCGATACGCGTACTGGATCCTGACCTAGGTTTGAGCAGTTCTAGTACTTG	*pyrEF* cassette SSO fw *Mlu*I, *Bam*HI, *Avr*II
937	CCCGATATGCATGCTTATCTTTTTTTACCTACTCCTTGTGG	*Saci mbp* promoter fw *Nsi*I
932	GCTATTTGGAAATGAGTACATCGGGTTAACTTAATCACGTAATTATTTATAAAC	*Saci mbp* promoter rev ol to *Sso lacS*
933	CGTGATTAAGTTAACCCGATGTACTCATTTCCAAATAGCTTTAGG	*Sso lacS* fw ol to *Saci mbp* promoter
940	CCCTAATGCATTTAGTGCCTTAATGGCTTTACTGGAGG	*lacS* cassette *SSO* reverse *NsiI*
PRIMERS FOR ΔupsE WITH DIFFERENT METHODS		
2010	GTAGGGCCCGTGTATAATGATGACCTATTTAGCTG	Δ*upsE* upstr fw *Apa*I
2011	CTAATATTTTCAAGCCATAAGAAGGAAATATTAAAAG	Δ*upsE* upstr rev ol
2012	CTTCTTATGGCTTGAAAATATTAGCATGTGATATATTC	Δ*upsE* downstr fw ol
2013	GTCGGATCCCTTAATCTATCCTTAAGCGAAACGC	Δ*upsE* downstr rev *Bam*HI
532	GGGGGGGATCCTCACATGCTAATATTTTCAACCTGATCACC	*upsE* fw *Bam*HI
639	GCGCTGCAGACGAGCTGATAACTGCATAC	*upsE* rev *Pst*I
2264	GATGGTACCTAGCTGGGAAGAGTTTAGTG	Δ*upsE* upstr fw *Kpn*I
2265	GATCCATGGTTAAGCGAAACGCCCATATC	Δ*upsE* downstr rev *Nco*I
2073	CTAGCTTTTTCCAATATTTTTCACC	Δ*upsE* check primer fw
2015	GTAAACTGGAAGCCTATAAGG	Δ*upsE* check primer rev
PRIMERS FOR ECTOPIC INTEGRATION OF GLUCOSE TRANSPORTER SYSTEM		
2256	GATGGTACCTAGCTGGGAAGAGTTTAGTG	*upsE* upstr fw + *Kpn*I site
2257	CTATCAGATATCCTCTGCTAATACCTGATC	*upsE* upstr rev ol to *Saci mbp* promoter
2258	GTATTAGCAGAGGATATCTGATAGTTGGAGAAATG	*Saci mbp* promoter fw ol to *upsE* upstr
986	CTATTACTGCTATTTGGGTATTGGATACTCC	*Saci mbp* promoter + ss rev ol to *Sso glcS*
987	CCAATACCCAAATAGCAGTAATAGTAGCAG	*Sso glcS* fw ol to *Saci mbp* promoter + ss
988	CCCTTCTTCATCAATTACTTCAAGAGATAGTATTTATTG	*Sso glcS* rev ol to *Sso glcT* + *glcV*
989	CTATCTCTTGAAGTAATTGATGAAGAAGGGTACAATAATC	*Sso glcT* fw ol to *Sso glcS* + *glcV*
2259	CTAGAACTGCTCGCTCTTTTTTTAGGGATTC	*Sso glcT* rev ol to *Sso pyrEF* cassette
2260	CTAAAAAAAGAGCGAGCAGTTCTAGTACTTG	*Sso pyrEF* cassette fw ol to *Sso glcV*
2261	GATTCACTGTCTCGCTCAAGCTATGCATGAC	*Sso pyrEF* cassette rev ol to *upsE* downstr
2262	CATAGCTTGAGCGAGACAGTGAATCAATAC	*upsE* downstr fw ol to *Sso pyrEF* cassette
2263	GATGGTACCCATTAAATCTCCGCGATGC	*upsE* downstr rev + *Kpn*I site
PRIMERS FOR IN GENOME TAGGING		
1818	CTCACTGGGCCCTAGAGGCATCGTATGCACCTCAAG	*Saci1274* fw *Apa*I
1819	TTACTTCTCAAATTGTGGATGACTCCAAGCTGTTGAAATTGTAGGTGGAATTATAAC	Saci1274 rev with Strep-tag ol to *Saci1274* downstr
1820	TGGAGTCATCCACAATTTGAGAAGTAACTTTACACCCTTTTTTAATTCGATTCTATATTTTAC	*Saci1274* downstr fw ol to *Saci1274* with Strep-tag
1821	CTCACTGGATCCACAAGTTCCCATCAGACGGAGAAG	*Saci1274* downstr rev *Bam*HI
1583	GGGCCATGGTAGGGCTTACCCTAACTATC	*Saci1210* fw *Nco*I
3106	GATGGTGATGCTTCTCAAATTGTGGATGACTCCACTCCTTTAATATTCGTACTATTGTCTG	*Saci1210* rev with His-tag
3107	CCACAATTTGAGAAGCATCACCATCATCACCATTGAGTTTATGACCATATCAGTTAAAGC	*Saci1210* downstr fw ol to *Saci1210* with His-tag
1578	GGGGGATCCATCTCGCCATACACTCTTAC	*Saci1210* downstr rev *Bam*HI
PRIMERS FOR COPPER PROMOTER		
3428	CACCCGCGGGATCGTTTTAAACTTGACTAGTAC	*copT* rev *Sac*II
3429	CCCCCCATGGGATTAACTCGCTAACAATAAATAAATAG	*copT* fw *copMA* promoter *Nco*I

Underline indicates restriction sites.

Table A2 | Strains and plasmids used in this study.

Strains	Genotype	Source/reference
STRAINS		
DH5α	*Escherichia coli* K-12 cloning strain	Gibco
DSM639	Wild type *Sulfolobus acidocaldarius*	DSMZ
MW001	DSM639 Δ*pyrE* (*Saci*1597; Δ91–412 bp)	This work
MW109	MW001 Δ*upsE* (*Saci*1494; 1–1410 bp)	This work
MW363	MW001 *Saci1210*::$Saci1210_{\text{Strep 6xHis}}$	This work
MW095	MW001 *aglB*:$aglB_{\text{One-Strep}}$	This work
MW009	MW001 Δ*upsE*:Pmbp_{Saci} SSmbp_{Saci} $glcS_{SSO}$ $glcT_{SSO}$ $glcU_{SSO}$ $glcV_{SSO}$	This work
PLASMIDS		
pGEM-T Easy		Promega
pSVA402	In-frame deletion of $pyrE_{Saci}$ cloned into pGEM-T Easy	This work
pSVA406	Gene targeting plasmid, pGEM-T Easy backbone, $pyrEF_{SSO}$ cassette; single crossover method	This work
pSVA407	Gene targeting plasmid, pGEM-T Easy backbone, $pyrEF_{SSO}$ and $lacS_{SSO}$ cassette; single crossover method	This work
pSVA431	Gene targeting plasmid, pGEM-T Easy backbone, $pyrEF_{SSO}$ and $lacS_{SSO}$ cassette; double crossover method	This work
pSVA1804	In-frame deletion of *upsE* cloned into pSVA406 with *Apa*I, *Bam*HI	This work
pSVA447	In-frame deletion of *upsE* cloned into pSVA407 with *Apa*I, *Bam*HI	This work
pSVA449	In-frame deletion of *upsE* cloned into pSVA431 with *Nco*I, *Kpn*I and the gen region with *Bam*HI, *Pst*I	This work
pSVA445	Inserting the glucose transporter system of *S. solfataricus* together with the $pyrEF_{SSO}$ cassette into the $upsE_{Saci}$ site	This work
pSVA1224	Genomically Strep-tag of aglB cloned into pSVA407 with *Apa*I, *Bam*HI	This work
pSVA1252	Insertion of One-Strep linker sequence into pSVA1224 with *Acc*I, *Pst*I	This work
pSVA1097	Genomically Strep-6xHis-tag of *Saci1210* cloned into pSVA406 with *Nco*I, *Bam*HI	This work
pCMalLacS	pRN1 based shuttle vector with $lacS_{SSO}$ reporter gene	Berkner et al. (2010)
pSVA1673	*copA* promoter replacing *mbp* promoter, cloned into pCMalLacS with *Sac*II, *Nco*I	This work

5´
GTATACATTATGGTCTATATATATCAATTCGTTATGCCAAACGTTATAATTCCACCTACAATTTCAA
CAGCTGGTGGAGGTGGAGGTTCATGGAGTCATCCTCAATTCGAGAAAGGAGGTGGATCAGGTG
GAGGTTCAGGTGGAGGTTCTTGGAGTCATCCACAATTTGAGAAGTAACTTTACACCCTTTTTTAAT
TCGATTCTATATTTTACATATTTATCGACGGGAGAAAAACGAGGAGGTATTGGTATGAATGTCTTT
TCACCACAAGTAGGACATCTTTCGCTCATCGTATAAGTATAGTCTTTATTACATCTTCTTATTTTGG
ATTTCACTTCTTAACCACCGAAAACGTTACGTTTTCATGCTGGCTTATCTCCTTAATCAAGTTGATT
AGTTCCTGTAAAGCATTTGATACCAACTTCTGATCTGTACCTATCACATCAATCCTATACCTGGGA
GCACCTATTGTATATATTTTAACTTCTACATCAAGGTCATCTATTTTTTCAACTGCAG 3'

FIGURE A1 | Codon optimized One-STrEP-tag sequence.

frontiers in
MICROBIOLOGY

REVIEW ARTICLE
published: 02 July 2012
doi: 10.3389/fmicb.2012.00207

Diversity and subcellular distribution of archaeal secreted proteins

Zalan Szabo*[1] *and Mechthild Pohlschroder*[2]

[1] *MicroDish BV, Utrecht, Netherlands*
[2] *Department of Biology, University of Pennsylvania, Philadelphia, PA, USA*

Edited by:
Frank T. Robb, University of California, USA

Reviewed by:
Thijs Ettema, Uppsala University, Sweden
Jerry Eichler, Ben Gurion University of the Negev, Israel

****Correspondence:***
Mechthild Pohlschroder, Department of Biology, University of Pennsylvania, 201 Leidy Laboratories, Philadelphia, PA 19104, USA.
e-mail: pohlschr@sas.upenn.edu

Secreted proteins make up a significant percentage of a prokaryotic proteome and play critical roles in important cellular processes such as polymer degradation, nutrient uptake, signal transduction, cell wall biosynthesis, and motility. The majority of archaeal proteins are believed to be secreted either in an unfolded conformation via the universally conserved Sec pathway or in a folded conformation via the Twin arginine transport (Tat) pathway. Extensive *in vivo* and *in silico* analyses of N-terminal signal peptides that target proteins to these pathways have led to the development of computational tools that not only predict Sec and Tat substrates with high accuracy but also provide information about signal peptide processing and targeting. Predictions therefore include indications as to whether a substrate is a soluble secreted protein, a membrane or cell wall anchored protein, or a surface structure subunit, and whether it is targeted for post-translational modification such as glycosylation or the addition of a lipid. The use of these *in silico* tools, in combination with biochemical and genetic analyses of transport pathways and their substrates, has resulted in improved predictions of the subcellular localization of archaeal secreted proteins, allowing for a more accurate annotation of archaeal proteomes, and has led to the identification of potential adaptations to extreme environments, as well as phyla-specific pathways among the archaea. A more comprehensive understanding of the transport pathways used and post-translational modifications of secreted archaeal proteins will also facilitate the identification and heterologous expression of commercially valuable archaeal enzymes.

Keywords: Tat transport, Sec transport, protein transport, archaea, lipoprotein, pili, cell surface structures, archaeosortase

INTRODUCTION

During protein translation at the ribosome, the presence of an N-terminal signal sequence or a transmembrane helix can target the corresponding protein for translocation into or across the cytoplasmic membrane (Driessen and Nouwen, 2008; Natale et al., 2008; Ellen et al., 2010b; Yuan et al., 2010; Calo and Eichler, 2011). Archaeal proteins targeted for secretion, including polymer degrading enzymes, antimicrobial porteins, or proteins involved in intercellular signaling may be released into the extracellular environment. However, most proteins remain associated with the cell, requiring either a direct interaction with the membrane, an indirect association with membrane-anchored proteins, or, as recently predicted, attachment to the cell wall (**Figure 1**; Gimenez et al., 2007; Albers and Pohlschroder, 2009; Storf et al., 2010; Haft et al., 2012). Proteins that remain associated with the cell surface are involved in a wide variety of cellular processes, including nutrient uptake, motility, and surface adhesion (Ellen et al., 2010b; Albers and Meyer, 2011).

In bacteria and archaea, the universally conserved Sec pathway is commonly used to transport secreted proteins to the cytoplasmic membrane (Nather and Rachel, 2004; Pohlschroder et al., 2005b; Driessen and Nouwen, 2008). Since these proteins are secreted in an unfolded conformation through a narrow proteinaceous pore, protein folding, post-translational modifications, such as glycosylation, and associations with cofactors, must occur in an extracytoplasmic environment devoid of ATP, and hence ATP-dependent chaperones. These complications are circumvented by substrates of the twin arginine transport (Tat) pathway, at least in part, because protein folding, cofactor incorporation, and possibly protein modifications, are accomplished in the cytoplasm (Pohlschroder et al., 2005a; Natale et al., 2008). Regardless of the transport pathway used, archaeal secreted proteins that are not released into the extracellular environment must be anchored to the cell, the only exception known to date being proteins of the archaeal genus *Ignicoccus*, which has an outer membrane that provides a periplasmic space that can retain secreted proteins (Huber et al., 2002; Nather and Rachel, 2004).

The results generated by several *in vivo* studies of archaeal secreted proteins and their transport pathways has allowed for the identification of the signal peptides required to target secreted proteins to a specific transport pathway as well as amino- or carboxy-terminal motifs within the protein that mediate substrate interactions with extracytoplasmic structures (Kobayashi et al., 1994; O'Connor and Shand, 2002; Rose et al., 2002; Dilks et al., 2005; Gimenez et al., 2007; De Castro et al., 2008; Albers and Pohlschroder, 2009; Ng et al., 2009; Kwan and Bolhuis, 2010; Storf et al., 2010; Calo and Eichler, 2011). In turn, the identification of

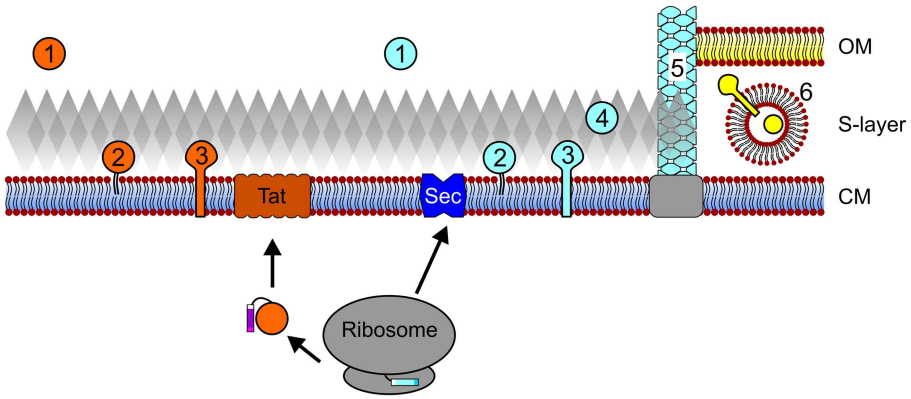


FIGURE 1 | Archaeal protein secretion and subcellular localization. Proteins that contain Tat signal peptides pass the cytoplasmic membrane (CM) through the Tat translocon after translation and folding. Conversely, protein translocation through the Sec pore can occur co- and possibly post-translationally. Upon secretion and signal peptide processing, Sec and Tat substrates can be released into the extracellular milieu (1), be embedded into the membrane via a lipid anchor (2) or a C-terminal transmembrane segment (3). *In silico* data also suggest that Sec substrates may be anchored to the cell wall in an archaeosortase-dependent manner (4) and a number of type IV pilin-like proteins have been shown to assemble into cell appendages (5). In *I. hospitalis*, an additional outer membrane (OM) is present and periplasmic vesicles (6) are thought to play a role in the trafficking of secreted and outer membrane proteins across the periplasmic space. See text for details. Cell components are not drawn to scale.

these motifs in a large number of substrates has allowed for the development of software programs that facilitate *in silico* analyses of secreted protein sequences encoded by a large number and variety of archaeal genomes (Rose et al., 2002; Szabo et al., 2007a; Bagos et al., 2009; Storf et al., 2010). These *in silico* analyses have already lead to important insights into the strategies used to secrete and anchor proteins to the cell surface, which vary in part depending on the function of the secreted protein. Ultimately, the trends revealed by *in silico* analyses may clarify how an organism adapts to the selective pressures imposed on it, and may make clear which aspects of the environment had the greatest impact on the evolution of the organism.

Besides the well-studied protein transport pathways, archaea may use additional, poorly understood or currently unknown, means by which to facilitate the egress of proteins across the membrane. For example, some archaeal species produce vesicles that are released into the extracellular environment (Soler et al., 2008; Ellen et al., 2010b), or in the case of *Ignicoccus*, are present in the periplasmic space (Nather and Rachel, 2004). Future studies may clarify the role of these mechanisms in protein transport.

In this review, we describe the highly diverse array of archaeal secreted proteins. We consider the pathways used to transport these proteins to the cytoplasmic membrane, the ultimate destinations of substrates, and whether these proteins are secreted into the extracellular environment or are anchored to the cytoplasmic membrane or cell wall. We also explore the mechanisms that facilitate the tethering of secreted substrates to the cell surface, and discuss the types of functions displayed by cell-associated and secreted proteins. Finally, we describe the *in silico* approaches used to predict the subcellular localization of substrates and the post-translational modifications that these substrates undergo and identify potential trends in the use of these pathways and modifications in various microorganisms.

Sec AND Tat SUBSTRATE TARGETING, SECRETION, AND POST-TRANSLATIONAL MODIFICATION

In the following section we briefly describe the two main routes for archaeal protein transport across the cytoplasmic membrane, the Sec and the Tat pathways, while focusing on the processing and modification of the substrates transported by these pathways. These systems have recently been more extensively reviewed elsewhere (Pohlschroder et al., 2005b; Ellen et al., 2010b; Yuan et al., 2010; Calo and Eichler, 2011).

Sec SUBSTRATE RECOGNITION AND TRANSPORT

In silico and *in vivo* analyses indicate that all species, whether eukaryotic or prokaryotic, transport proteins by way of the universally conserved Sec pathway, which acts as a conduit for inserting proteins into the cytoplasmic membrane or secreting them into the extracytoplasmic environment (Yuan et al., 2010; Calo and Eichler, 2011). Proteins are targeted to the Sec pathway by conserved amino-terminal signal peptides that have a tripartite structure consisting of a charged amino-terminus, a hydrophobic stretch and a signal peptidase recognition motif (Bardy et al., 2003; Ng et al., 2007; Zimmermann et al., 2011; **Table 1**). The Sec pathway consists of several components, including the signal recognition particle (SRP). The SRP recognizes either the signal peptide or transmembrane segments in the substrate as the nascent peptide chain emerges from the ribosome, which results in a translational arrest. Subsequently, the SRP–ribosome nascent chain complex is targeted to the proteinaceous Sec pore where the substrate is translocated across the membrane co-translationally (Grudnik et al., 2009). Conversely, SRP-independent post-translational Sec transport requires chaperones to maintain the precursor in an unfolded conformation to allow for the secretion through the approximately 20 Å Sec pore – a pore just large enough to allow for transport of an unfolded polypeptide (Van den Berg et al., 2004; Mori et al., 2010; Yuan et al., 2010). While the key components required for co-translational transport are universally

Table 1 | Frequencies of secretion and cell surface-anchoring motifs.

Kingdom	Euryarchaea			Crenarchaea			Other		
Signal Peptide*	*H. volcanii*	*M. kandleri*	*A. fulgidus*	*S. solfataricus*	*I. hospitalis*	*A. pernix*	*N. equitans*	*K. cryptofilum*	*N. maritimus*
Sec/Spase I N ++ AXA	++	++	++	++	++	++	++	++	++
Sec/Spase II N ++ LAGC	+	-	+	-	-	-	-	-	-
Sec/Spase III N +G/A	+	+	+	+	+	+	+	+	+
Tat/Spase I N RR AXA	+	-	+	+	+	+	-	+	-
Tat/Spase II N RR LAGC	++	-	+	-	-	-	-	-	-
PGF C-term PGF ++ C	+	-	+	-	-	-	-	-	-

Species abbreviations: *Haloferax volcanii*, *Methanopyrus kandleri*, *Archaeoglobus fulgidus*, *Sulfolobus solfataricus*, *Igniococcus hospitalis*, *Aeropyrum pernix*, *Nanoarchaeum equitans*, *Korarchaeum cryptofilum*, *Nitrosopumilus maritimus*

*Spaces between boxes represent SPase processing sites, except for PGF C-term, where it represents predicted archaeosortase processing at PGF-motif (exact processing site has not yet been determined)

conserved, post-translational transport across the ER membrane depends on a luminal ATPase, Bip/Kar2, whereas bacterial post-translational transport requires the cytoplasmic ATPase, SecA. There is evidence for archaeal post-translational translocation of Sec substrates (Dale et al., 2000; Ortenberg and Mevarech, 2000; Irihimovitch and Eichler, 2003). However, as archaea lack homologs of either of the two aforementioned ATPases, the energetics of archaeal post-translational transport is not understood to date.

Tat SUBSTRATE RECOGNITION AND TRANSPORT

The Tat pathway, which unlike the Sec pathway, is not universally conserved, being limited to prokaryotes, chloroplasts, and a few protists, transports proteins in a folded conformation (Pohlschroder et al., 2005a; Natale et al., 2008; Yuan et al., 2010; Robinson et al., 2011). Because they are folded prior to transport, Tat substrates are too large for the Sec pore to accommodate. Therefore, it is not surprising that the Tat pathway consists of components and mechanisms that are fundamentally different from those employed by the Sec pathway. However, just as Sec substrates are targeted to the Sec pathway by a signal peptide, a Tat signal peptide targets proteins to the Tat pathway (Berks et al., 2005; Robinson et al., 2011). Organisms containing a functional Tat pathway require at least TatA and TatC, which play particularly pivotal roles, with TatC (and TatB in many organisms) apparently being involved in substrate targeting, while multimers of TatA, a single transmembrane spanning protein, allow for the formation of pores of varying sizes to accommodate the diverse sizes of secreted substrates (Dilks et al., 2005; Gohlke et al., 2005; Leake et al., 2008; Robinson et al., 2011).

While the Sec and Tat pathways are mechanistically very different, the tripartite structures of the signal sequences that target substrates to these systems are surprisingly similar (**Table 1**). However, there are some subtle differences between Sec and Tat signal peptides. Among the most significant of these differences is the highly conserved twin arginine motif that lies within the charged region of the Tat signal peptide. The Tat signal peptide also has a hydrophobic stretch that is generally less hydrophobic than that of the Sec signal sequence (Rose et al., 2002; Bendtsen et al., 2005).

Sec AND Tat SIGNAL PEPTIDE PROCESSING

In addition to the regions of signal peptides that are critical to targeting secreted proteins to a protein transport pathway, Sec and Tat signal peptides also usually contain a recognition sequence that targets it for cleavage by a signal peptidase, although unprocessed Sec substrate signal peptides occasionally serve as amino-terminal membrane anchors (Eichler, 2002; Paetzel et al., 2002; Tuteja, 2005; Ng et al., 2007). Substrates having signal peptides that are processed by the universally conserved signal peptidase I (SPase I) are either released from the membrane or anchored to the cell by way of a carboxy-terminal membrane anchor or protein–protein interaction (Rose et al., 2002; Dilks et al., 2005; Tuteja, 2005; Gimenez et al., 2007; De Castro et al., 2008; Uthandi et al., 2010).

Recent *in silico* data suggest that several euryarchaeal SPase I processed Sec substrates have an additional carboxy-terminal sorting signal that targets these proteins to the cell wall or to another cell surface structure (Haft et al., 2012). Originally identified in Gram-positive bacteria, an enzyme known as a sortase processes these sorting signals, which are characterized by a signature motif containing a hydrophobic transmembrane α-helix and a cluster of basic amino acids (Spirig et al., 2011; **Table 1**). Upon cleavage, substrate proteins are typically covalently linked to the peptidoglycan cell wall by the sortase. Several types of sortases, each processing distinct motifs within the conserved tripartite structure have now been identified in bacteria, and similar C-terminal tripartite sorting signals have also been found in archaea. Furthermore, using comparative genomics, potential archaeosortases have been identified (Haft et al., 2012). However, sorting signal processing and modification in archaea awaits *in vitro* and *in vivo* confirmation.

Bacteria and archaea, but to the best of our knowledge not eukaryotes, express proteins that have signal peptides processed by signal peptidases distinct from the universally conserved SPase I. Bacterial signal peptides recognized by SPase II are largely similar to those processed by SPase I, containing a charged amino-terminus and a hydrophobic stretch; however, they also contain a short motif known as a lipobox that lies subsequent to the hydrophobic stretch (**Table 1**). The lipobox contains an invariant cysteine residue that is acylated by a prolipoprotein diacylglyceryl transferase prior to signal peptidase processing by SPase II, which then cleaves the precursor immediately upstream of the lipid-modified cysteine. The acylated cysteine residue anchors the processed substrate to the cytoplasmic membrane (Sankaran and Wu, 1994; Hutchings et al., 2009; Thompson et al., 2010; Okuda and Tokuda, 2011). Although archaea, particularly euryarchaea, express Sec and Tat substrates containing lipobox motifs, neither an archaeal SPase II homolog nor a prolipoprotein diacylglyceryl transferase homolog has yet been identified (Mattar et al., 1994; Dilks et al., 2005; Falb et al., 2005; Gimenez et al., 2007; Storf et al., 2010).

The highly conserved nature of SPase II processing site sequences across prokaryotic domains makes the apparent absence of an archaeal SPase II homolog even more intriguing (Storf et al., 2010). Perhaps the mechanisms involved in archaeal lipoprotein biosynthesis are unique to archaea, possibly owing to the distinct membrane lipid composition in the two prokaryotic domains. Studies performed on bacterial and archaeal lipoprotein mutants in which the conserved lipobox cysteine was replaced with a serine lend credence to the hypothesis that this cysteine is critical in both domains. Mutant lipoproteins are not processed in either archaea or bacteria, and like in bacteria, unprocessed archaeal substrates often remain cell-associated while a subset of similar mutant lipoprotein precursors are secreted into the supernatant (Hutchings et al., 2009; Storf et al., 2010).

SPase III specifically processes the subunits of a wide array of type IV pilus-like surface structures; included among these are the flagella of archaea (see below; Pohlschroder et al., 2011). Similar to SPase I and SPase II, SPase III recognizes a processing site located adjacent to the hydrophobic stretch of the N-terminal signal peptide that targets the substrate to the transport pathway (Arts et al., 2007; Francetic et al., 2007). However, unlike signal peptides processed by SPase I or SPase II, the processing site is N-terminal to the hydrophobic stretch, which hence remains part of the mature protein where it is essential for subunit–subunit interactions that allow the formation of an α-helical scaffold facilitating the assembly of type IV pili and pilus-like structures (Bardy and Jarrell, 2002; Albers et al., 2003; **Table 1**). Although the translocation of archaeal type IV pilin-like proteins into the cytoplasmic membrane has not been studied, in Gram-negative bacteria, they were shown to be secreted by the Sec pathway in a signal particle-dependent fashion (Arts et al., 2007; Francetic et al., 2007).

N- AND O-GLYCOSYLATION

Additional post-translational modifications of archaeal secreted proteins, other than signal peptide processing, include both O- and N-glycosylation (Eichler and Adams, 2005; Calo et al., 2010; Jarrell et al., 2010). The first non-eukaryotic N-glycosylated protein reported was the surface S-layer glycoprotein of the haloarchaeon *Halobacterium salinarum*, a protein that was subsequently also shown to be O-glycosylated (Mescher et al., 1974; Sumper et al., 1990). While little is known about the O-linked glycosylation process, genetic and biochemical studies of the S-layer, as well as of the flagella subunits in two methanogenic archaea and a haloarchaeon, revealed a set of archaeal glycosylation (Agl) enzymes that catalyze the assembly and attachment of N-linked glycans to target proteins (Calo et al., 2010; Jarrell et al., 2010). As in eukaryotes, in the final step of glycosylation, an oligosaccharyltransferase homolog covalently links an assembled oligosaccharide onto the asparagine residue of a target motif (N-X-T/S) in the protein. Interestingly, aside from the oligosaccharyltransferase, the Agl components of haloarchaea and methanogens, or even closely related methanogens, show no homology to each other, indicating a wide diversity in the specifics of post-translational modification by N-glycosylation across the archaea (Calo et al., 2010; Jarrell et al., 2010). Initial studies on N-glycosylated proteins in crenarchaea support this hypothesis; however, much less is known about the components required for N-glycosylation in these organisms (Peyfoon et al., 2010; Meyer et al., 2011).

DIVERSITY OF SECRETED PROTEINS

Secreted proteins play a crucial role in the interaction of cells with their environment. They fulfill a variety of functions such as degradation of complex polymeric substances (e.g., carbohydrates, proteins), nutrient uptake, signal transduction, or formation of surface complexes such as the cell wall or pili (for examples see **Table 2**). While for the majority of secreted proteins no function can be predicted, many secreted proteins encoded in model archaea or heterologously expressed in a model archaeon have been characterized, leading to a better understanding of the physiology of the organisms, as well as the transport pathways and post-translational modification (i.e., some have been used as reporters). In the following section, several examples are given for the variety of tasks that secreted archaeal proteins fulfill and what is known about their secretion and localization.

DEFENSE

Several archaea produce peptides or proteins with antimicrobial activity that are collectively termed archaeocins. Thus far they have

Table 2 | Archaeal secreted proteins.

Secreted protein	Secretion pathway	Signal peptide processing	Localization				
			Extracellular	CW-anchor	TM-anchor	Lipid anchor	Peripheral[#]
ARCHAEOCINS							
HalH4, *Hfx. mediterranei* R4	Tat	SPase I	+				
HalC8, haloarchaeal stain AS7092	Tat	SPase I	+				
SulAB, *S. acidocaldarius*	Sec	SPase I					+
ENZYMES							
Pyrolysin, *P. furiosus*	Sec	SPase I	+				
α-Amylase, *S. solfataricus*	Sec	SPase I	+		+		
Protease, *S. marinus*	Sec	SPase I		+			
Protease (Nep), *N. magadii*	Tat	SPase I	+				
Halocyanin I, *Hfx. volcanii*	Tat	SPase I			+		
Halocyanin, *Hbt. salinarum*	Tat	SPase II				+	
Laccase, *Hfx. volcanii*	Tat	SPase I	+				
Alkaline phosphatase, *Hbt. salinarum*	Tat	SPase I	+				
SUBSTRATE-BINDING PROTEINS							
Glucose-binding protein, *S. solfataricus*	Sec	SPase III					+
Iron-binding protein, *Hfx. volcanii*	Tat	SPase II				+	
Trehalose-binding protein, *T. litoralis*	Sec	SPase II				+	
Oligopeptide-binding protein, *A. pernix*	Sec	SPase I			+		
SURFACE STRUCTURES							
S-layer glycoprotein *Hfx. volcanii*	Sec	SPase I		+			
S-layer glycoprotein, *S. solfataricus*	Sec	SPase I			+		
S-layer glycoprotein, *M. fervidus*	Sec	SPase I					+
Flagellin, *M. maripaludis*	Sec	SPase III					+
Type IV pilins, *S. acidocaldarius*	Sec	SPase III					+
Mth60 pilin, *M. thermoautotrophicus*	Sec	SPase I					+

Information based on in vivo and/or in silico studies (see text for details and references); [#]mode of surface-anchoring not known.

been identified in halophiles (halocins) and crenarchaea from the genus *Sulfolobus* (sulfolobicins; Ellen et al., 2011). Halocins are plasmid-encoded, secreted antimicrobials that can be subdivided into protein or peptide (micro) halocins (O'Connor and Shand, 2002).

HalH4 from *Haloferax mediterranei* is the best-characterized protein halocin. It is a 35 kDa protein predicted to be secreted by the Tat pathway that is active against other haloarchaea (Cheung et al., 1997). Microhalocins are peptides ranging from about 3 to 8 kDa in mass that are released from a larger prepro-protein that also contains a predicted Tat signal peptide (Price and Shand, 2000; O'Connor and Shand, 2002). They are effective against other haloarchaea, and in some cases, against thermophilic crenarchaea (Haseltine et al., 2001). In the case of halocin 8, the C-terminal 76 amino acid peptide carries antimicrobial activity while the N-terminal protein portion termed HalI, which remains membrane-associated, confers immunity against the halocin, probably by sequestration (Sun et al., 2005).

Sulfolobicins, proteins (about 20 kDa) associated with the membrane and extracellular vesicles, were first identified in *Sulfolobus islandicus* (Prangishvili et al., 2000). The genes encoding sulfolobicins were recently identified in *S. acidocladarius* and *S. tokodaii*, *sulA* and *sulB*, both of which are necessary for antimicrobial activity (**Figure 2A**). The gene products carry putative SPase I processed Sec signal peptides and were suggested to be peripheral membrane proteins (Ellen et al., 2011).

SECRETED ENZYMES

Nutrient sources in the environment are often in a form that cannot be directly taken up by cells. Polymeric molecules such as carbohydrates must be cleaved into smaller subunits, proteins degraded into peptides, and inorganic phosphate liberated from organic compounds (Worthington et al., 2003; Wende et al., 2010; Xu et al., 2011). Many archaeal enzymes that perform these tasks have been identified and characterized, either by sequence homology search and heterologous expression, or by purification of the native enzymes from cell culture supernatants of the host (Moracci et al., 2007). In the case of native enzymes, protein identification by N-terminal sequencing provided valuable information regarding the signal peptidase cleavage site and in some cases post-translational modifications by glycosylation or lipid modification. In addition, functional assays can also be used as reporters for *in vivo* studies on secretion, using either plate assays or measuring enzyme activity in cell culture supernatant (**Figure 2B**; Rose et al.,

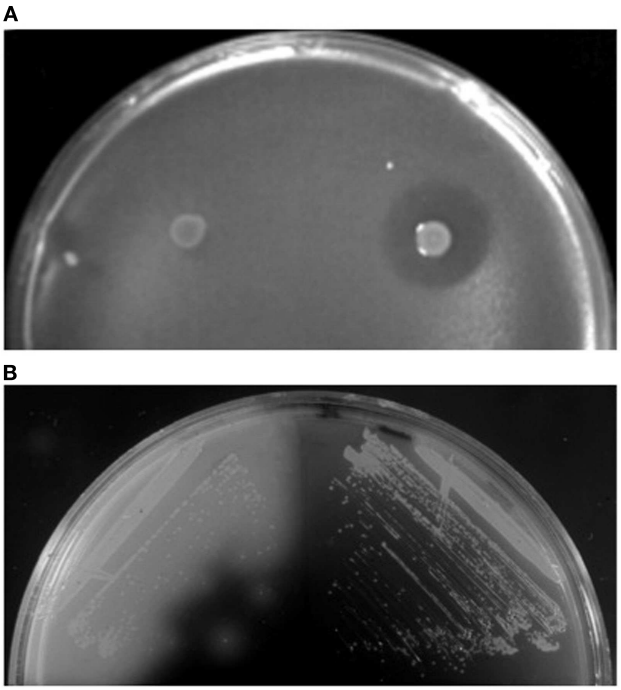


FIGURE 2 | Detection of extracellular protein activities. (A) Overlay assay with colonies of *S. acidocaldarius* wt (right) and Δ*sulAB* mutant (left) grown on a lawn of *S. solfataricus* strain P2. Clear halos surrounding the colonies indicate killing of archaea by the sulfolobicins. Image reproduced with permission from Ellen et al. (2011). **(B)** Iodine vapor staining of *H. volcanii* expressing wild-type (left) or signal sequence mutated (right) α-amylase grown on rich medium supplemented with 0.2% soluble starch. Clear halos surrounding colonies indicate starch hydrolysis by extracytoplasmic α-amylase (Rose et al., 2002).

2002; Worthington et al., 2003; Hutcheon et al., 2005; Wende et al., 2010; Xu et al., 2011).

Examples of secreted enzymes include proteases such as halolysins and pyrolysins (Kamekura et al., 1992; de Vos et al., 2001; Shi et al., 2006; De Castro et al., 2008), proteins involved in starch degradation such as α-amylase and cyclodextrin glycosyltransferase (Kobayashi et al., 1994; Hutcheon et al., 2005; Bautista et al., 2012), and a copper-containing oxidase (laccase; Uthandi et al., 2010). In some cases, secretion and localization have been studied in more detail. These studies showed that not all secreted enzymes are soluble extracellular proteins, with some being anchored to the membrane via a lipid anchor (Mattar et al., 1994; Gimenez et al., 2007; Storf et al., 2010) or a C-terminal transmembrane segment (Gimenez et al., 2007). Interestingly, an α-amylase from *S. solfataricus* that contains a C-terminal membrane anchor in addition to an SPase I processed signal sequence was detected in both membrane and extracellular fractions (Worthington et al., 2003; Ellen et al., 2010a). Apparently, part of the protein pool is released by an unknown mechanism, which may contribute to more efficient substrate degradation. Direct attachment to the S-layer was demonstrated for a protease from *Staphylothermus marinus* (Mayr et al., 1996), and tetrathionate hydrolase from *Acidianus ambivalens* was suggested to be localized to the pseudo-periplasmic space, i.e., between the cytoplasmic membrane and the S-layer (Protze et al., 2011). More recently, another tetrathionate hydrolase from *A. hospitalis* was found to be assembled into "zipper-like particles" on the cell surface and extracellular space (Krupovic et al., 2012) (**Figure 3A**). The secretion mechanism of enzymes has only been addressed in haloarchaea. Secretion by the Tat pathway was demonstrated for the protease SptA from *Natrinema* sp. J7 (Shi et al., 2006), arabinanase, halocyanin 2 and 3, a DsbA-like protein, and two hypothetical proteins from *Haloferax volcanii* (Gimenez et al., 2007; Storf et al., 2010) as well as α-amylases from *Natronococcus* sp. strain Ah36 and *Haloarcula hispanica* (Rose et al., 2002; Hutcheon et al., 2005).

Enzymes, especially those of archaeal hyperthermophiles, have potential industrial applications, in particular hydrolytic enzymes that degrade complex polymers. At high temperatures, starch and cellulose are more accessible and the risk of microbial contamination is low. Most known hyperthermophiles (cells that grow at 80°C and above) are archaea and their enzymes can withstand the harsh conditions employed in industrial processes (Egorova and Antranikian, 2005). For example, an α-amylase with high homology to a protein from *Thermococcus kodakarensis*, which was identified in an environmental DNA library, is stable and highly active under the conditions applied for starch liquefaction (Richardson et al., 2002).

SUBSTRATE BINDING PROTEINS

Nutrients such as peptides and oligosaccharides produced by extracellular enzymes are taken up by dedicated transporters. Archaea make extensive use of binding protein-dependent ATP binding cassette (ABC) transporters. Binding proteins have high affinity for their substrates. Although substrate binding proteins of Gram-negative bacteria reside in the periplasmic space, they are usually associated with the cytoplasmic membrane in archaea. Different mechanisms are employed for anchoring.

The class of oligopeptide-binding proteins characterized in the crenarchaeal species *S. solfataricus* and *Aeropyrum pernix* contains an N-terminal secretory signal sequence with an SPase I cleavage site and an additional C-terminal transmembrane segment (Elferink et al., 2001; Goglietto et al., 2010; Balestrieri et al., 2011). The membrane anchor is preceded by a stretch of glycosylated serine and threonine residues. Several euryarchaeal binding proteins were shown to be lipid anchored after processing by the yet to be identified SPase II. These lipobox-containing proteins are associated with transport of maltose in *Thermococcus litoralis* (Horlacher et al., 1998), or predicted to bind iron and maltose, respectively, in *H. volcanii* (Gimenez et al., 2007). In addition, a putative phosphate binding protein from *Hfx. volcanii* was predicted to be processed by an archaeosortase and may therefore have a C-terminal lipid anchor (Haft et al., 2012). Interestingly, BasA and CosA from *Hbt. salinarum* are two lipid anchored binding proteins shown to be involved in chemotaxis rather than substrate uptake despite their homology to ABC transporter solute binding proteins (Kokoeva et al., 2002).

A third class of binding proteins contains a prepilin-like signal sequence that is cleaved by SPase III. Removal of the short positively charged leader peptide has been shown for *S. solfataricus* glucose and arabinose binding proteins (Albers et al., 1999, 2003; Elferink et al., 2001). *In silico* analyses predicted substrate binding proteins with SPase III processing sites in many Eury-

and Crenarchaea, however these predictions await experimental confirmation (Szabo et al., 2007b). Similar to pilins and archaeal flagellins that are assembled into cell appendages, sugar binding proteins are thought to be incorporated into a macromolecular structure called the bindosome (see below).

CELL SURFACE STRUCTURES

Most archaeal cells are surrounded by a membrane-anchored, crystalline S-layer that provides stability and osmoprotection to the cells (Engelhardt, 2007a,b; Albers and Meyer, 2011). Known S-layers are composed of a single protein or two subunits that are secreted by the Sec pathway and whose signal peptides are processed by SPase I. In Sulfolobales, anchoring to the cell membrane occurs via a C-terminal transmembrane segment, which is preceded by a stalk-like structure creates a pseudo-periplasmic space. While haloarchaeal S-layer proteins also contain a C-terminal hydrophobic stretch, it has recently been proposed that it is part of a C-terminal conserved tripartite sorting signal that is cleaved and lipid-modified by a yet uncharacterized system possibly involving the predicted archaeosortase enzyme (Haft et al., 2012). Consistent with this observation, *Hfx. volcanii* and *Hbt. salinarum* S-layer glycoproteins contain lipid modifications (Kikuchi et al., 1999; Konrad and Eichler, 2002). Conversely, the S-layer glycoprotein of *Methanothermus fervidus* does not contain a C-terminal transmembrane segment and the mechanism of cell surface attachment is elusive (Brockl et al., 1991).

Archaea also produce a diverse array of cell appendages that provide functions such as motility and surface adhesion. Most characterized cell appendage subunits have an N-terminal type IV pilin-like signal peptide (see above; Albers et al., 2003; Bardy et al., 2003; Szabo et al., 2007b; Tripepi et al., 2010).

The best-characterized archaeal cell surface structure is the flagellum, which is required for swimming motility. In contrast to bacterial flagellar biogenesis, which uses a type III secretion-like mechanism, the archaeal flagellum biosynthesis machinery includes homologs of the bacterial type IV pilus biogenesis components, including the archaeal prepilin-peptidase homolog PibD/FlaK (Journet et al., 2005; Ghosh and Albers, 2011). In addition to conferring swimming motility, flagella also play a role in surface adhesion in *S. solfataricus*, *M. maripaludis*, and *Pyrococcus furiosus* (Nather et al., 2006; Zolghadr et al., 2010; Jarrell et al., 2011). In fact, the thermal stability and adhesive properties led to the suggestion that *Pyrococcus* flagella be used as "molecular glue" (US Patent US 2008/0305524). Finally, motility assays developed for several flagellated archaea have been used to monitor successful secretion and post-translational modification of flagellins (Chaban et al., 2007; Tripepi et al., 2010; Lassak et al., 2012).

Additional genes encoding archaeal type IV pilin-like proteins were identified, using a program trained on flagellin subunits (see below), ultimately leading to the identification and characterization of non-flagellar pilus-like structures. For example, *S. acidocaldarius*, use Aap and Ups-pili, along with flagella, for surface adhesion and biofilm formation (Henche et al., 2012). DNA damage such as that caused by UV light irradiation strongly induces expression of the subunits of the Ups-pili, cell surface structures that subsequently facilitate cell aggregation. Once aggregated, these cells are thought to exchange chromosomal DNA, allowing efficient repair of the damaged DNA (Frols et al., 2008; Ajon et al., 2011). Interestingly, as noted above, *S. solfataricus* also contains sugar binding proteins that are processed by the same SPase III as the flagellin/pilin subunits. It is believed that these proteins are also assembled into a large pilus-like structure, the bindosome, since homologs of pilus-biosynthesis genes are also required for bindosome function, i.e., the growth of cells on certain sugars (Zolghadr et al., 2007).

Most archaea appear to only have one SPase III homolog; some, however, contain additional paralogs with specialized functions. *M. maripaludis* expresses a prepilin-peptidase, EppA, that recognizes type IV pilins with a specific SPase III processing site (previously designated as domain of unknown function 361), along with FlaK/PibD (Szabo et al., 2007b; Wang et al., 2008; Ng et al., 2011) (**Figure 3B**).

The only known archaeal pili that do not contain subunits with a type IV pilin signal peptide are the Mth60 fimbriae produced by *Methanothermobacter thermoautotrophicus*, which are involved in adhesion to solid surfaces (Thoma et al., 2008; **Figure 3C**). Mth60 is a small (16 kDa) glycoprotein with no known homologs in any other species. It contains a predicted SPase I processed Sec signal peptide and while it is not known how Mth60 fimbriae are assembled, the recombinant protein was shown to polymerize into fibers at the optimal growth temperature of *M. thermoautotrophicus*.

Halomucin is the largest known archaeal protein, produced by the extremely halophilic archaeon *Haloquadratum walsbyi* (**Figure 3D**). The 9195 amino acid protein with homology to eukaryotic mucin, is predicted to be secreted via the Sec pathway and its likely function is to provide a hydration shell in their low water activity environment (Bolhuis et al., 2006; Sublimi Saponetti et al., 2011).

Other unusual cell surface structures have been described, but are not yet characterized in molecular detail. One fascinating example is the hamus produced by the euryarchaeon SM1, which grows in cold sulfidic springs (Rudolph et al., 2001; Moissl et al., 2005). Each cell of strain SM1 produces about 100 hami. The hami are highly complex structures having a helical basic structure from which prickles emanate at regular distances, forming a barbwire-like appearance. These structures consist of large subunits of about 120 kDa. The N-terminal sequence of this protein is not known and therefore the mode of secretion cannot be assessed (Moissl et al., 2005). Finally, cannulae from the crenarchaeaon *Pyrodictium abyssi* are tube-like extracellular structures that cross the S-layer and the pseudo-periplasmic space, but do not protrude into the cytoplasm. Cannulae are composed of at least three related glycoproteins, but their sequences have not been published (Mai et al., 1998). While most currently characterized archaeal surface appendages appear to follow the type IV pilin assembly route, a more diverse variety of specialized secretion systems can be expected from the molecular analysis of structures that have so far only been studied on a morphological level.

SUBCELLULAR LOCALIZATION PREDICTION

As described above, the fate of a secretory protein upon translocation is determined by specific sequence motifs within the signal peptide, such as those for signal peptidase processing and lipid modification (Bardy et al., 2003; Ng et al., 2007). Additionally, the

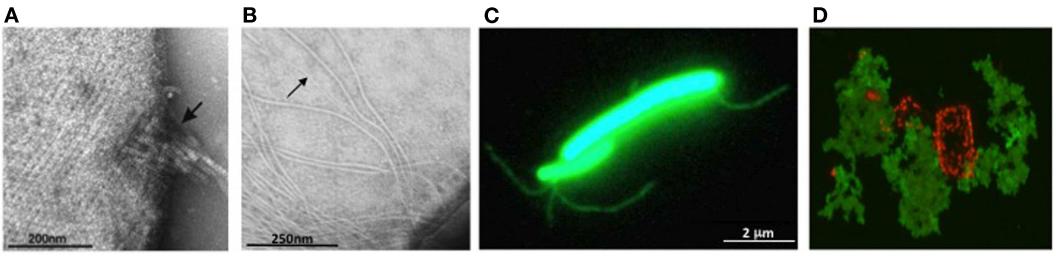


FIGURE 3 | Archaeal surface structures. (A) Tetrathionate hydrolase from *A. hospitalis* YS8 assembles into zipper-like particles on the cell surface. Image reproduced with permission from Krupovic et al. (2012). **(B)** Electron micrographs of *M. maripaludis* cells expressing type IV pilus-like structures. Arrows indicate EppA-processed pili, while additional, thicker structures, are PibD-processed flagella. Samples were negatively stained with 2% phosphotungstic acid. Image reproduced with permission from (VanDyke et al., 2008). **(C)** Non-type IV Mth60 fimbriae of planktonic *M. thermoautotrophicus* cells by staining with AlexaFluor®-488 (Thoma et al., 2008). Image courtesy of R. Wirth, University of Regensburg, Germany. **(D)** Secreted halomucin complexes (stained green with specific antibody coupled to fluorescein) surrounds quadratic *Haloarcula marismortui* cell (stained red by Nile blue for polyhydroxy butyrate). Unpublished image, courtesy of D. Oesterhelt, Max Planck Institute of Biochemistry, Martinsried, Germany.

presence of a C-terminal membrane anchor or a putative sorting signal is a determinant for the localization of the mature protein. Because most of these motifs are relatively well conserved, as has been confirmed by numerous *in vivo* studies, it was possible to develop *in silico* programs that can provide more detailed information indicating whether a substrate is secreted into the extracellular environment, is tethered to a membrane by an amino-terminal transmembrane or lipid anchor, or is incorporated into a surface structure. Predictions such as these, in conjunction with *in vivo* and *in vitro* confirmations, can also deepen our understanding of the extent to which specific transport pathways and signal peptidases are used, which may point to the evolutionary pressures encountered by a particular species (see below).

A variety of software programs have been developed to predict specific signal peptide classes. Phobius, for example, is a combined transmembrane helix and SPase I processed Sec signal peptide prediction program (Kall et al., 2004). As laid out above, prokaryotic Sec substrates can contain signal peptides cleaved by either SPase I, SPase II, or SPase III, and in many prokaryotes structurally similar signal peptides can differentially target proteins to the Sec or Tat pathway (Bardy et al., 2003; Ng et al., 2007; Calo et al., 2010). Hence, to accurately predict the pathways employed to transport specific prokaryotic proteins, and their ultimate destinations, Phobius needs to be used in conjunction with programs that predict whether signal peptides contain a lipobox or a prepilin-peptidase processing site, as well as programs that can distinguish between Tat and Sec substrates.

TatFind and TatP are two programs that were specifically trained to identify Tat substrates (Rose et al., 2002; Bendtsen et al., 2005). Although structurally conserved, Tat signal peptide sequences vary somewhat across prokaryotic species. In some species, a small fraction of Tat substrates tolerate replacement of one of the arginines in the twin arginine motif with a lysine, while in others replacing either arginine disrupts transport (Berks et al., 2005). There is also some variability in the specificity of the amino acid residues allowed near the twin arginines, both among bacterial species and between prokaryotic domains (McDonough et al., 2008). However, TatFind, which was trained on a large set of predicted haloarchaeal Tat substrates, and TatP, have been validated by *in vivo* verification of a large number of substrates in archaea as well as bacteria, suggesting that Tat signal peptide sequences are reasonably similar across bacterial and archaeal species (Rose et al., 2002; Dilks et al., 2003, 2005; Gimenez et al., 2007; McDonough et al., 2008; Bagos et al., 2009; Storf et al., 2010; Uthandi et al., 2010). Consistent with this finding, Coulthurst et al. (2012) identified a Tat signal peptide recognition system in the euryarchaeon *Archaeoglobus fulgidus* resembling that of *Escherichia coli*.

Lipoprotein prediction programs such as LipoP (Juncker et al., 2003) and pred-lipo (Bagos et al., 2008), the latter being trained specifically on Gram-positive bacterial lipoproteins, were developed to identify bacterial Sec signal peptides containing SPase II processing sites. *In silico* and limited *in vivo* analyses suggest that the bacterial and archaeal lipoboxes are similar, and therefore, these programs can be used to identify archaeal lipoproteins (Storf et al., 2010). Using these programs in tandem gives the best results since although both programs identify most putative lipoproteins, each misses a subset. When used consecutively, TatFind and lipoprotein prediction programs can also predict Tat lipoproteins. However, the only prediction program trained on Tat substrates processed by SPase II is TatLipo, which was trained on haloarchaeal Tat substrates and is not yet fully tested on non-haloarchaeal proteins (Storf et al., 2010). Tat substrates not identified by LipoP or pred-lipo are assumed to be processed by SPase I.

Unlike the other types of signal peptides discussed here, signal peptides processed by SPase III show readily apparent differences in bacteria and archaea, requiring independent prediction programs be developed to specifically detect archaeal (FlaFind; Szabo et al., 2007b) and bacterial type IV pilin-like proteins (PilFind; Imam et al., 2011). In archaeal SPase III substrates, the most important residues for cleavage to occur are the amino acids at positions −2, −1, and +1, relative to the cleavage site. In particular, a positively charged amino acid residue at position −2 appears to be specific to archaeal SPase III recognition sites. Conversely, many archaeal SPase III substrates do not contain a glutamate residue at the +5 position, which appears to be a nearly universal feature of the signal peptides of type IV pilin-like proteins in bacteria (Imam et al., 2011). As noted above, Tat signal peptides containing SPase III processing sites have not been identified. Clearly, applying prediction programs in concert can clarify subcellular localization

of processed substrates by ensuring more accurate predictions of the class of the signal peptide that each substrate contains.

DIVERSE USE OF PROTEIN TRANSPORT PATHWAYS AND SURFACE-ANCHORING STRATEGIES

When applied to the large number of genome sequences completed in recent years, the prediction tools discussed above have enabled us to confirm trends first observed several years ago, and have also helped identify patterns in the usage of secretion pathways and protein anchoring mechanisms. These patterns may have significant implications for the evolutionary history or relationships of the organisms in which they are observed.

Genomic analyses indicate that the Sec protein transport pathway is conserved, and very likely essential, in all organisms (Pohlschroder et al., 2005b). The requirement of the Sec pathway for protein transport across the membrane as well as for the insertion of membrane proteins, strongly suggests that it evolved early in the history of living organisms with the insertion of membrane proteins preceding the transport of proteins with a SPase I-cleavable signal peptide. The Tat pathway likely evolved later, rerouting Sec substrates to this pathway, hence the similarities of their signal peptides. The fact that a highly diverse set of bacterial and archaeal species use the Tat pathway to transport secreted proteins suggests that this pathway was also present before the divergence of bacteria and archaea from the common ancestor (Dilks et al., 2003; Storf et al., 2010). The lack of this pathway in most eukaryotes is likely due to the transport of proteins into the lumen of the ER, an ATP-containing compartment that lacks many of the challenges of the extracytoplasmic environment prokaryotic secreted proteins face. Most prokaryotic species are predicted to use the Sec pathway to transport a large majority of their secreted proteins (Dilks et al., 2003; Storf et al., 2010). Interestingly, *in vivo* studies strongly support *in silico* analyses of nearly 20 haloarchaeal genomes showing that the species of this euryarchaeal class transport nearly half of their secreted proteins through the Tat pathway (Bolhuis, 2002; Rose et al., 2002; Dilks et al., 2005; Falb et al., 2005; Gimenez et al., 2007; Storf et al., 2010). Similar analyses performed on over a 100 non-haloarchaeal genomes indicated that the use of the Tat pathway to transport such a large percentage of secreted proteins may be unique to the halophilic archaea (Dilks et al., 2003). Perhaps, the extensive use of the Tat pathway is an adaptation to the high salt environments that haloarchaea inhabit. Highly negatively charged surfaces, providing them with a hydration shell, are a prerequisite for the stability of proteins exposed to high salt concentration and hence require rapid, efficient protein folding, a process that may be best accomplished in the cytoplasm, which contains ATP-dependent chaperones (Frolow et al., 1996; Kennedy et al., 2001). It should be noted that halophilic bacteria appear to use the Tat pathway for transporting only a very limited number of substrates. However, although they face the same high salt concentration challenges as haloarchaea, as noted above, halophilic bacteria express a Sec transport designated ATPase (SecA), possibly resulting in fundamental differences in Sec transport in these two prokaryotic domains that might have lead to distinctly different adaptations to high salt (Mongodin et al., 2005).

Analysis of the signal peptidase processing sites also suggest that the vast majority of haloarchaeal Tat substrates are lipoproteins, possibly because lipid anchoring avoids the need to insert a hydrophobic stretch of the secreted protein into the cytoplasmic membrane (Storf et al., 2010). For haloarchaea, transporting a carboxy-terminal hydrophobic anchor across a membrane and inserting it into that membrane in an unfolded conformation is problematic since the solubility of hydrophobic amino acids is low in a high salt environment – it is unlikely that the Tat pathway is able to laterally insert the transmembrane segments of their substrates as is observed for Sec substrates. Under these conditions, using chaperones to maintain the solubility of the long hydrophobic stretch of amino acids might be counterproductive since hydrophobic interactions between the carboxy-terminal anchor and the chaperones may be enhanced, ultimately resulting in the chaperones becoming affixed to the secreted substrate. The observation that only a small subset of non-haloarchaeal Tat substrates contain a predicted lipobox also supports the notion that the extensive use of lipid anchors by Tat substrates is likely an adaptation to high salt. Conversely, similar portions of non-haloarchaeal and haloarchaeal Sec substrates appear to have signal peptides that contain lipoboxes.

The genomes of most euryarchaeal species encode proteins having lipoboxes, again indicating the presence of lipid anchoring in the common ancestor of bacteria and archaea. The absence of archaeal homologs of the bacterial lipoprotein biosynthesis components may be due to the distinct cytoplasmic membrane compositions of bacteria and archaea. Interestingly, to date, *in silico* analyses of crenarchaeal genomes suggest that the crenarchaea lack lipoproteins (Storf et al., 2010). Perhaps crenarchaeal secreted proteins, compared to their euryarchaeal counterparts, are more frequently anchored to the membrane via amino- or carboxy-terminal transmembrane segments, or perhaps crenarchaea use as yet unidentified mechanisms to anchor or retain proteins, possibly by forming a periplasmic space (Ellen et al., 2010a). This is consistent with the fact that crenarchaea also appear to lack both archaeosortase homologs and substrates that contain the carboxy-terminal motifs targeted by an archaeosortase (Haft et al., 2012).

Conversely, species in nearly all archaeal kingdoms express type IV pilin-like proteins, although not every species within a given kingdom necessarily does, strongly suggesting that type IV pili are ancient surface structures (Szabo et al., 2007b; Pohlschroder et al., 2011). Interestingly, even the extraordinarily small genome of the symbiont *Nanoarchaeum equitans*, a species that cannot even produce its own lipids, encodes homologs of type IV pilus-biosynthesis pathway as well as FlaFind positive substrates (Huber et al., 2002; Szabo et al., 2007b). It is possible that these proteins play an integral role in their attachment to their symbiotic host *I. hospitalis* (Moissl-Eichinger and Huber, 2011). The lack of both lipobox-containing proteins and type IV pilins in eukaryotes is not surprising, again because proteins are transported into the ER-lumen, rather than directly across the cytoplasmic membrane.

To date, archaea adapted to extreme conditions other than high salt, such as high temperature or low pH, have not shown a preference for either of the known protein transport pathways, nor do they appear to have a preferred mechanism for anchoring secreted proteins to the cell surface (Storf et al., 2010).

OUTLOOK

The combination of *in vivo* and *in vitro* experiments with *in silico* predictions has greatly extended our understanding of the diversity of archaeal protein secretion and the localization of Sec and Tat substrates. However, additional strategies may have evolved for the export of proteins and their targeting. One possible approach toward the exploration of alternative secretion routes involves the proteomics of secreted proteins. Proteomic studies aim to generate a global view of all proteins present in a sample, and because quantitative methods are available, the relative expression of certain gene products under a variety of conditions can also be evaluated.

In recent years, there has been some progress on the proteomics of archaeal extracellular, surface exposed and membrane proteins (Chong and Wright, 2005; Saunders et al., 2006; Burghardt et al., 2008; Ellen et al., 2009, 2010b; Williams et al., 2010a,b). One typical observation was that proteins expected to be cytoplasmic, based on functional homology to cytoplasmic proteins, or the lack of a predicted Tat or Sec signal peptide are frequently recovered from the extracellular fractions. These proteins may originate from lysed cells, but in some cases could indicate alternative secretion strategies. Difficulties in determining which secreted proteins are transported through as yet undefined secretion pathways include an inability to identify motifs that target proteins to these pathways. Furthermore, although many predicted proteins are homologous to proteins having known functions, making the qualifier "hypothetical" obsolete, their functions have not yet been characterized. While computational analyses do allow us to make predictions, based on homology, concerning the general functions of these proteins, thereby indicating new directions for further investigations of their physiological function, homologous proteins sometimes have very different functions and hence predictions based on homology may be misleading with regard to subcellular localization. Most significantly though, the majority of predicted proteins show no homology to known proteins, making investigations of their functions and subcellular localization much more complicated.

Vesicles may be one possible mechanism for transporting proteins across the hydrophobic membrane in a Sec and Tat-independent manner. Extracellular vesicles have been detected in various archaea, including *Thermococcales* and *Sulfolobus* species as well as *I. hospitalis* (Prangishvili et al., 2000; Reysenbach et al., 2006; Soler et al., 2008; Ellen et al., 2009). Proteomic analysis of *Sulfolobus* derived S-layer coated vesicles revealed the presence of homologs of the eukaryotic endosomal sorting complex (ESCRT-III) proteins (Lindas et al., 2008; Samson et al., 2008). These proteins play an important role in cell division of *Sulfolobus* (Samson and Bell, 2011) and may thus have an additional role in eukaryotic-like vesicle budding and release. Unfortunately, the physiological roles of extracellular vesicles in archaea are not yet clear, although sulfolobicins are associated with these structures (Prangishvili et al., 2000; Ellen et al., 2011). Interestingly, *Sulfolobus* turreted icosahedral virus (STIV) particles released from the host cells contain ESCRT-III protein homologs, and the expression of these proteins in the host cell is up-regulated during viral infection. It is tempting to speculate that STIV hijacks the cellular vesicle release mechanism for its own biogenesis (Snyder and Young, 2011; Maaty et al., 2012).

The vesicles present in the periplasmic space of *I. hospitalis* bud off the inner membrane and probably fuse with the outer membrane (**Figure 4A**). Whether this is a strategy by which *I. hospitalis* outer membrane proteins are transported across the periplasmic space is not yet known. In fact, *I. hospitalis* requires the transport of many outer membrane proteins as this crenarchaeon is the only reported organism with an energized outer membrane and ATP synthesis within the periplasmic space (Kuper et al., 2010; **Figure 4A**). The presence of the outer membrane of the only diderm archaeon identified to date raises numerous intriguing questions such as why the type IV Iho670 pili required for *I. hospitalis* surface adhesion are secreted across the outer membrane (Muller et al., 2009), or how secreted outer membrane proteins that are likely to be required for cell–cell adhesion with the *I. hospitalis* symbiont *N. equitans* are being targeted to and transported across this outer membrane (Junglas et al., 2008; **Figure 4B**).

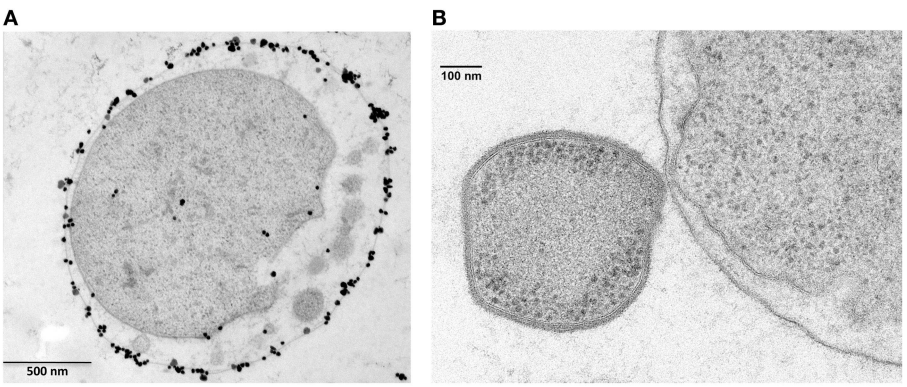


FIGURE 4 | ***Ignicoccus hospitalis*** **protein transport into and across the outer membrane. (A)** Localization of A_1A_O ATP synthase on *I. hospitalis* outer membrane by EM of ultrathin sections and labeled with antibodies specifically raised against the purified 440-kDa ATPase complex (Kuper et al., 2010). Image courtesy of T. Heimerl, H. Huber, and R. Rachel, University of Regensburg, Germany. **(B)** Electron micrographs of *I. hospitalis* ultrathin sections. Outer membrane of *I. hospitalis* and the cell surface of *N. equitans* in direct contact (Junglas et al., 2008). Image courtesy of T. Heimerl, H. Huber, and R. Rachel, and image **(B)** courtesy of H. Huber and R. Rachel, University of Regensburg, Germany.

In conclusion, further experimental studies, including proteomic analyses, will lead to the improvement of existing prediction programs and may allow the identification of phyla-specific sequences. However, along with a better understanding of "classical" secretory routes via the Tat and Sec pathways, the recent development of several eury- and crenarchaeal models (Leigh et al., 2011) combined with *in silico* analyses, such as the recent comparative genomics that identified the putative archaeosortases (Haft et al., 2012), may lead to the discovery of novel protein transport and anchoring strategies.

Exciting new trends have emerged from recent analyses of archaeal genomes that are shaping studies designed to identify and characterize the secreted proteins and the mechanisms underlying the transport, processing, and modification of secreted proteins. The most promising of these approaches combine *in vivo* and *in silico* analyses. Finally, researchers have identified a plethora of prokaryotic proteins having useful, commercially valuable enzymatic activities. Many of these enzymes are secreted proteins. Knowledge about the secretion pathways as well as signal peptide processing and other post-translational modifications will be invaluable in designing strategies for efficiently producing these proteins.

ACKNOWLEDGMENTS

Research at Microdish is in part supported by the EU Seventh Framework Large Scale Integrated Project "HotZyme". Mechthild Pohlschroder was supported by a National Science Foundation grant MCB02-MT and by a National Aeronautics and Space Administration grant NNX10AR84G.

REFERENCES

Ajon, M., Frols, S., Van Wolferen, M., Stoecker, K., Teichmann, D., Driessen, A. J., Grogan, D. W., Albers, S. V., and Schleper, C. (2011). UV-inducible DNA exchange in hyperthermophilic archaea mediated by type IV pili. *Mol. Microbiol.* 82, 807–817.

Albers, S. V., Konings, W. N., and Driessen, A. J. M. (1999). A unique short signal sequence in membrane-anchored proteins of archaea. *Mol. Microbiol.* 31, 1595–1596.

Albers, S. V., and Meyer, B. H. (2011). The archaeal cell envelope. *Nat. Rev. Microbiol.* 9, 414–426.

Albers, S. V., and Pohlschroder, M. (2009). Diversity of archaeal type IV pilin-like structures. *Extremophiles* 13, 403–410.

Albers, S. V., Szabo, Z., and Driessen, A. J. (2003). Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J. Bacteriol.* 185, 3918–3925.

Arts, J., van Boxtel, R., Filloux, A., Tommassen, J., and Koster, M. (2007). Export of the pseudopilin XcpT of the Pseudomonas aeruginosa type II secretion system via the signal recognition particle-Sec pathway. *J. Bacteriol.* 189, 2069–2076.

Bagos, P. G., Tsirigos, K. D., Liakopoulos, T. D., and Hamodrakas, S. J. (2008). Prediction of lipoprotein signal peptides in Gram-positive bacteria with a hidden Markov model. *J. Proteome Res.* 7, 5082–5093.

Bagos, P. G., Tsirigos, K. D., Plessas, S. K., Liakopoulos, T. D., and Hamodrakas, S. J. (2009). Prediction of signal peptides in archaea. *Protein Eng. Des. Sel.* 1, 27–35.

Balestrieri, M., Gogliettino, M., Fiume, I., Pocsfalvi, G., Catara, G., Rossi, M., and Palmieri, G. (2011). Structural and functional insights into *Aeropyrum pernix* OppA, a member of a novel archaeal OppA subfamily. *J. Bacteriol.* 193, 620–630.

Bardy, S. L., Eichler, J., and Jarrell, K. F. (2003). Archaeal signal peptides – a comparative survey at the genome level. *Protein Sci.* 12, 1833–1843.

Bardy, S. L., and Jarrell, K. F. (2002). FlaK of the archaeon *Methanococcus maripaludis* possesses preflagellin peptidase activity. *FEMS Microbiol. Lett.* 208, 53–59.

Bautista, V., Esclapez, J., Perez-Pomares, F., Martinez-Espinosa, R. M., Camacho, M., and Bonete, M. J. (2012). Cyclodextrin glycosyltransferase: a key enzyme in the assimilation of starch by the halophilic archaeon *Haloferax mediterranei*. *Extremophiles* 16, 147–159.

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005). Prediction of twin-arginine signal peptides. *BMC Bioinformatics* 6, 167. doi:10.1186/1471-2105-6-167

Berks, B. C., Palmer, T., and Sargent, F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Curr. Opin. Microbiol.* 8, 174–181.

Bolhuis, A. (2002). Protein transport in the halophilic archaeon *Halobacterium* sp. NRC-1: a major role for the twin-arginine translocation pathway? *Microbiology* 148, 3335–3346.

Bolhuis, H., Palm, P., Wende, A., Falb, M., Rampp, M., Rodriguez-Valera, F., Pfeiffer, F., and Oesterhelt, D. (2006). The genome of the square archaeon *Haloquadratum walsbyi*: life at the limits of water activity. *BMC Genomics* 7, 169. doi:10.1186/1471-2164-7-169

Brockl, G., Behr, M., Fabry, S., Hensel, R., Kaudewitz, H., Biendl, E., and Konig, H. (1991). Analysis and nucleotide sequence of the genes encoding the surface-layer glycoproteins of the hyperthermophilic methanogens *Methanothermus fervidus* and *Methanothermus sociabilis*. *Eur. J. Biochem.* 199, 147–152.

Burghardt, T., Saller, M., Gurster, S., Muller, D., Meyer, C., Jahn, U., Hochmuth, E., Deutzmann, R., Siedler, F., Babinger, P., Wirth, R., Huber, H., and Rachel, R. (2008). Insight into the proteome of the hyperthermophilic crenarchaeon *Ignicoccus hospitalis*: the major cytosolic and membrane proteins. *Arch. Microbiol.* 190, 379–394.

Calo, D., and Eichler, J. (2011). Crossing the membrane in archaea, the third domain of life. *Biochim. Biophys. Acta* 1808, 885–891.

Calo, D., Kaminski, L., and Eichler, J. (2010). Protein glycosylation in archaea: sweet and extreme. *Glycobiology* 20, 1065–1076.

Chaban, B., Ng, S. Y. M., Kanbe, M., Saltzman, L., Nimmo, G., Aizawa, S. I., and Jarrell, K. F. (2007). Systematic deletion analyses of the fla genes in the flagella operon identify several genes essential for proper assembly and function of flagella in the archaeon, *Methanococcus maripaludis*. *Mol. Microbiol.* 66, 596–609.

Cheung, J., Danna, K. J., O'Connor, E. M., Price, L. B., and Shand, R. F. (1997). Isolation, sequence, and expression of the gene encoding halocin H4, a bacteriocin from the halophilic archaeon *Haloferax mediterranei* R4. *J. Bacteriol.* 179, 548–551.

Chong, P. K., and Wright, P. C. (2005). Identification and characterization of the *Sulfolobus solfataricus* P2 proteome. *J. Proteome Res.* 4, 1789–1798.

Coulthurst, S. J., Dawson, A., Hunter, W. N., and Sargent, F. (2012). Conserved signal peptide recognition systems across the prokaryotic domains. *Biochemistry* 51, 1678–1686.

Dale, H., Angevine, C. M., and Krebs, M. P. (2000). Ordered membrane insertion of an archaeal opsin in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7847–7852.

De Castro, R. E., Ruiz, D. M., Gimenez, M. I., Silveyra, M. X., Paggi, R. A., and Maupin-Furlow, J. A. (2008). Gene cloning and heterologous synthesis of a haloalkaliphilic extracellular protease of *Natrialba magadii* (Nep). *Extremophiles* 12, 677–687.

de Vos, W. M., Voorhorst, W. G. B., Dijkgraaf, M., Kluskens, L. D., Van Der Oost, J., and Siezen, R. J. (2001). Purification, characterization, and molecular modeling of pyrolysin and other extracellular thermostable serine proteases from hyperthermophilic microorganisms. *Methods Enzymol.* 330, 383–393.

Dilks, K., Gimenez, M. I., and Pohlschroder, M. (2005). Genetic and biochemical analysis of the twin-arginine translocation pathway in halophilic archaea. *J. Bacteriol.* 187, 8104–8113.

Dilks, K., Rose, R. W., Hartmann, E., and Pohlschroder, M. (2003). Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. *J. Bacteriol.* 185, 1478–1483.

Driessen, A. J., and Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. *Annu. Rev. Biochem.* 77, 643–667.

Egorova, K., and Antranikian, G. (2005). Industrial relevance of thermophilic archaea. *Curr. Opin. Microbiol.* 8, 649–655.

Eichler, J. (2002). Archaeal signal peptidases from the genus *Thermoplasma*: structural and mechanistic hybrids of the bacterial and eukaryal enzymes. *J. Mol. Evol.* 54, 411–415.

Eichler, J., and Adams, M. W. (2005). Posttranslational protein modification in archaea. *Microbiol. Mol. Biol. Rev.* 69, 393–425.

Elferink, M. G. L., Albers, S. V., Konings, W. N., and Driessen, A. J. M. (2001). Sugar transport in *Sulfolobus solfataricus* is mediated by two families of binding protein-dependent ABC transporters. *Mol. Microbiol.* 39, 1494–1503.

Ellen, A. F., Albers, S. V., and Driessen, A. J. (2010a). Comparative study of the extracellular proteome of *Sulfolobus* species reveals limited secretion. *Extremophiles* 14, 87–98.

Ellen, A. F., Zolghadr, B., Driessen, A. M., and Albers, S. V. (2010b). Shaping the archaeal cell envelope. *Archaea* 2010, 608243.

Ellen, A. F., Albers, S. V., Huibers, W., Pitcher, A., Hobel, C. F. V., Schwarz, H., Folea, M., Schouten, S., Boekema, E. J., Poolman, B., and Driessen, A. J. M. (2009). Proteomic analysis of secreted membrane vesicles of archaeal *Sulfolobus* species reveals the presence of endosome sorting complex components. *Extremophiles* 13, 67–79.

Ellen, A. F., Rohulya, O. V., Fusetti, F., Wagner, M., Albers, S. V., and Driessen, A. J. (2011). The sulfolobicin genes of *Sulfolobus acidocaldarius* encode novel antimicrobial proteins. *J. Bacteriol.* 193, 4380–4387.

Engelhardt, H. (2007a). Are S-layers exoskeletons? The basic function of protein surface layers revisited. *J. Struct. Biol.* 160, 115–124.

Engelhardt, H. (2007b). Mechanism of osmoprotection by archaeal S-layers: a theoretical study. *J. Struct. Biol.* 160, 190–199.

Falb, M., Pfeiffer, F., Palm, P., Rodewald, K., Hickmann, V., Tittor, J., and Oesterhelt, D. (2005). Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res.* 15, 1336–1343.

Francetic, O., Buddelmeijer, N., Lewenza, S., Kumamoto, C. A., and Pugsley, A. P. (2007). Signal recognition particle-dependent inner membrane targeting of the PulG Pseudopilin component of a type II secretion system. *J. Bacteriol.* 189, 1783–1793.

Frolow, F., Harel, M., Sussman, J. L., Mevarech, M., and Shoham, M. (1996). Insights into protein adaptation to a saturated salt environment from the crystal structure of a halophilic 2Fe-2S ferredoxin. *Nat. Struct. Biol.* 3, 452–458.

Frols, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., Boekema, E. J., Driessen, A. J., Schleper, C., and Albers, S. V. (2008). UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. *Mol. Microbiol.* 70, 938–952.

Ghosh, A., and Albers, S. V. (2011). Assembly and function of the archaeal flagellum. *Biochem. Soc. Trans.* 39, 64–69.

Gimenez, M. I., Dilks, K., and Pohlschroder, M. (2007). *Haloferax volcanii* twin-arginine translocation substates include secreted soluble, C-terminally anchored and lipoproteins. *Mol. Microbiol.* 66, 1597–1606.

Gogliettino, M., Balestrieri, M., Pocsfalvi, G., Fiume, I., Natale, L., Rossi, M., and Palmieri, G. (2010). A highly selective oligopeptide binding protein from the archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 192, 3123–3131.

Gohlke, U., Pullan, L., Mcdevitt, C. A., Porcelli, I., De Leeuw, E., Palmer, T., Saibil, H. R., and Berks, B. C. (2005). The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10482–10486.

Grudnik, P., Bange, G., and Sinning, I. (2009). Protein targeting by the signal recognition particle. *Biol. Chem.* 390, 775–782.

Haft, D. H., Payne, S. H., and Selengut, J. D. (2012). Archaeosortases and exosortases are widely distributed systems linking membrane transit with posttranslational modification. *J. Bacteriol.* 194, 36–48.

Haseltine, C., Hill, T., Montalvo-Rodriguez, R., Kemper, S. K., Shand, R. F., and Blum, P. (2001). Secreted euryarchaeal microhalocins kill hyperthermophilic crenarchaea. *J. Bacteriol.* 183, 287–291.

Henche, A. L., Koerdt, A., Ghosh, A., and Albers, S. V. (2012). Influence of cell surface structures on crenarchaeal biofilm formation using a thermostable green fluorescent protein. *Environ. Microbiol.* 14, 779–793.

Horlacher, R., Xavier, K. B., Santos, H., Diruggiero, J., Kossmann, M., and Boos, W. (1998). Archaeal binding protein-dependent ABC transporter: molecular and biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis*. *J. Bacteriol.* 180, 680–689.

Huber, H., Hohn, M. J., Rachel, R., Fuchs, T., Wimmer, V. C., and Stetter, K. O. (2002). A new phylum of archaea represented by a nanosized hyperthermophilic symbiont. *Nature* 417, 63–67.

Hutcheon, G. W., Vasisht, N., and Bolhuis, A. (2005). Characterisation of a highly stable alpha-amylase from the halophilic archaeon *Haloarcula hispanica*. *Extremophiles* 9, 487–495.

Hutchings, M. I., Palmer, T., Harrington, D. J., and Sutcliffe, I. C. (2009). Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em. *Trends Microbiol.* 17, 13–21.

Imam, S., Chen, Z., Roos, D. S., and Pohlschroder, M. (2011). Identification of surprisingly diverse type IV pili, across a broad range of Gram-positive bacteria. *PLoS ONE* 6, e28919. doi:10.1371/journal.pone.0028919

Irihimovitch, V., and Eichler, J. (2003). Post-translational secretion of fusion proteins in the halophilic archaea *Haloferax volcanii*. *J. Biol. Chem.* 278, 12881–12887.

Jarrell, K. F., Jones, G. M., and Nair, D. B. (2010). Biosynthesis and role of N-linked glycosylation in cell surface structures of archaea with a focus on flagella and s layers. *Int. J. Microbiol.* 2010, 470138.

Jarrell, K. F., Stark, M., Nair, D. B., and Chong, J. P. (2011). Flagella and pili are both necessary for efficient attachment of *Methanococcus maripaludis* to surfaces. *FEMS Microbiol. Lett.* 319, 44–50.

Journet, L., Hughes, K. T., and Cornelis, G. R. (2005). Type III secretion: a secretory pathway serving both motility and virulence (review). *Mol. Membr. Biol.* 22, 41–50.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662.

Junglas, B., Briegel, A., Burghardt, T., Walther, P., Wirth, R., Huber, H., and Rachel, R. (2008). *Ignicoccus hospitalis* and *Nanoarchaeum equitans*: ultrastructure, cell-cell interaction, and 3D reconstruction from serial sections of freeze-substituted cells and by electron cryotomography. *Arch. Microbiol.* 190, 395–408.

Kall, L., Krogh, A., and Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338, 1027–1036.

Kamekura, M., Seno, Y., Holmes, M. L., and Dyallsmith, M. L. (1992). Molecular cloning and sequencing of the gene for halophilic alkaline serine protease (halolysin) from an unidentified halophilic strain (172P1) and expression of the gene in *Haloferax volcanii*. *J. Bacteriol.* 174, 736–742.

Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L., and Dassarma, S. (2001). Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* 11, 1641–1650.

Kikuchi, A., Sagami, H., and Ogura, K. (1999). Evidence for covalent attachment of diphytanylglyceryl phosphate to the cell-surface glycoprotein of *Halobacterium halobium*. *J. Biol. Chem.* 274, 18011–18016.

Kobayashi, T., Kanai, H., Aono, R., Horikoshi, K., and Kudo, T. (1994). Cloning, expression, and nucleotide sequence of the α-amylase gene from the haloalkaliphilic archaeon *Natronococcus* sp. strain Ah-36. *J. Bacteriol.* 176, 5131–5134.

Kokoeva, M. V., Storch, K. F., Klein, C., and Oesterhelt, D. (2002). A novel mode of sensory transduction in archaea: binding protein-mediated chemotaxis towards osmoprotectants and amino acids. *EMBO J.* 21, 2312–2322.

Konrad, Z., and Eichler, J. (2002). Lipid modification of proteins in archaea: attachment of a mevalonic acid-based lipid moiety to the surface-layer glycoprotein of *Haloferax vokanii* follows protein translocation. *Biochem. J.* 366, 959–964.

Krupovic, M., Peixeiro, N., Bettstetter, M., Rachel, R., and Prangishvili, D. (2012). Archaeal tetrathionate hydrolase goes viral: secretion of a sulfur metabolism enzyme in the form of virus-like particles. *Appl. Environ. Microbiol.* doi: 10.1128/AEM.01186-12. [Epub ahead of print].

Kuper, U., Meyer, C., Muller, V., Rachel, R., and Huber, H. (2010). Energized outer membrane and spatial separation of metabolic processes in the hyperthermophilic archaeon *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 3152–3156.

Kwan, D., and Bolhuis, A. (2010). Analysis of the twin-arginine motif of a haloarchaeal Tat substrate. *FEMS Microbiol. Lett.* 308, 138–143.

Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S. V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124.

Leake, M. C., Greene, N. P., Godun, R. M., Granjon, T., Buchanan, G., Chen, S., Berry, R. M., Palmer, T., and Berks, B. C. (2008). Variable stoichiometry of the TatA component of the twin-arginine protein transport

system observed by in vivo single-molecule imaging. *Proc. Natl. Acad. Sci. U.S.A.* 105, 15376–15381.

Leigh, J. A., Albers, S. V., Atomi, H., and Allers, T. (2011). Model organisms for genetics in the domain archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol. Rev.* 35, 577–608.

Lindas, A. C., Karlsson, E. A., Lindgren, M. T., Ettema, T. J., and Bernander, R. (2008). A unique cell division machinery in the Archaea. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18942–18946.

Maaty, W. S., Selvig, K., Ryder, S., Tarlykov, P., Hilmer, J. K., Heinemann, J., Steffens, J., Snyder, J. C., Ortmann, A. C., Movahed, N., Spicka, K., Chetia, L., Grieco, P. A., Dratz, E. A., Douglas, T., Young, M. J., and Bothner, B. (2012). Proteomic analysis of *Sulfolobus solfataricus* during *Sulfolobus* turreted icosahedral virus infection. *J. Proteome Res.* 11, 1420–1432.

Mai, B., Frey, G., Swanson, R. V., Mathur, E. J., and Stetter, K. O. (1998). Molecular cloning and functional expression of a protein-serine/threonine phosphatase from the hyperthermophilic archaeon *Pyrodictium abyssi* TAG11. *J. Bacteriol.* 180, 4030–4035.

Mattar, S., Scharf, B., Kent, S. B., Rodewald, K., Oesterhelt, D., and Engelhard, M. (1994). The primary structure of halocyanin, an archaeal blue copper protein, predicts a lipid anchor for membrane fixation. *J. Biol. Chem.* 269, 14939–14945.

Mayr, J., Lupas, A., Kellermann, J., Eckerskorn, C., Baumeister, W., and Peters, J. (1996). A hyperthermostable protease of the subtilisin family bound to the surface layer of the Archaeon *Staphylothermus marinus*. *Curr. Biol.* 6, 739–749.

McDonough, J. A., Mccann, J. R., Tekippe, E. M., Silverman, J. S., Rigel, N. W., and Braunstein, M. (2008). Identification of functional Tat signal sequences in *Mycobacterium tuberculosis* proteins. *J. Bacteriol.* 190, 6428–6438.

Mescher, M. F., Strominger, J. L., and Watson, S. W. (1974). Protein and carbohydrate composition of the cell envelope of *Halobacterium salinarium*. *J. Bacteriol.* 120, 945–954.

Meyer, B. H., Zolghadr, B., Peyfoon, E., Pabst, M., Panico, M., Morris, H. R., Haslam, S. M., Messner, P., Schaffer, C., Dell, A., and Albers, S. V. (2011). Sulfoquinovose synthase – an important enzyme in the N-glycosylation pathway of *Sulfolobus acidocaldarius*. *Mol. Microbiol.* 82, 1150–1163.

Moissl, C., Rachel, R., Briegel, A., Engelhardt, H., and Huber, R. (2005). The unique structure of archaeal "hami," highly complex cell appendages with nano-grappling hooks. *Mol. Microbiol.* 56, 361–370.

Moissl-Eichinger, C., and Huber, H. (2011). Archaeal symbionts and parasites. *Curr. Opin. Microbiol.* 14, 364–370.

Mongodin, E. F., Nelson, K. E., Daugherty, S., Deboy, R. T., Wister, J., Khouri, H., Weidman, J., Walsh, D. A., Papke, R. T., Sanchez Perez, G., Sharma, A. K., Nesbo, C. L., Macleod, D., Bapteste, E., Doolittle, W. F., Charlebois, R. L., Legault, B., and Rodriguez-Valera, F. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152.

Moracci, M., Cobucci-Ponzano, B., Perugino, G., and Rossi, M. (2007). "Biotechnology," in *Archaea: Molecular and Cellular Biology*, ed. R. Cavicchioli (Washington: ASM Press), 478–495.

Mori, T., Ishitani, R., Tsukazaki, T., Nureki, O., and Sugita, Y. (2010). Molecular mechanisms underlying the early stage of protein translocation through the Sec translocon. *Biochemistry* 49, 945–950.

Muller, D. W., Meyer, C., Gurster, S., Kuper, U., Huber, H., Rachel, R., Wanner, G., Wirth, R., and Bellack, A. (2009). The Iho670 fibers of *Ignicoccus hospitalis*: a new type of archaeal cell surface appendage. *J. Bacteriol.* 191, 6465–6468.

Natale, P., Bruser, T., and Driessen, A. J. (2008). Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane – distinct translocases and mechanisms. *Biochim. Biophys. Acta* 1778, 1735–1756.

Nather, D. J., and Rachel, R. (2004). The outer membrane of the hyperthermophilic archaeon *Ignicoccus*: dynamics, ultrastructure and composition. *Biochem. Soc. Trans.* 32, 199–203.

Nather, D. J., Rachel, R., Wanner, G., and Wirth, R. (2006). Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts. *J. Bacteriol.* 188, 6915–6923.

Ng, S. Y., Chaban, B., Vandyke, D. J., and Jarrell, K. F. (2007). Archaeal signal peptidases. *Microbiology* 153, 305–314.

Ng, S. Y., Vandyke, D. J., Chaban, B., Wu, J., Nosaka, Y., Aizawa, S., and Jarrell, K. F. (2009). Different minimal signal peptide lengths recognized by the archaeal prepilin-like peptidases FlaK and PibD. *J. Bacteriol.* 191, 6732–6740.

Ng, S. Y., Wu, J., Nair, D. B., Logan, S. M., Robotham, A., Tessier, L., Kelly, J. F., Uchida, K., Aizawa, S., and Jarrell, K. F. (2011). Genetic and mass spectrometry analyses of the unusual type IV-like pili of the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 193, 804–814.

O'Connor, E. M., and Shand, R. F. (2002). Halocins and sulfolobicins: the emerging story of archaeal protein and peptide antibiotics. *J. Ind. Microbiol. Biotechnol.* 28, 23–31.

Okuda, S., and Tokuda, H. (2011). Lipoprotein sorting in bacteria. *Annu. Rev. Microbiol.* 65, 239–259.

Ortenberg, R., and Mevarech, M. (2000). Evidence for post-translational membrane insertion of the integral membrane protein bacterioopsin expressed in the heterologous halophilic archaeon *Haloferax volcanii*. *J. Biol. Chem.* 275, 22839–22846.

Paetzel, M., Karla, A., Strynadka, N. C. J., and Dalbey, R. E. (2002). Signal peptidases. *Chem. Rev.* 102, 4549–4580.

Peyfoon, E., Meyer, B., Hitchen, P. G., Panico, M., Morris, H. R., Haslam, S. M., Albers, S. V., and Dell, A. (2010). The S-layer glycoprotein of the crenarchaeote *Sulfolobus acidocaldarius* is glycosylated at multiple sites with chitobiose-linked N-glycans. *Archaea.* PMID: 20936123.

Pohlschroder, M., Ghosh, A., Tripepi, M., and Albers, S. V. (2011). Archaeal type IV pilus-like structures – evolutionarily conserved prokaryotic surface organelles. *Curr. Opin. Microbiol.* 14, 357–363.

Pohlschroder, M., Gimenez, M. I., and Jarrell, K. F. (2005a). Protein transport in archaea: Sec and twin arginine translocation pathways. *Curr. Opin. Microbiol.* 8, 713–719.

Pohlschroder, M., Hartmann, E., Hand, N. J., Dilks, K., and Haddad, A. (2005b). Diversity and evolution of protein translocation. *Annu. Rev. Microbiol.* 59, 91–111.

Prangishvili, D., Holz, I., Stieger, E., Nickell, S., Kristjansson, J. K., and Zillig, W. (2000). Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus *Sulfolobus*. *J. Bacteriol.* 182, 2985–2988.

Price, L. B., and Shand, R. F. (2000). Halocin S8: a 36-amino-acid microhalocin from the haloarchaeal strain S8a. *J. Bacteriol.* 182, 4951–4958.

Protze, J., Muller, F., Lauber, K., Nass, B., Mentele, R., Lottspeich, F., and Kletzin, A. (2011). An extracellular tetrathionate hydrolase from the thermoacidophilic archaeon *Acidianus ambivalens* with an activity optimum at pH 1. *Front. Microbiol.* 2:68. doi:10.3389/fmicb.2011.00068

Reysenbach, A. L., Liu, Y. T., Banta, A. B., Beveridge, T. J., Kirshtein, J. D., Schouten, S., Tivey, M. K., Von Damm, K. L., and Voytek, M. A. (2006). A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. *Nature* 442, 444–447.

Richardson, T. H., Tan, X. Q., Frey, G., Callen, W., Cabell, M., Lam, D., Macomber, J., Short, J. M., Robertson, D. E., and Miller, C. (2002). A novel, high performance enzyme for starch liquefaction – discovery and optimization of a low pH, thermostable α-amylase. *J. Biol. Chem.* 277, 26501–26507.

Robinson, C., Matos, C. F., Beck, D., Ren, C., Lawrence, J., Vasisht, N., and Mendel, S. (2011). Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria. *Biochim. Biophys. Acta* 1808, 876–884.

Rose, R. W., Bruser, T., Kissinger, J. C., and Pohlschroder, M. (2002). Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway. *Mol. Microbiol.* 45, 943–950.

Rudolph, C., Wanner, G., and Huber, R. (2001). Natural communities of novel archaea and bacteria growing in cold sulfurous springs with a string-of-pearls-like morphology. *Appl. Environ. Microbiol.* 67, 2336–2344.

Samson, R. Y., and Bell, S. D. (2011). Cell cycles and cell division in the archaea. *Curr. Opin. Microbiol.* 14, 350–356.

Samson, R. Y., Obita, T., Freund, S. M., Williams, R. L., and Bell, S. D. (2008). A role for the ESCRT system in cell division in archaea. *Science* 322, 1710–1713.

Sankaran, K., and Wu, H. C. (1994). Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol. *J. Biol. Chem.* 269, 19701–19706.

Saunders, N. F. W., Ng, C., Raftery, M., Guilhaus, M., Goodchild, A., and Cavicchioli, R. (2006).

Proteomic and computational analysis of secreted proteins with type 1 signal peptides from the antarctic archaeon *Methanococcoides burtonii*. *J. Proteome Res.* 5, 2457–2464.

Shi, W. L., Tang, X. F., Huang, Y. P., Gan, F., Tang, B., and Shen, P. (2006). An extracellular halophilic protease SptA from a halophilic archaeon *Natrinema* sp J7: gene cloning, expression and characterization. *Extremophiles* 10, 599–606.

Snyder, J. C., and Young, M. J. (2011). Potential role of cellular ESCRT proteins in the STIV life cycle. *Biochem. Soc. Trans.* 39, 107–110.

Soler, N., Marguet, E., Verbavatz, J. M., and Forterre, P. (2008). Virus-like vesicles and extracellular DNA produced by hyperthermophilic archaea of the order Thermococcales. *Res. Microbiol.* 159, 390–399.

Spirig, T., Weiner, E. M., and Clubb, R. T. (2011). Sortase enzymes in Gram-positive bacteria. *Mol. Microbiol.* 82, 1044–1059.

Storf, S., Pfeiffer, F., Dilks, K., Chen, Z. Q., Imam, S., and Pohlschroder, M. (2010). Mutational and bioinformatic analysis of haloarchaeal lipobox-containing proteins. *Archaea* 2010, 11.

Sublimi Saponetti, M., Bobba, F., Salerno, G., Scarfato, A., Corcelli, A., and Cucolo, A. (2011). Morphological and structural aspects of the extremely halophilic archaeon *Haloquadratum walsbyi*. *PLoS ONE* 6, e18653. doi:10.1371/journal.pone.0018653

Sumper, M., Berg, E., Mengele, R., and Strobel, I. (1990). Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. *J. Bacteriol.* 172, 7111–7118.

Sun, C., Li, Y., Mei, S., Lu, Q., Zhou, L., and Xiang, H. (2005). A single gene directs both production and immunity of halocin C8 in a haloarchaeal strain AS7092. *Mol. Microbiol.* 57, 537–549.

Szabo, Z., Sani, M., Groeneveld, M., Zolghadr, B., Schelert, J., Albers, S. V., Blum, P., Boekema, E. J., and Driessen, A. J. (2007a). Flagellar motility and structure in the hyperthermoacidophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.* 189, 4305–4309.

Szabo, Z., Stahl, A. O., Albers, S. V., Kissinger, J. C., Driessen, A. J., and Pohlschroder, M. (2007b). Identification of diverse archaeal proteins with class III signal peptides cleaved by distinct archaeal prepilin peptidases. *J. Bacteriol.* 189, 772–778.

Thoma, C., Frank, M., Rachel, R., Schmid, S., Nather, D., Wanner, G., and Wirth, R. (2008). The Mth60 fimbriae of *Methanothermobacter thermoautotrophicus* are functional adhesins. *Environ. Microbiol.* 10, 2785–2795.

Thompson, B. J., Widdick, D. A., Hicks, M. G., Chandra, G., Sutcliffe, I. C., Palmer, T., and Hutchings, M. I. (2010). Investigating lipoprotein biogenesis and function in the model Gram-positive bacterium *Streptomyces coelicolor*. *Mol. Microbiol.* 77, 943–957.

Tripepi, M., Imam, S., and Pohlschroder, M. (2010). *Haloferax volcanii* flagella are required for motility but are not involved in PibD-dependent surface adhesion. *J. Bacteriol.* 192, 3093–3102.

Tuteja, R. (2005). Type I signal peptidase: an overview. *Arch. Biochem. Biophys.* 441, 107–111.

Uthandi, S., Saad, B., Humbard, M. A., and Maupin-Furlow, J. A. (2010). LccA, an archaeal laccase secreted as a highly stable glycoprotein into the extracellular medium by *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76, 733–743.

Van den Berg, B., Clemons, W. M. Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., and Rapoport, T. A. (2004). X-ray structure of a protein-conducting channel. *Nature* 427, 36–44.

VanDyke, D. J., Wu, J., Ng, S. Y., Kanbe, M., Chaban, B., Aizawa, S., and Jarrell, K. F. (2008). Identification of a putative acetyltransferase gene, MMP0350, which affects proper assembly of both flagella and pili in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 190, 5300–5307.

Wang, Y. A., Yu, X., Ng, S. Y., Jarrell, K. F., and Egelman, E. H. (2008). The structure of an archaeal pilus. *J. Mol. Biol.* 381, 456–466.

Wende, A., Johansson, P., Vollrath, R., Dyall-Smith, M., Oesterhelt, D., and Grininger, M. (2010). Structural and biochemical characterization of a halophilic archaeal alkaline phosphatase. *J. Mol. Biol.* 400, 52–62.

Williams, T. J., Burg, D. W., Ertan, H., Raftery, M. J., Poljak, A., Guilhaus, M., and Cavicchioli, R. (2010a). Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcoides burtonii*. Part II: the effect of different methylated growth substrates. *J. Proteome Res.* 9, 653–663.

Williams, T. J., Burg, D. W., Raftery, M. J., Poljak, A., Guilhaus, M., Pilak, O., and Cavicchioli, R. (2010b). Global proteomic analysis of the insoluble, soluble, and supernatant fractions of the psychrophilic archaeon *Methanococcoides burtonii*. Part I: the effect of growth temperature. *J. Proteome Res.* 9, 640–652.

Worthington, P., Hoang, V., Perez-Pomares, F., and Blum, P. (2003). Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataficus*. *J. Bacteriol.* 185, 482–488.

Xu, Z., Du, X., Li, T., Gan, F., Tang, B., and Tang, X.-F. (2011). Functional insight into the C-terminal extension of halolysin SptA from haloarchaeon *Natrinema* sp J7. *PLoS ONE* 6, e23562. doi:10.1371/journal.pone.0023562

Yuan, J., Zweers, J. C., Van Dijl, J. M., and Dalbey, R. E. (2010). Protein transport across and into cell membranes in bacteria and archaea. *Cell. Mol. Life Sci.* 67, 179–199.

Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011). Protein translocation across the ER membrane. *Biochim. Biophys. Acta* 1808, 912–924.

Zolghadr, B., Klingl, A., Koerdt, A., Driessen, A. J., Rachel, R., and Albers, S. V. (2010). Appendage-mediated surface adherence of *Sulfolobus solfataricus*. *J. Bacteriol.* 192, 104–110.

Zolghadr, B., Weber, S., Szabo, Z., Driessen, A. J., and Albers, S. V. (2007). Identification of a system required for the functional surface localization of sugar binding proteins with class III signal peptides in *Sulfolobus solfataricus*. *Mol. Microbiol.* 64, 795–806.

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

Received: 16 March 2012; accepted: 21 May 2012; published online: 02 July 2012. Citation: Szabo Z and Pohlschroder M (2012) Diversity and subcellular distribution of archaeal secreted proteins. Front. Microbio. ***3****:207. doi: 10.3389/fmicb.2012.00207*
This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 18 June 2012
doi: 10.3389/fmicb.2012.00224

Genetic and biochemical identification of a novel single-stranded DNA-binding complex in *Haloferax volcanii*

Amy Stroud[1], Susan Liddell[2] and Thorsten Allers[1]*

[1] School of Biology, Queen's Medical Centre, University of Nottingham, Nottingham, UK
[2] Division of Animal Sciences, University of Nottingham, Loughborough, UK

Edited by:
Zvi Kelman, University of Maryland, USA

Reviewed by:
Jocelyne DiRuggiero, The Johns Hopkins University, USA
Stuart MacNeill, University of St Andrews, UK

***Correspondence:**
Thorsten Allers, School of Biology, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK.
e-mail: thorsten.allers@nottingham.ac.uk

Single-stranded DNA (ssDNA)-binding proteins play an essential role in DNA replication and repair. They use oligonucleotide/oligosaccharide-binding (OB)-folds, a five-stranded β-sheet coiled into a closed barrel, to bind to ssDNA thereby protecting and stabilizing the DNA. In eukaryotes the ssDNA-binding protein (SSB) is known as replication protein A (RPA) and consists of three distinct subunits that function as a heterotrimer. The bacterial homolog is termed SSB and functions as a homotetramer. In the archaeon *Haloferax volcanii* there are three genes encoding homologs of RPA. Two of the *rpa* genes (*rpa1* and *rpa3*) exist in operons with a novel gene specific to Euryarchaeota; this gene encodes a protein that we have termed RPA-associated protein (*rpap*). The *rpap* genes encode proteins belonging to COG3390 group and feature OB-folds, suggesting that they might cooperate with RPA in binding to ssDNA. Our genetic analysis showed that *rpa1* and *rpa3* deletion mutants have differing phenotypes; only Δ*rpa3* strains are hypersensitive to DNA damaging agents. Deletion of the *rpa3*-associated gene *rpap3* led to similar levels of DNA damage sensitivity, as did deletion of the *rpa3* operon, suggesting that RPA3 and RPAP3 function in the same pathway. Protein pull-downs involving recombinant hexahistidine-tagged RPAs showed that RPA3 co-purifies with RPAP3, and RPA1 co-purifies with RPAP1. This indicates that the RPAs interact only with their respective associated proteins; this was corroborated by the inability to construct *rpa1 rpap3* and *rpa3 rpap1* double mutants. This is the first report investigating the individual function of the archaeal COG3390 RPA-associated proteins (RPAPs). We have shown genetically and biochemically that the RPAPs interact with their respective RPAs, and have uncovered a novel single-stranded DNA-binding complex that is unique to Euryarchaeota.

Keywords: archaea, *Haloferax volcanii*, RPA single-strand DNA-binding protein, COG3390 RPA-associated protein, DNA repair, protein overexpression, Cdc48d

INTRODUCTION

Genomic DNA must be unwound in order to be replicated or repaired, leaving it vulnerable to nuclease and chemical attack as well as open to the possibility of forming secondary structures. Binding of the single-stranded DNA (ssDNA)-binding proteins (SSB) RPA and SSB prevents any of these events from occurring (Lu et al., 2009). The SSB is denominated SSB in bacteria and replication protein A (RPA) in eukaryotes; they bind to ssDNA with high affinity and to dsDNA and RNA with low affinity (Wobbe et al., 1987; Wold et al., 1989; Kim et al., 1992). They play a vital organizational role in the central genome maintenance of the cell, providing docking platforms for a wide range of enzymes to gain access to genomic substrates (Lu and Keck, 2008). The bacteriophage T4 gene 32 monomer was the first SSB to be identified (Alberts and Frey, 1970). RPA was first identified as an essential protein for DNA replication in the eukaryotic simian virus (SV40; Wobbe et al., 1987) by stimulating the T antigen-mediated unwinding of the SV40 origin of replication (Kenny et al., 1989). RPA and SSB have now been established as essential proteins for DNA metabolism including DNA replication, recombination, and repair in all domains of life (Wobbe et al., 1987; Heyer et al., 1990; Coverley et al., 1991, 1992; Moore et al., 1991; Wold, 1997). The basic architecture of RPA and SSB is based on the oligonucleotide/oligosaccharide-binding (OB)-fold, a five-stranded β-sheet coiled into a closed barrel, but the number of OB-folds present varies from species to species (Bochkarev and Bochkareva, 2004; Fanning et al., 2006).

Unlike in bacteria and eukaryotes, the architecture of SSBs present in archaea is not uniform. There is wide diversity of SSBs in the two main archaeal phyla, Euryarchaea, and Crenarchaea. Crenarchaea possess SSBs similar to those of bacteria, consisting of a single subunit with one OB-fold and an acidic C-terminus tail (Rolfsmeier and Haseltine, 2010). Euryarchaea have RPA-like proteins that show homology with the eukaryotic RPA, but from species to species the architecture of euryarchaeal RPAs varies dramatically from a single polypeptide RPA to an RPA made up of several subunits. Each of these RPAs can contain up to four OB-folds as well as a zinc finger motif (Rolfsmeier and Haseltine, 2010).

Pyrococcus furiosus RPA consists of three subunits RPA41, 14, and 32, denominated RPA1, 2, and 3, respectively, which form a heterotrimer as seen in eukaryotes. Strand exchange and immunoprecipitation assays have shown that *P. furiosus* heterotrimeric RPA stimulates strand exchange, and interacts with the clamp loader RFC and both DNA polymerases B and D (Komori and Ishino, 2001). The heterotrimeric complex seen in *P. furiosus* is also found in *P. abyssi* and *P. horikoshii*. However, in other archaeal species the *rpa* genes have undergone lineage-specific duplications, resulting in differing numbers of SSBs with diverse structures. Unlike the RPA complex found in *Pyrococcus* spp., or eukaryotic RPA, these do not form trimeric complexes (Robbins et al., 2004).

Methanosarcina acetivorans possesses three RPA subunits, MacRPA1, 2, and 3, which are unlikely to form a heterotrimeric complex as seen in *P. furiosus* and in eukaryotes. MacRPA1 contains four DNA-binding domains (DBD) containing OB-folds, MacRPA2 and 3 both have two OB-fold containing DBDs. Each of the three MacRPAs can function as SSBs, and are able to stimulate primer extension by *M. acetivorans* DNA polymerase BI (Robbins et al., 2004, 2005). This demonstrates an element of redundancy between the three MacRPAs, and suggests that the heterotrimeric RPA structure observed in *P. furiosus* is the exception and not the rule. Lin et al. (2008) suggest that intramolecular recombination between RPA homologs may have led to the diversity of RPAs found in euryarchaea, which can function in different pathways or cellular processes.

A similar pattern of lineage-specific gene duplication is seen with the archaeal MCM helicase, where the number and type of MCM subunits that make up the hexameric helicase differ between archaeal species. The genes encoding the MCM subunits fall into distinct phylogenetic clades, but these do not correspond to specific subunits of eukaryotic MCM. Instead they have arisen through lineage-specific gene family expansion (Chia et al., 2010). Such gene duplication might allow different archaeal species to refine the structure and function of MCM (and potentially RPA) for differing conditions and specialized roles.

Sulfolobus solfataricus has a bacterial-like SSB consisting of a small 20 kDa peptide containing one OB-fold and an acidic C-terminus tail (Haseltine and Kowalczykowski, 2002; Rolfsmeier and Haseltine, 2010). The *S. solfataricus* SSB quaternary structure is similar to that of *E. coli* SSB, however the primary structure of the OB-fold shows greater homology to that of the eukaryotic RPA70 DNA-binding domain B (DBDB). This suggests that crenarchaeal SSBs may be structurally similar to bacterial SSB but at a protein sequence level show homology to the eukaryotic RPA (Haseltine and Kowalczykowski, 2002; Kerr et al., 2003). In *S. solfataricus* there is an absence of DNA damage recognition proteins such as homologs of XPA or XPC to initiate NER. The ability of *S. solfataricus* SSB to specifically bind and melt damaged duplex DNA *in vitro* suggests SSB may play a role in the identification and binding of damaged DNA, followed by the subsequent recruitment of NER repair proteins (Cubeddu and White, 2005).

Haloferax volcanii encodes three RPA genes *rpa1*, *rpa2*, and *rpa3* (Hartman et al., 2010). Recent studies have shown RPA2 to be essential while RPA1 and RPA3 are not (Skowyra and MacNeill, 2012). Note that these authors used the nomenclature *rpaA1*, *A2*, *B1*, *B2*, *B3*, and *C* to refer to *rpa3*, *rpap3*, *rpa1*, *rpap1*, *rpe*, and *rpa2*, respectively, while we have chosen to maintain the official nomenclature as described in Table 4 of the *H. volcanii* genome paper (Hartman et al., 2010). Both *rpa1* and *rpa3* are in operons with other genes; *rpa1* is in an operon with genes encoding an OB-fold containing protein (hereby designated RPA-associated protein or RPAP) and a calcineurin-like phosphoesterase, while only one OB-fold rpa-associated protein (*rpap*) gene is present in the *rpa3* operon (**Figure 1**). The presence of an *rpap* gene in the same operon as *rpa* can be found in other euryarchaeota, including *Halobacterium marismortui*, *Halobacterium salinarum*, and *Natronomonas pharaonis*, as well as in *M. mazei* and *M. barkeri*. The *rpap* gene has been assigned to the cluster of orthologous groups (COG) 3390 (Berthon et al., 2008).

To examine if RPA1 and 3, as well as RPAP1 and RPAP3 play a role in DNA repair, as is true for both the bacterial SSB and eukaryotic RPA, DNA damage assays were performed using the single and operon deletion mutants. Cells with deletions of the *rpa1* and *rpa3* operons had previously been examined by Skowyra and MacNeill (2012). However, this is the first report investigating the individual function of the archaeal COG3390 RPAP. We show genetically and biochemically that the RPAPs interact with their respective RPAs, and have thereby uncovered a novel SSB complex that is unique to Euryarchaeota.

MATERIALS AND METHODS

All chemicals were from Sigma and restriction enzymes from New England Biolabs, unless stated otherwise. Standard molecular techniques were used (Sambrook and Russell, 2001).

STRAINS AND PLASMIDS

Haloferax volcanii strains (**Table 1**) were grown at 45°C on complete (Hv-YPC), casamino acids (Hv-Ca), or minimal (Hv-Min) agar, or in Hv-YPC or Hv-Ca broth as described previously. Isolation of genomic and plasmid DNA, as well as transformation of *H. volcanii* were carried out as described previously (Allers et al., 2004).

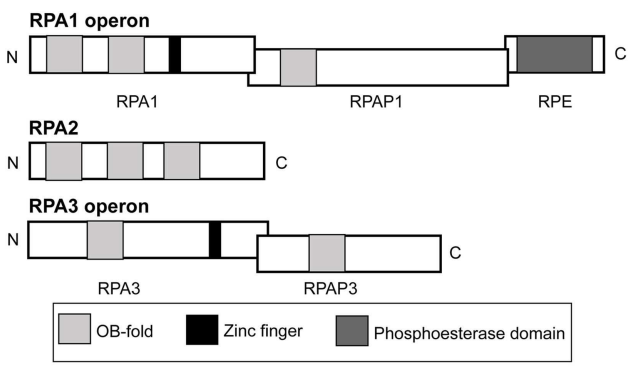


FIGURE 1 | Operon and domain structures of *H. volcanii* single-stranded DNA-binding proteins. Genes for RPA1 and RPA3 are in operons with genes for RPA-associated proteins, RPAP1 and RPAP3, respectively. The gene for RPA1 phosphoesterase (RPE) is present in the *rpa1* operon. Domains (not to scale) comprising OB-folds, zinc fingers and a phosphoesterase motif are shown.

Table 1 | *Haloferax volcanii* strains.

Strain	Relevant genotype*	Source or reference
DS2	Wild-type	Mullakhanbhai and Larsen (1975)
H195	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB*	Guy et al. (2006)
H1209	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr*	Allers et al. (2010)
H1216	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpap1*::*trpA*+	H195 pTA1170
H1217	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa1*::*trpA*+	H195 pTA1166
H1244	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa3*::*trpA*+	H195 pTA1174
H1246	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa1 operon*	H195 pTA1189
H1260	Δ*pyrE2* Δ*hdrB bgaHa-Bb* Δ*rpa3 operon*::*trpA*+ *leuB-Ag1* Δ*trpA*	H195 pTA1207
H1280	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpap1*	H1216 pTA1217
H1281	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa1*	H1217 pTA1141
H1282	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa1 operon rpa3 operon*+::[Δ*rpa3 operon*::*trpA*+, *pyrE2*+]	H1246 pTA1207 pop-in
H1326	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpap1 rpa3*+::[Δ*rpa3*::*trpA*+ *pyrE2*+]	H1280 pTA1174 pop-in
H1333	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr Hsa-cdc48d*	H1209 pTA1240
H1390	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa1 operon rpa3 operon*+::[Δ*rpa3 operon*::*trpA*+, *pyrE2*+] <*rpa1* operon+*hdrB*+*pyr*E2>	H1282 pTA1265 pop-in
H1410	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpap3*	H195 pTA1284
H1424	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr cdc48d-Ct*	H1333 pTA1294
H1430	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr cdc48d-Ct* <*p.tna*::*his tag*+*pyrE2*+ *hdrB*+>	H1424 pTA963
H1473	Δ*pyrE2 bgaHa-Bb leuB-Ag1* Δ*trpA* Δ*hdrB* Δ*rpa1 rpap3*+::[Δ*rpap3*::*trpA*+ *pyrE2*+]	H1281 pTA1284 pop-in
H1480	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr cdc48d-Ct* <*p.tnaA*::*his tag-rpa3 rpap3 pyrE2*+ *hdrB*+>	H1424 pTA1280
H1481	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr cdc48d-Ct* <*p.tnaA*::*rpa3 his tag-rpap3 pyrE2*+ *hdrB*+>	H1424 pTA1281
H1482	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr cdc48d-Ct* <*p.tnaA*::*his tag-rpa1 rpap1 pyrE2*+ *hdrB*+>	H1424 pTA1326
H1483	Δ*pyrE2* Δ*hdrB Nph-pitA* Δ*mrr cdc48d-Ct* <*p.tnaA*::*rpa1 his tag-rpap1 pyrE2*+ *hdrB*+>	H1424 pTA1327

**Genes shown within <> are present on an episomal plasmid, genes shown within [] are present on an integrated plasmid (pop-in).*

CONSTRUCTION OF MUTANT STRAINS

Deletion mutants were constructed as described previously (Allers et al., 2004). Plasmids for gene deletion are shown in **Table 2**, and were generated by PCR using oligonucleotides shown in **Table 3**. Template DNA for the PCRs was isolated from genomic DNA.

CONSTRUCTION OF PROTEIN OVEREXPRESSION STRAINS

Protein overexpression strains were constructed by transformation with episomal overexpression plasmids as described previously (Allers et al., 2010). Plasmids for protein expression are shown in **Table 2**, and were generated by PCR using oligonucleotides shown in **Table 3**. Template DNA for the PCRs was isolated from genomic DNA.

UV IRRADIATION ASSAYS

UV irradiation assays were carried out as described previously (Delmas et al., 2009).

MITOMYCIN C ASSAYS

Mitomycin C (MMC) assays were carried out as described previously (Lestini et al., 2010).

PROTEIN OVEREXPRESSION AND PURIFICATION

Protein overexpression was carried out as described previously (Allers et al., 2010) with the following amendments: cultures were incubated at 45°C overnight to an OD_{650} of 0.5, when protein expression was induced by adding 3 mM Trp to the culture followed by incubation at 45°C, with shaking for a further 1 h until $OD_{650} \approx 0.7$.

PROTEIN PRECIPITATION

Deoxycholate was added to 0.015%, vortexed, and incubated for 10 min at room temperature. Trichloroacetic acid was added to 7.2% and incubated at room temperature for 5 min. Samples were centrifuged at 14,000 × *g* at room temperature for 8 min. Supernatant was removed and precipitated protein resuspended in 15 µl resuspension buffer (330 mM Tris-HCl pH 7.2, 2.6% SDS, 17 mM NaOH, 5% glycerol, 0.25 mg/ml bromophenol blue). Samples were heated for 10 min at 94°C and cooled on ice before loading onto an SDS-PAGE gel.

MASS SPECTROMETRY

Mass spectrometry of excised protein bands was carried out as described previously (Allers et al., 2010). Details of protein identification are given in the **Table A1** in Appendix.

RESULTS

RPA3 BUT NOT RPA1 FUNCTIONS IN DNA REPAIR

In eukaryotes, specifically *Saccharomyces cerevisiae*, all three RPA subunits have been shown to be essential for cell survival (Brill and Stillman, 1991). Work by Skowyra and MacNeill (2012) has shown that *H. volcanii rpa2* is essential, which is in agreement with our fruitless attempts to delete *rpa2* (data not shown). To examine if the other *rpa* genes of *H. volcanii* are also essential,

Table 2 | Plasmids.

Plasmid	Relevant properties	Source or reference
pBluescript II SK+	Standard cloning vector	Stratagene
pTA131	Integrative vector based on pBluescript II, with *pyr*E2 marker	Allers et al. (2004)
pTA409	Shuttle vector containing ampicillin, *pyrE2* and *hdrB* markers, and pHV1/4 replication origin	Delmas et al. (2009)
pTA884	pBluescript II with *H. volcanii* 5,038-bp *Eco*RI/*Not*I genomic fragment containing *rpa3* operon	This study
pTA898	pBluescript II with *H. volcanii* 7,335-bp *Eco*RI/*Not*I genomic fragment containing *rpa2*	This study
pTA937	pBluescript II with *H. volcanii* 8,565-bp *Bsp*EI genomic fragment containing *rpa1* operon	This study
pTA963	Overexpression vector with 6xHis-tag, *pyr*E2 and *hdrB* markers, and pHV2 origin	Allers et al. (2010)
pTA1141	pTA131 containing *rpa1* deletion construct inserted at *Kpn*I and *Xba*I sites, contains an internal *Nde*I site	This study
pTA1142	pTA131 containing *rpa3* deletion construct inserted at *Eco*RI and *Kpn*I sites, contains an internal *Nde*I site	This study
pTA1166	*rpa1* deletion construct pTA1141 with *trpA* marker, amplified from pTA298 introducing *Nde*I restriction sites to insert at internal *Nde*I restriction site in pTA1141	This study
pTA1170	Deletion construct of *rpap1* containing *trpA* marker from pTA298 inserted at *Eco*RI and *Kpn*I sites in pTA131	This study
pTA1174	*rpa3* deletion construct containing *trpA* marker from pTA1166 inserted at *Nde*I restriction site	This study
pTA1180	pTA131 with *cdc48d* deletion construct	Allers et al. (2010)
pTA1189	pTA131 with *rpa1* operon deletion construct inserted at restriction sites *Xba*I and *Eco*RI with an internal *Nde*I site	This study
pTA1196	*rpa3* operon deletion construct, using *Nde*I/*Eco*RI downstream fragment from pTA1282 (*rpap3* deletion construct) inserted at *Nde*I/*Eco*RI sites in pTA1142 (*rpa3* deletion construct), to replace the downstream fragment of the *rpa3* deletion construct	This study
pTA1207	Deletion construct of *rpa3* operon pTA1196 with insertion of the *trpA* marker from pTA1166 at internal *Nde*I site	This study
pTA1217	*RPAP1* deletion construct pTA1170 with upstream and *trpA* fragment replaced with the upstream fragment amplified from pTA937 by PCR, to introduce compatible *Sph*I sites, generating non-*trpA*-marked deletion construct	This study
pTA1218	pTA963 with *rpa3* inserted downstream of His-tag. *Ase*I inserted after *rpa3* stop codon to allow insertion of His-tagged *rpa3* upstream of His-tagged *rpap3* (*Ase*I is *Nde*I compatible)	This study
pTA1222	pTA963 with *rpa1* N-terminally His-tagged, has an *Ase*I site downstream of *rpa1* to allow insertion of *rpap1* (*Nde*I compatible)	This study
pTA1223	pTA963 overexpression vector with *rpap1* N-terminally His-tagged inserted at *Psi*I and *Bam*HI sites. *rpap1* was amplified by PCR from pTA937 introducing *Bsp*HI and *Bam*HI sites	This study
pTA1224	pTA963 with *rpap3* N-terminally His-tagged inserted at *Psi*I and *Eco*RI sites. *RPAP3* was amplified by PCR from pTA884 introducing *Bsp*HI and *Eco*RI sites	This study
pTA1240	Gene replacement construct with insertion of 896 bp *Hsa-cdc48d* gene (amplified from *H. salinarum* DNA) between upstream and downstream flanking regions of *H. volcanii cdc48d* deletion construct pTA1180	This study
pTA1265	pTA409 with insertion of *rpa1* operon from pTA937 at *Eco*RV site	This study
pTA1280	pTA1218 with *rpap3* amplified from pTA884 by PCR and inserted at *Bst*EII and *Eco*RI sites after the N-terminally His-tagged *rpa3*, maintaining reading frame	This study
pTA1281	pTA1224 with *rpa3* amplified from pTA884 by PCR and inserted upstream of N-terminally His-tagged *rpap3* at *Nde*I site	This study
pTA1282	*rpap3* deletion construct with upstream and downstream regions amplified from genomic clone pTA884, introducing external *Kpn*I and *Eco*RI sites, used to ligate into pTA131, and internal *Nde*I site	This study
pTA1284	*rpap3* deletion construct pTA1282 with *trpA* marker digested from pTA1166 using *Nde*I and inserted at *Nde*I site in pTA1282, generating *trpA*-marked *rpap3* deletion construct	This study
pTA1288	pBluescript II with *H. volcanii* 3,299-bp *Sal*I/*Bsp*HI genomic fragment containing *cdc48d* gene	This study
pTA1294	pTA131 with 2,247 bp *Hvo-cdc48d-Ct* gene replacement construct amplified from pTA1288: 1,797 bp *Eco*RI-*Nhe*I fragment with C-terminally truncated *cdc48d* plus upstream region, ligated to 485 bp *Nhe*I-*Kpn*I fragment with downstream region of *cdc48d*, inserted at *Eco*RI and *Kpn*I sites	This study
pTA1326	pTA1222 with *rpap1*, amplified from pTA937 introducing *Bst*EII and *Bam*HI sites, and inserted downstream of His-tagged *rpa1* at *Bst*EII and *Bam*HI sites	This study
pTA1327	pTA1223 with *rpa1* inserted upstream of His-tagged *rpap1* at *Nde*I site. *rpa1* was amplified from pTA937 introducing *Nde*I and *Ase*I (*Nde*I compatible) sites	This study

Table 3 | Oligonucleotides.

Oligonucleotide	Sequence (5′–3′)	Relevant properties	Use (plasmid generated)
Rpa1CF DS	GTTCGAGGTACCGTTCGGGGAGC	Δ*rpa1* external downstream primer, *Kpn*I site	pTA1141
Rpa1CR DS	AGGTGCGCATATGAGCGCCTCGC	Δ*rpa1* internal downstream primer, *Nde*I site	pTA1141
Rpa1 CR US	TACTACGTCTAGACGGACCTGTTCG	Δ*rpa1* external upstream primer, *Xba*I site	pTA1141
Rpa1 CF US	GGTCGAGTTCCATATGGTCGGGATTCGCC	Δ*rpa1* internal upstream primer, *Nde*I site	pTA1141
Rpa3KpnI F	GCCGGTGGTACCACAGCCTC	Δ*rpa3* external upstream primer, *Kpn*I site	pTA1142
Rpa3NdeIR	GCAAATCAGTCATATGCTACCTCGCC	Δ*rpa3* internal upstream primer, *Nde*I site	pTA1142
Rpa3EcoRIR	GACGGTGGAATTCGGCCGTCG	Δ*rpa3* external downstream primer, *Eco*RI site	pTA1142
Rpa3NdeI FC	GCGAGGTCGATGCATATGAGTTCCAACG	Δ*rpa3* internal downstream primer, *Nde*I site	pTA1142
trpANdeIF	CTCTGCACATATGTCGCTCGAAGACGC	*trpA* forward primer containing *Nde*I site	pTA1166
trpANdeIR	TGCATGCCATATGCGTTATGTGCG	*trpA* reverse primer containing *Nde*I	pTA1166
RPAP11kpnIus	CCGCGAGTGGTACCGCAAGCCCG	Δ*rpap1* external upstream primer, *Kpn*I site	pTA1170
RPAP11nsiIus	CGACGACCGGCGATGCATTCATGCGCGC	Δ*rpap1* internal upstream primer, *Nsi*I site	pTA1170
RPAP11sphIds	GCTGAAGGGCATGCGAGGCCGTGC	Δ*rpap1* internal downstream primer, *Sph*I site	pTA1170
RPAP11ecoRIds	CGGCGAGAGAATTCCCTGCCCGGG	Δ*rpap1* external downstream primer, *Eco*RI site	pTA1170
PEecorI F	GCCCGAATTCCGTCTGATTG	Δ*rpa1 operon* external downstream primer, *Eco*RI site	pTA1189
Rpa1CR US	TACTACGTCTAGACGGACCTGTTCG	Δ*rpa1 operon* external upstream primer, *Xba*I site	pTA1189
RPEndeI R DS	CTACCGGAACATATGACTCGGGTCG	Δ*rpa1 operon* internal downstream primer, *Nde*I site	pTA1189
Rpa1ndeI F2	GTTGGACCCATATGTCGAACGACG	Δ*rpa1 operon* internal upstream primer, *Nde*I site	pTA1189
RPAP11SphI US	GCGATTTCCCGCATGCCGACGACCG	Δ*rpap1* internal upstream primer, *Sph*I site	pTA1217
RPAP11 kpnI us	CCGCGAGTGGTACCGCAAGCCCG	Δ*rpap1* external upstream primer, *Kpn*I site	pTA1217
Rpa3BspHI F	AGGTAGATCATGACTGATTTGC	*rpa3* forward primer, *Bsp*HI site	pTA1218
Rpa3 RAseI	CGAGTGGGGAATTCGTTGGAATTAATTTACATC	*rpa3* reverse primer, *Ase*I site	pTA1218
Rpa1F *Nco*I	CCCGACTCCATGGAACTCGACC	*rpa1* forward primer, *Nco*I site	pTA1222
Rpa1AseI/*Eco*RI	CGGCGGCGAATTCGCGGTAGGCGATTAATCGCGTGC pTA1327	*rpa1* reverse primer, *Ase*I and *Eco*RI sites	pTA1222
RPAP1F *Bsp*HI	GGTGCGCTCATGAGCGCCTCG	*rpap1* forward primer, *Bsp*HI site	pTA1223
RPAP1BamHI	CGTTCGGGGGATCCGCGCCTGC pTA1326	*rpap1* reverse primer, *Bam*HI site	pTA1223
RPAP3BspHI F	GTCGATGTTCATGAGTTCCAACG	*rpap3* forward primer, *Bsp*HI site	pTA1224
RPAP3EcoRI R	CGGTCGGAATTCAGGCCGAC pTA1280	*rpap3* reverse primer, *Eco*RI site	pTA1224
HsaCdc48F	GTTCTTGGCATATGACCGAGGCTCTC	Forward primer for *Hsa-cdc48d*, *Nde*I site	pTA1240, Probe **Figure 5B**
HsaCdc48R	CTGACAGATCTCGCAGTCACAGC	Reverse primer for *Hsa-cdc48d*, *Bg*III site	pTA1240, Probe **Figure 5B**
Rpa3BstEII	GATGCGCGGTGACCTCGTGG	*rpap3* forward primer, native *Bst*EII site	pTA1280
Rpa3NdeI	CGAGGTAGCATATGACTGATTTGCG	*rpa3* forward primer, *Nde*I site	pTA1281
RPAP3 gitF	CTCCCAATGGGTACCAAGGTGGAGGC	Δ*rpap3* internal upstream primer, *Nde*I site	pTA1282
RPAP3 gitR	TCGTTGGACATATGTTACATCGACCTCGC	Δ*rpap3* external upstream primer, *Kpn*I site	pTA1282
RPAP3 F DS	CTCGCTGAATTCGGTGGGTGC	Δ*rpap3* external downstream primer, *Eco*RI site	pTA1282
RPAP3 R DS	CTGAGCGCATATGCGGGCGTCTCG	Δ*rpap3* internal downstream primer, *Nde*I site	pTA1282
cdc48dUF	ACGGGTACCCACGTTGCTGG	*Hvo-cdc48d* external upstream primer, *Kpn*I site	pTA1294
cdc48dDR	GCCGAATTCGAGCCGAGGTGG	*Hvo-cdc48d* external downstream primer, *Eco*RI site	pTA1294
cdc48d-CtrR	CGGCGCGCTAGCCGGACCGGTTACGC	Internal reverse primer to generate C-terminally truncated *Hvo-cdc48d*, *Nhe*I site at *cdc48d* stop codon	pTA1294
cdc48d-CtrF	CTGTGGTGCTAGCCGTCGTCCGACCCCG	Internal forward primer to generate C-terminal truncated *Hvo-cdc48d*, *Nhe*I site at *cdc48d* stop codon	pTA1294
cdc48dSeqF	GGAAAAAGGGGCAGATGGTG	Forward primer to downstream flanking region of *Hvo-cdc48d*	PCR **Figure 5C**
cdc48dHvSeqR	CGACGACATCTCGCTGATTCG	Reverse primer to *Hvo-cdc48d* gene	PCR **Figure 5C**
cdc48dHsalSeqR	GGTCAACACGCTGCTGAAGTCC	Reverse primer to *Hsa-cdc48d* gene	PCR **Figure 5C**
Rpa1BstEII	CCGGCACGGTGACCGCCATCC	*rpap1* forward primer, native *Bst*EII site	pTA1326
Rpa1NdeI	CCCGACCATATGGAACTCGACC	*rpa1* forward primer, *Nde*I site	pTA1327

genomic deletions of *rpa1* and *rpa3* were generated using the counter selective pop-in/pop-out method (Allers et al., 2004). To generate the deletion constructs by PCR, *rpa1* and *rpa3* operons were first isolated from wild-type (WT) *H. volcanii* using native *Bsp*EI and *Eco*RI/*Not*I restriction sites, respectively, to generate genomic libraries. These were then screened for the presence of the *rpa1* and *rpa3* operons, individually, using colony hybridization. The isolated plasmids, pTA937 (*rpa1* operon) and pTA884 (*rpa3* operon) were confirmed by DNA sequencing. Deletion constructs for *rpa1* and *rpa3* were designed to avoid polar effects on the expression of the downstream *rpap* genes by maintaining the reading frame. Genomic deletions of both *rpa1* and *rpa3* (*trpA*-marked) were successful, generating strains H1217 and H1244, respectively (**Figures 2** and **3**, respectively). The ability to delete both *rpa1* and *rpa3* with relative ease, but not *rpa2*, indicates that the cellular requirement for each RPA is not equal, making it unlikely that they function collectively.

Both eukaryotic and bacterial SSB are involved in DNA repair. To examine if *H. volcanii* RPA1 and RPA3 function in DNA repair, the effects of DNA damage on cell survival of H1217 and H1244 were examined. UV irradiation results in the formation of cyclobutane pyrimidine dimers and 6-4 pyrimidine-pyrimidone dimer photoproducts, as well as ssDNA nicks that indirectly generate double-stranded DNA breaks (DSBs). The latter require repair by homologous recombination (HR) or single-strand DNA annealing (Fousteri and Mullenders, 2008; Rouillon and White, 2011). MMC is a chemotherapeutic agent that reacts with DNA generating covalent interstrand cross-links, requiring removal by nucleotide excision repair (NER) and HR (Tomasz et al., 1987). The Δ*rpa1* mutant H1217 was no more sensitive than the WT to UV and MMC-induced DNA damage, however the Δ*rpa3* mutant H1244 exhibited moderate sensitivity to both UV and MMC-induced DNA damage (**Figure 4**).

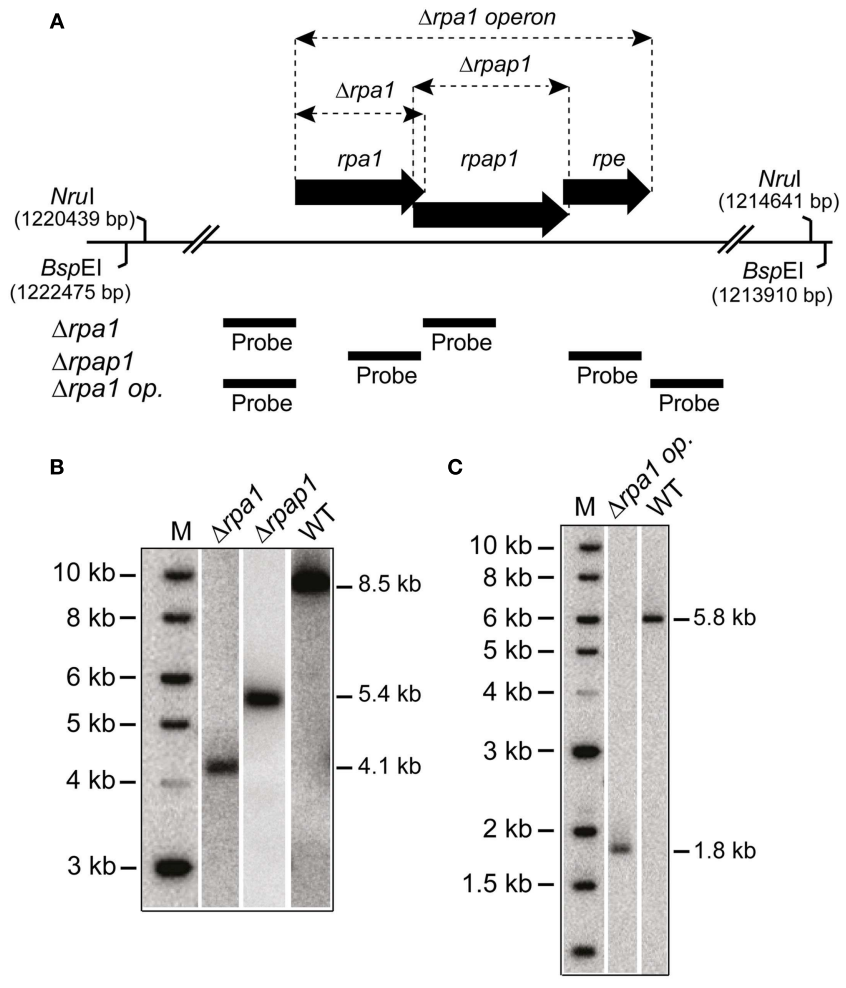


FIGURE 2 | (A) Map of *rpa1* operon indicating location of Δ*rpa1*, Δ*rpap1*, and Δ*rpa1* operon deletions, as well as the *Bsp*EI and *Nru*I sites, and probes used to verify the deletions. (**B**) Southern blot of genomic DNA cut with *Bsp*EI and probed with flanking regions of *rpa1* and *rpap1*, as shown in **(A)**, to indicate deletion of *rpa1* and *rpap1*, respectively. **(C)** Southern blot of genomic DNA cut with *Nru*I and probed with flanking regions of *rpa1* operon (*rpa1 op.*), as shown in **(A)**, to indicate deletion of the *rpa1* operon.

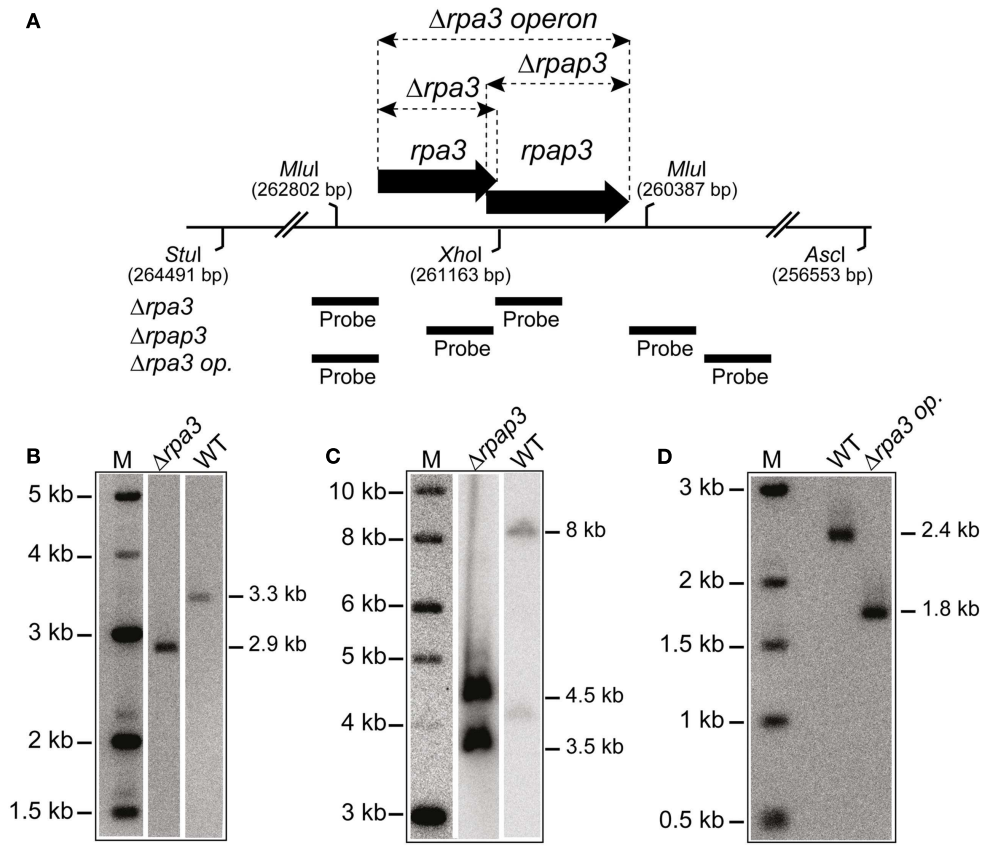


FIGURE 3 | (A) Map of *rpa3* operon indicating location of Δ*rpa3*, Δ*rpap3*, and Δ*rpa3* operon deletions, as well as the *Asc*I, *Mlu*I, *Stu*I, and *Xho*I sites and probes used to verify the deletions. **(B)** Southern blot of genomic DNA cut with *Stu*I and *Xho*I, and probed with flanking regions of *rpa3*, as shown in **Figure 2A**, to indicate deletion of *rpa3*. **(C)** Southern blot of genomic DNA cut with *Asc*I and *Stu*I, and probed with flanking regions of *rpap3*, as shown in **Figure 2A**, to indicate deletion of the *rpap3*. **(D)** Southern blot of genomic DNA cut with *Mlu*I and probed with flanking regions of *rpa3* operon (*rpa3 op.*), as shown in **Figure 2A**, to indicate deletion of the *rpa3* operon.

RPAP3 BUT NOT RPAP1 FUNCTIONS IN DNA REPAIR

Analysis of predicted protein domains indicates that RPA1 and RPA3 both possess zinc finger domains, and that RPA1 has three OB-folds compared to the single OB-fold present in RPA3 (**Figure 1**). Both COG3390 RPAPs RPAP1 and RPAP3 possess a single OB-fold suggesting a possible role in DNA binding. The RPA1 phosphoesterase (RPE) has a calcineurin-like phosphoesterase domain, and was not investigated individually. However, our results and those of Skowyra and MacNeill (2012) show that *rpe* is a non-essential gene.

To study the roles of RPAP1 and RPAP3 in DNA repair, Δ*rpap1* (H1216) and Δ*rpap3* (H1410) mutants were generated, both using *trpA*-marked deletion constructs (**Figures 2** and **3**, respectively). As with Δ*rpa1* strain H1217, the Δ*rpap1* mutant H1216 showed no increased sensitivity to UV irradiation or to MMC-induced DNA damage. However, the Δ*rpap3* deletion mutant H1410 was hypersensitive to both types of DNA damage, and the level of sensitivity was similar to that exhibited by the Δ*rpa3* mutant H1244.

We examined whether the absence of both RPA and RPAP results in a synergistic deficiency in DNA repair. Genomic deletions of the *rpa1* and *rpa3* operons were generated in strains H1246 and H1260, respectively, with only the latter being a *trpA*-marked deletion (**Figures 2** and **3**, respectively); deletions of the *rpa1* and *rpa3* operons have previously been reported by Skowyra and MacNeill (2012). The Δ*rpa1* operon mutant showed no increased sensitivity to UV irradiation or to MMC-induced DNA damage. However the Δ*rpa3* operon deletion mutant was hypersensitive to both types of DNA damage, and the level of sensitivity was similar to that exhibited by the single Δ*rpa3* and Δ*rpap3* mutants H1244 and H1410, respectively (**Figure 4**). This result suggests that RPA3 and RPAP3 function in the same pathway(s) of DNA repair.

REDUNDANCY BETWEEN RPA1 AND RPA3 OPERONS

In order to test for redundancy between the two RPAs, an attempt was made to generate a double Δ*rpa1* operon Δ*rpa3* operon deletion. This involved constructing the strain H1282, which contained the pop-in of a *trpA*-marked Δ*rpa3* operon construct (pTA1207) in an unmarked Δ*rpa1* operon background (H1246). An episomal plasmid (pTA1265), marked with *pyr*E2 and providing in *trans* expression of the *rpa1* operon was used for complementation during the pop-out step (note that this episomal plasmid is lost during counter-selection with 5-FOA). Neither of the two pop-outs generated from this strain (H1390) yielded the desired Δ*rpa1* operon

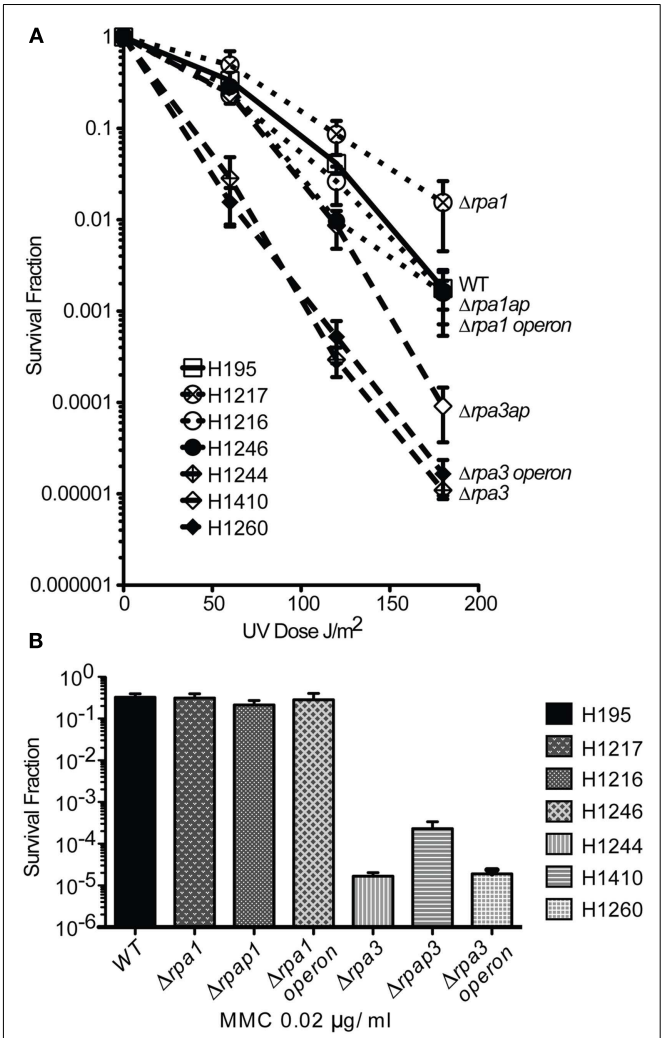


FIGURE 4 | (A) Wild-type (WT), *rpa1*, *rpap1*, *rpa1* operon, *rpa3*, *rpap3*, *rpa3* operon (H195, H1217, H1216, H1246, H1244, H1410, and H1260, respectively), were plated out and exposed to ultraviolet (UV) radiation. The average and standard error (SE) of three experiments are shown. **(B)** The strains shown in **(A)** were plated out and exposed to mitomycin C (MMC). The average and SE of three experiments are shown.

Δ*rpa3* operon mutant (see **Figure A1** in Appendix). This indicates that the cell requires either RPA1 or RPA3 (and/or their respective RPAPs) for survival.

Next we attempted to generate Δ*rpa1* Δ*rpap3* and Δ*rpa3* Δ*rpap1* deletion mutants. This would test whether the RPAPs can complement each other, or whether they are instead specific for their respective RPAs. The *trpA*-marked Δ*rpap3* construct (pTA1284) was used in an unmarked Δ*rpa1* background (H1280), and the *trpA*-marked Δ*rpa3* construct (pTA1207) was used in an unmarked Δ*rpap1* background (H1281). In both cases two pop-outs were generated but none proved to be the desired deletions (see **Figure 1** in Appendix). This suggests that the putative RPA:RPAP complex is dependent upon specific RPA:RPAP interactions for functionality.

CONSTRUCTION OF PROTEIN OVEREXPRESSION STRAIN WITH C-TERMINAL TRUNCATION OF CDC48D

In a previous publication (Allers et al., 2010), we constructed a strain of *H. volcanii* where the histidine-rich *pitA* gene is replaced by the ortholog from *N. pharaonis*. The latter protein lacks the histidine-rich linker region found in *H. volcanii* PitA and does not co-purify with His-tagged recombinant proteins. The absence of Hvo-PitA revealed an additional co-purifying protein, which we identified as Cdc48d (HVO_1907) and features a histidine-rich C-terminus (**Figure 5A**). We were unable to delete *cdc48d*, indicating that this gene is essential (Allers et al., 2010). The presence of this contaminating protein was problematic for purification of His-tagged RPA1 and RPAP1, due to similar molecular weights (Cdc48d, 53 kDa; RPA1, 46 kDa; RPAP1, 65 kDa).

All orthologs of Cdc48d from haloarchaea feature a histidine-rich C-terminus, however Cdc48d from *Haloarcula marismortui* and *H. salinarum* have only three and four histidines, respectively, compared to six in *H. volcanii* (**Figure 5A**). Therefore, we replaced the *H. volcanii cdc48d* gene in H1209 (Allers et al., 2010) with orthologous genes from *H. marismortui* and *H. salinarum*, generating *H. volcanii* strains H1405 and H1333, respectively. Unfortunately these strains grew poorly and were not suitable for recombinant protein overexpression. Instead, we generated a truncated allele of *H. volcanii cdc48d*, encoding a Cdc48d protein lacking the histidine-rich C-terminus (Cdc48d-Ct; **Figure 5A**). The *cdc48d-Ct* allele was used to replace the *H. salinarum cdc48d* gene in *H. volcanii* H1333, generating H1424 (**Figures 5B,C**). This strain exhibits normal cell growth and the Cdc48d-Ct protein no longer co-purifies with His-tagged recombinant proteins (**Figure 5D**). A number of minor histidine-rich contaminants are now apparent, which have been identified by mass spectrometry.

DIRECT RPAP INTERACTION WITH RESPECTIVE RPA

The genetic analysis of *rpa1* and *rpa3* and their respective *rpap* genes indicates not only that RPA3 and RPAP3 function in the same DNA repair pathway(s), but also that they function together as a specific RPA:RPAP complex. To establish whether this is achieved via a direct RPA:RPAP interaction, affinity pull-downs were employed (Allers et al., 2010). The *rpa1* and *rpa3* operons were cloned under control of the tryptophanase promoter in plasmid pTA963, where either the RPA or the RPAP was tagged with a hexahistidine tag.

Histidine-tagged RPA1 and RPA3 pulled down their respective RPAPs, and histidine-tagged RPAP1 and RPAP3 pulled down their respective RPAs (**Figure 6**). However, histidine-tagged RPA1 did not pull down RPAP3, and vice versa. This confirms that the RPAs interact specifically with their respective RPAPs, supporting our conclusions based on the failure to generate Δ*rpa1* Δ*rpap3* and Δ*rpa3* Δ*rpap1* deletion mutants. Neither RPA1 nor RPA3 pulled down RPA2, and histidine-tagged RPA2 did not pull down RPA1 or 3, or either of the RPAPs (data not shown). This supports the suggestion that the three RPAs of *H. volcanii* do not form a heterotrimer as observed in eukaryotes and *P. furiosus*, but instead form three separate ssDNA-binding factors.

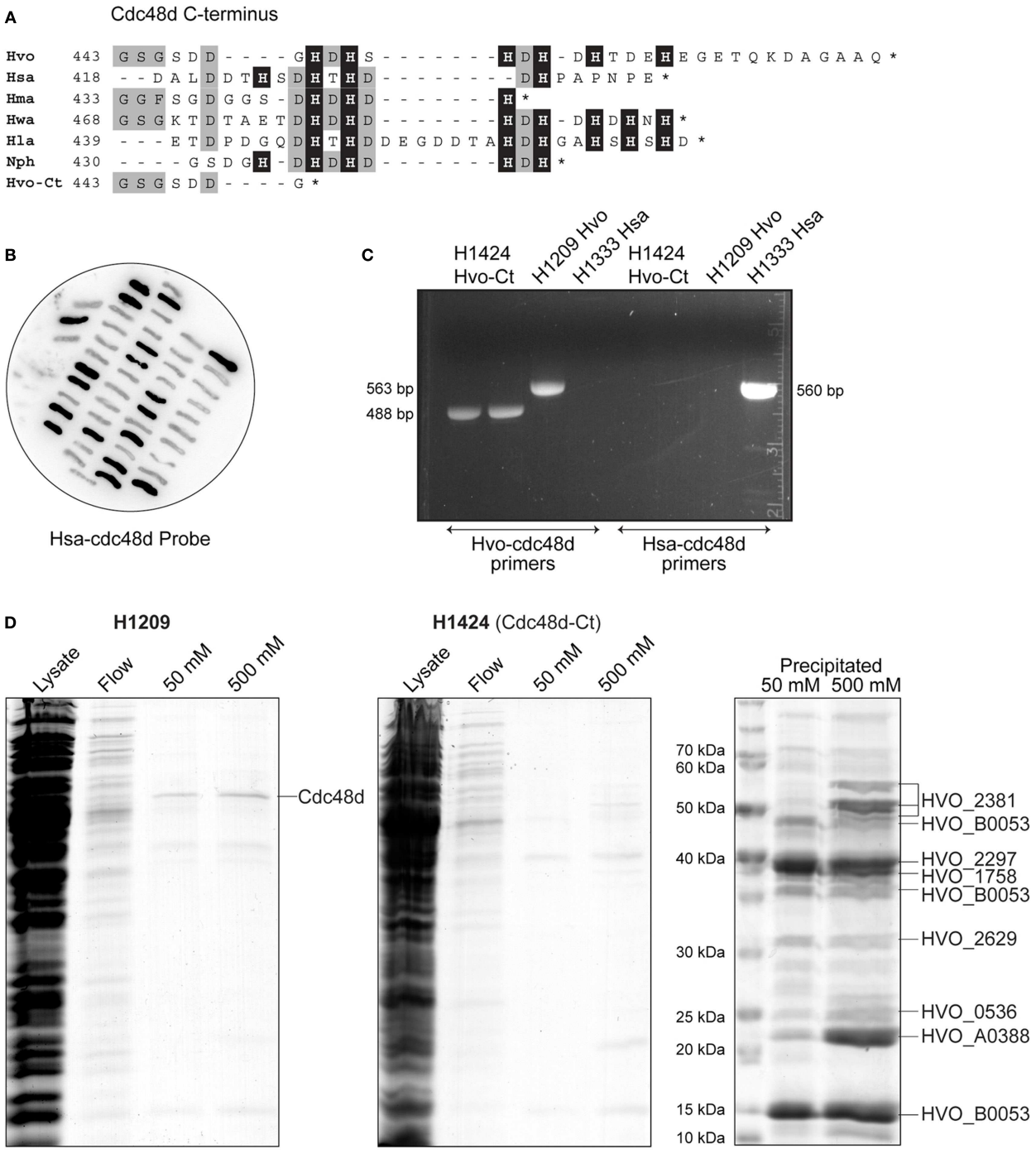


FIGURE 5 | (A) Protein sequence alignment of C-terminus of Cdc48d from selected species of haloarchaea (Hvo, *H. volcanii*; Hsa, *Halobacterium salinarum*; Hma, *Haloarcula marismortui*; Hwa, *Haloquadratum walsbyi*; Hla, *Halorubrum lacusprofundi*; Nph, *N. pharaonis*; Hvo-Ct, *H. volcanii* C-terminal truncation Cdc48d-Ct). Histidine residues are indicated by a black background. **(B)** Colony hybridization of 5-FOA-resistant clones of *H. volcanii* H1333, after pop-in/pop-out gene replacement with pTA1294. *H. salinarum cdc48d* sequences (*Hsa-cdc48d*) were used as a probe, clones failing to hybridize therefore carry the truncated *H. volcanii cdc48d-Ct* allele present in pTA1294. **(C)** Verification of truncated *cdc48d-Ct* allele in H1424 by PCR (488 bp product), with primers specific to either *H. volcanii* or *H. salinarum* genes. H1209 genomic DNA was used as a control for wild-type *H. volcanii cdc48d* (563 bp product), and H1333 was used as a control for *H. salinarum cdc48d* (560 bp product). **(D)** *H. volcanii* strains H1209 and H1424 containing empty vector pTA963 (Allers et al., 2010) were used in mock protein overexpression. Histidine-rich cellular proteins were purified from the soluble fraction (lysate) by affinity chromatography on a Ni^{2+} chelating column, samples were taken from the flow-through (flow) and bound proteins were eluted using 50 and 500 mM imidazole. Precipitation using trichloroacetic acid and deoxycholate was used to enhance visualization and identification of the eluted proteins by mass spectrometry. Cdc48d (HVO_1907) eluted from cell extracts of H1209 but not from H1424 (*cdc48d-Ct*).

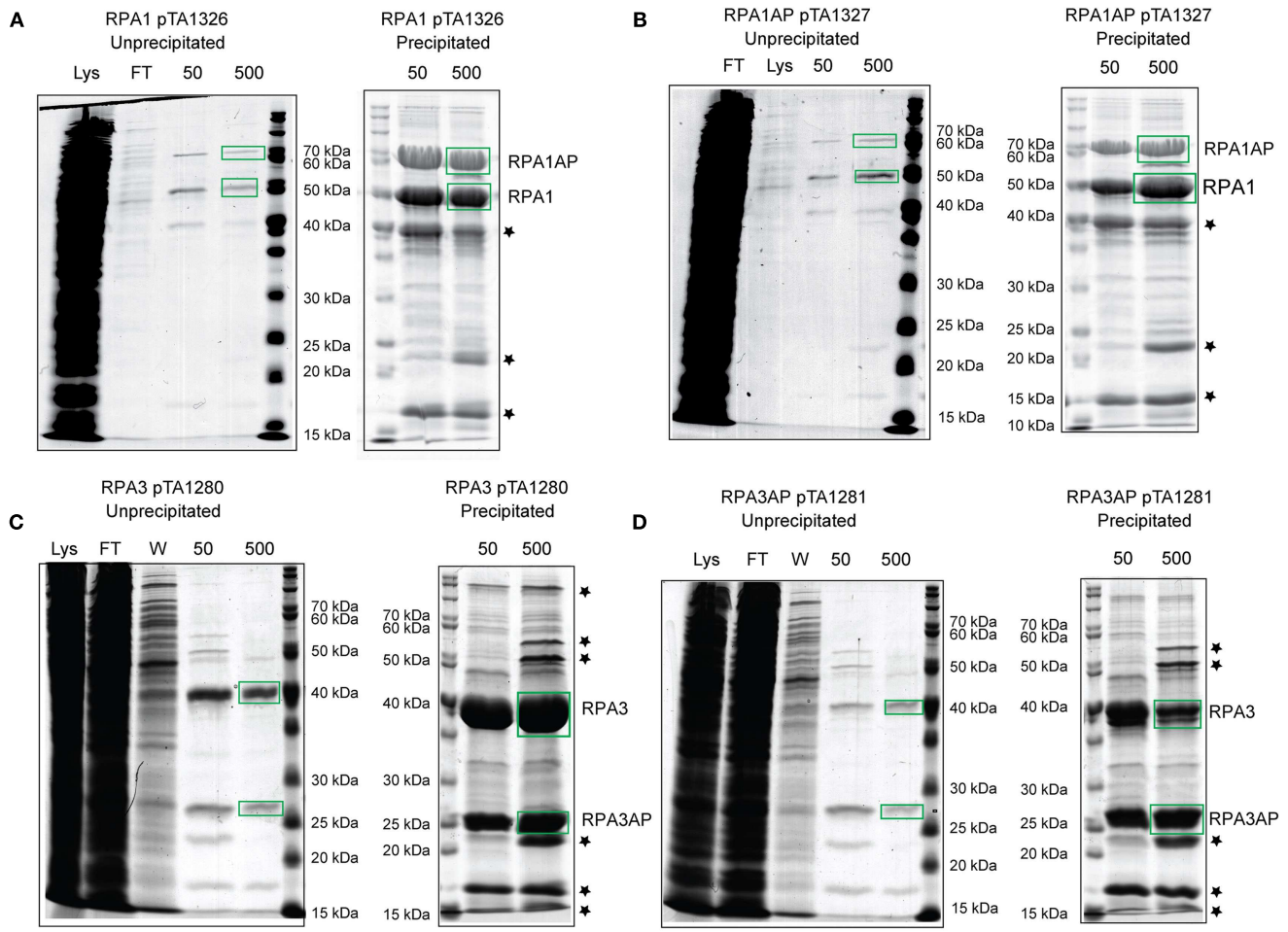


FIGURE 6 | *H. volcanii* strains containing (A) histidine-tagged RPA1 (H1482), (B) histidine-tagged RPAP1 (H1483), (C) histidine-tagged RPA3 (H1480), and (D) histidine-tagged RPAP3 (H1481) were used in protein overexpression. The histidine-tagged recombinant proteins were purified from the soluble fraction (lysate) by affinity chromatography on a Ni^{2+} chelating column, samples were taken from the flow-through (FT) and bound proteins were eluted using 50 and 500 mM imidazole. Precipitation using trichloroacetic acid and deoxycholate was used to enhance visualization and identification of the eluted proteins by mass spectrometry. Stars denote contaminants shown in **Figure 5D**.

DISCUSSION

There is a unifying theme in archaea of a great variety in the number and type of proteins involved in DNA replication, and repair, whose counterparts in eukaryotes are much more uniform. This has been shown to be the case for RPA, where eukaryotes possess three subunits that all form unified clades in a phylogenetic analysis, but in archaea the number and structure of subunits varies widely. Some euryarchaea possess differing numbers of RPA subunits and some possess differing numbers of RPAs and RPAPs. Crenarchaea also possess varying numbers of SSB, however in both euryarchaea and crenarchaea none of the RPAs, RPAPs, or SSBs fall into unified clades. Again this is seen in the case of MCM, where eukaryotes possess six MCM subunits that each form unified clades, but in archaea there is a vast range in the number of MCM subunits, differing between individual species, and none of which fall into uniform clades (Chia et al., 2010). Characterizing the RPA-RPAP complexes of *H. volcanii* will shed light on how the RPAs and RPAPs function together in binding and stabilizing ssDNA. This in turn will provide insight for other RPAs and RPAPs in archaea, but also offer reasoning behind the driving force of such non-uniform evolution of archaea.

The genetic and biochemical analysis presented here indicates the three RPAs of *H. volcanii* do not form a heterotrimeric complex as in *P. furiosus* and eukaryotes. Instead, RPA1 and RPA3 form complexes with their respective RPAPs. Unlike *rpa2*, both *rpa1* and *rpa3* genomic deletions were generated with relative ease, showing that the latter are not essential for cell survival and supporting the hypothesis that the three RPAs do not form a heterotrimeric complex.

The ease at which the *rpa1*, *rpap1*, and *rpa1* operon deletion mutants were made, coupled with a lack DNA damage sensitivity, signifies the *rpa1* operon does not play a major role in DNA replication or repair. The moderate DNA damage sensitivity shown by the individual *rpa3*, *rpap3*, and the *rpa3* operon mutants indicates that the efficient repair of UV and MMC-induced DNA damage requires the products of the *rpa3* operon but not the

rpa1 operon. However, it proved impossible to generate a double Δ*rpa1* operon Δ*rpa3* operon deletion, showing that cellular growth requires either RPA1 or RPA3.

Both single Δ*rpa3* and Δ*rpap3* mutants showed a similar DNA damage sensitivity to each other, and to the *rpa3* operon mutant, providing genetic evidence that RPA3 and RPAP3 act in the same DNA repair pathway. Furthermore, we were unable to generate Δ*rpa1* Δ*rpap3* and Δ*rpa3* Δ*rpap1* deletion mutants, indicating that RPAP1 could not substitute for RPAP3 (and vice versa), and suggesting that the RPA3 interacts specifically with RPAP3 (and likewise for RPA1 and RPAP1). However it is unclear what role the associated proteins play, since the presence of an OB-fold does not necessarily indicate direct ssDNA binding. Instead, the RPAPs may provide a platform for protein:protein interactions. This is seen in eukaryotes, where the RPA 14 kDa subunit possesses a single OB-fold, this subunit is essential for formation of the RPA heterotrimer by facilitating protein:protein interactions (Fanning et al., 2006).

The co-purification of histidine-tagged RPA1 and RPA3 with their respective untagged RPAPs (and vice versa) supports our hypothesis that *H. volcanii* RPA1 and RPA3 form complexes with their respective RPAPs. This observation, and the differing outcomes of *rpa1*, *rpa2*, and *rpa3* deletions, indicates that the three RPAs of *H. volcanii* do not function as a heterotrimer. Similar results have been obtained in *M. acetivorans*, where the three RPAs are able to bind ssDNA individually, in addition to stimulating primer extension by *M. acetivorans* DNA polymerase BI *in vitro*. (Robbins et al., 2004).

This study has shown genetically and biochemically that RPAPs interact with RPAs, and that this interaction is RPA-specific. This is the first report investigating the function of the archaeal COG3390 RPA-associated proteins (RPAPs), thus providing an important insight of the structure and function of *H. volcanii* single-strand DNA-binding proteins.

AUTHOR CONTRIBUTIONS

Amy Stroud and Thorsten Allers wrote the paper; Amy Stroud and Thorsten Allers designed the experiments; Amy Stroud and Thorsten Allers performed the microbiological and biochemical experiments; Susan Liddell carried out the mass spectrometry; Amy Stroud, Susan Liddell, and Thorsten Allers analyzed the data.

ACKNOWLEDGMENTS

We are grateful to the BBSRC for a PhD studentship awarded to Amy Stroud and the Royal Society for a University Research Fellowship awarded to Thorsten Allers. We thank Kayleigh Wardell for help with strain construction, and Ed Bolt, Geoff Briggs, and Karen Bunting for their advice.

REFERENCES

Alberts, B. M., and Frey, L. (1970). T4 bacteriophage gene 32: a structural protein in the replication and recombination of DNA. *Nature* 227, 1313–1318.

Allers, T., Barak, S., Liddell, S., Wardell, K., and Mevarech, M. (2010). Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76, 1759–1769.

Allers, T., Ngo, H. P., Mevarech, M., and Lloyd, R. G. (2004). Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the leuB and trpA genes. *Appl. Environ. Microbiol.* 70, 943–953.

Berthon, J., Cortez, D., and Forterre, P. (2008). Genomic context analysis in Archaea suggests previously unrecognized links between DNA replication and translation. *Genome Biol.* 9, R71.

Bochkarev, A., and Bochkareva, E. (2004). From RPA to BRCA2: lessons from single-stranded DNA binding by the OB-fold. *Curr. Opin. Struct. Biol.* 14, 36–42.

Brill, S. J., and Stillman, B. (1991). Replication factor-A from *Saccharomyces cerevisiae* is encoded by three essential genes coordinately expressed at S phase. *Genes Dev.* 5, 1589–1600.

Chia, N., Cann, I., and Olsen, G. J. (2010). Evolution of DNA replication protein complexes in eukaryotes and Archaea. *PLoS ONE* 5, e10866. doi:10.1371/journal.pone.0010866

Coverley, D., Kenny, M. K., Lane, D. P., and Wood, R. D. (1992). A role for the human single-stranded DNA binding protein HSSB/RPA in an early stage of nucleotide excision repair. *Nucleic Acids Res.* 20, 3873–3880.

Coverley, D., Kenny, M. K., Munn, M., Rupp, W. D., Lane, D. P., and Wood, R. D. (1991). Requirement for the replication protein SSB in human DNA excision repair. *Nature* 349, 538–541.

Cubeddu, L., and White, M. F. (2005). DNA damage detection by an archaeal single-stranded DNA-binding protein. *J. Mol. Biol.* 353, 507–516.

Delmas, S., Shunburne, L., Ngo, H. P., and Allers, T. (2009). Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination. *PLoS Genet.* 5, e1000552. doi:10.1371/journal.pgen.1000552

Fanning, E., Klimovich, V., and Nager, A. R. (2006). A dynamic model for replication protein A (RPA) function in DNA processing pathways. *Nucleic Acids Res.* 34, 4126–4137.

Fousteri, M., and Mullenders, L. H. (2008). Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* 18, 73–84.

Guy, C. P., Haldenby, S., Brindley, A., Walsh, D. A., Briggs, G. S., Warren, M. J., Allers, T., and Bolt, E. L. (2006). Interactions of RadB, a DNA repair protein in Archaea, with DNA and ATP. *J. Mol. Biol.* 358, 46–56.

Hartman, A. L., Norais, C., Badger, J. H., Delmas, S., Haldenby, S., Madupu, R., Robinson, J., Khouri, H., Ren, Q., Lowe, T. M., Maupin-Furlow, J., Pohlschroder, M., Daniels, C., Pfeiffer, F., Allers, T., and Eisen, J. A. (2010). The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLoS ONE* 5, e9605. doi:10.1371/journal.pone.0009605

Haseltine, C. A., and Kowalczykowski, S. C. (2002). A distinctive single-strand DNA-binding protein from the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* 43, 1505–1515.

Heyer, W. D., Rao, M. R., Erdile, L. F., Kelly, T. J., and Kolodner, R. D. (1990). An essential *Saccharomyces cerevisiae* single-stranded DNA binding protein is homologous to the large subunit of human RP-A. *EMBO J.* 9, 2321–2329.

Kenny, M. K., Lee, S. H., and Hurwitz, J. (1989). Multiple functions of human single-stranded-DNA binding protein in simian virus 40 DNA replication: single-strand stabilization and stimulation of DNA polymerases alpha and delta. *Proc. Natl. Acad. Sci. U.S.A.* 86, 9757–9761.

Kerr, I. D., Wadsworth, R. I., Cubeddu, L., Blankenfeldt, W., Naismith, J. H., and White, M. F. (2003). Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein. *EMBO J.* 22, 2561–2570.

Kim, C., Snyder, R. O., and Wold, M. S. (1992). Binding properties of replication protein A from human and yeast cells. *Mol. Cell. Biol.* 12, 3050–3059.

Komori, K., and Ishino, Y. (2001). Replication protein A in *Pyrococcus furiosus* is involved in homologous DNA recombination. *J. Biol. Chem.* 276, 25654–25660.

Lestini, R., Duan, Z., and Allers, T. (2010). The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in *Haloferax volcanii*. *DNA Repair (Amst.)* 9, 994–1002.

Lin, Y., Lin, L. J., Sriratana, P., Coleman, K., Ha, T., Spies, M., and Cann, I. K. (2008). Engineering of functional replication protein a homologs based on insights into the evolution of oligonucleotide/oligosaccharide-binding folds. *J. Bacteriol.* 190, 5766–5780.

Lu, D., and Keck, J. L. (2008). Structural basis of *Escherichia coli* single-stranded DNA-binding protein

stimulation of exonuclease I. *Proc. Natl. Acad. Sci. U.S.A.* 105, 9169–9174.

Lu, D., Windsor, M. A., Gellman, S. H., and Keck, J. L. (2009). Peptide inhibitors identify roles for SSB C-terminal residues in SSB/exonuclease I complex formation. *Biochemistry* 48, 6764–6771.

Moore, S. P., Erdile, L., Kelly, T., and Fishel, R. (1991). The human homologous pairing protein HPP-1 is specifically stimulated by the cognate single-stranded binding protein hRP-A. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9067–9071.

Mullakhanbhai, M. F., and Larsen, H. (1975). *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. *Arch. Microbiol.* 104, 207–214.

Robbins, J. B., McKinney, M. C., Guzman, C. E., Sriratana, B., Fitz-Gibbon, S., Ha, T., and Cann, I. K. (2005). The euryarchaeota, nature's medium for engineering of single-stranded DNA-binding proteins. *J. Biol. Chem.* 280, 15325–15339.

Robbins, J. B., Murphy, M. C., White, B. A., Mackie, R. I., Ha, T., and Cann, I. K. (2004). Functional analysis of multiple single-stranded DNA-binding proteins from *Methanosarcina acetivorans* and their effects on DNA synthesis by DNA polymerase BI. *J. Biol. Chem.* 279, 6315–6326.

Rolfsmeier, M. L., and Haseltine, C. A. (2010). The single-stranded DNA binding protein of *Sulfolobus solfataricus* acts in the presynaptic step of homologous recombination. *J. Mol. Biol.* 397, 31–45.

Rouillon, C., and White, M. F. (2011). The evolution and mechanisms of nucleotide excision repair proteins. *Res. Microbiol.* 162, 19–26.

Sambrook, J., and Russell, D. W. (2001). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Skowyra, A., and MacNeill, S. A. (2012). Identification of essential and non-essential single-stranded DNA-binding proteins in a model archaeal organism. *Nucleic Acids Res.* 40, 1077–1090.

Tomasz, M., Lipman, R., Chowdary, D., Pawlak, J., Verdine, G. L., and Nakanishi, K. (1987). Isolation and structure of a covalent cross-link adduct between mitomycin C and DNA. *Science* 235, 1204–1208.

Wobbe, C. R., Weissbach, L., Borowiec, J. A., Dean, F. B., Murakami, Y., Bullock, P., and Hurwitz, J. (1987). Replication of simian virus 40 origin-containing DNA in vitro with purified proteins. *Proc. Natl. Acad. Sci. U.S.A.* 84, 1834–1838.

Wold, M. S. (1997). Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* 66, 61–92.

Wold, M. S., Weinberg, D. H., Virshup, D. M., Li, J. J., and Kelly, T. J. (1989). Identification of cellular proteins required for simian virus 40 DNA replication. *J. Biol. Chem.* 264, 2801–2809.

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

Received: 15 March 2012; accepted: 31 May 2012; published online: 18 June 2012.

Citation: Stroud A, Liddell S and Allers T (2012) Genetic and biochemical identification of a novel single-stranded DNA-binding complex in Haloferax volcanii. Front. Microbio. ***3****:224. doi: 10.3389/fmicb.2012.00224*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

APPENDIX

Table A1 | Identification of proteins present in cellular soluble fraction after purification by affinity chromatography on a Ni^{2+} chelating column.

Prot accession	Protein name	HVO_#	Predicted MW	Observed MW	MASCOT score	Number of peptides	% coverage	Peptide sequences
gi[292655491]	RPAP1	1337	64,829	57,781	671	12	18	8
gi[292655492]	RPA1	1338	45,954	36,960	940	15	40	11
gi[292655491]	RPAP1	1337	64,829	62,780	578	10	21	9
gi[292655492]	RPA1	1338	45,954	45,686	808	15	40	12
gi[292654471]	RPAP3	0291	21,979	15,074	435	10	53	7
gi[292654471]	RPAP3	0291	21,979	16,217	484	10	59	8
gi[292654472]	RPA3	0292	34,562	28,262	796	13	43	10
gi[292654472]	RPA3	0292	34,562	31,741	735	13	48	11
gi[292656508]	Hypothetical protein	2381	52,319	55,300/50,200	240	5	11	5
gi[292493992]	Hypothetical protein	B0053	13,897	4,73,600	259	4	33	3
gi[292656425]	Deoxyhypusine synthase	2297	38,616	39,500	451	7	25	4
gi[292655899]	Thioredoxin reductase	1758	36,505	38,000	312	6	21	6
gi[292493992]	Hypothetical protein	B0053	13,897	36,300	208	6	40	4
gi[292656748]	Htr-like protein	2629	30,266	31,700	93	4	17	4
gi[292654704]	Ferritin	0536	19,892	24,900	66	1	6	1
gi[292653937]	Transcriptional regulator	A0388	20,201	22,300	586	15	54	8
gi[292493992]	Hypothetical protein	B0053	13,897	13,500	170	3	33	3

Prot Accession, the database entry, e.g., gi|292655491; predicted MW, predicted molecular weight (Da) of the protein sequence identified by MASCOT; observed MW, molecular weight estimated from migration on SDS-PAGE; MASCOT score, MASCOT score associated with protein identification; number of peptides, no. of peptides associated with protein identification by MASCOT; % coverage, % of the database sequence entry that is covered by the peptides matched to it in the MASCOT data. Peptide sequences, the number of distinct peptide sequences associated with the protein identified by MASCOT.

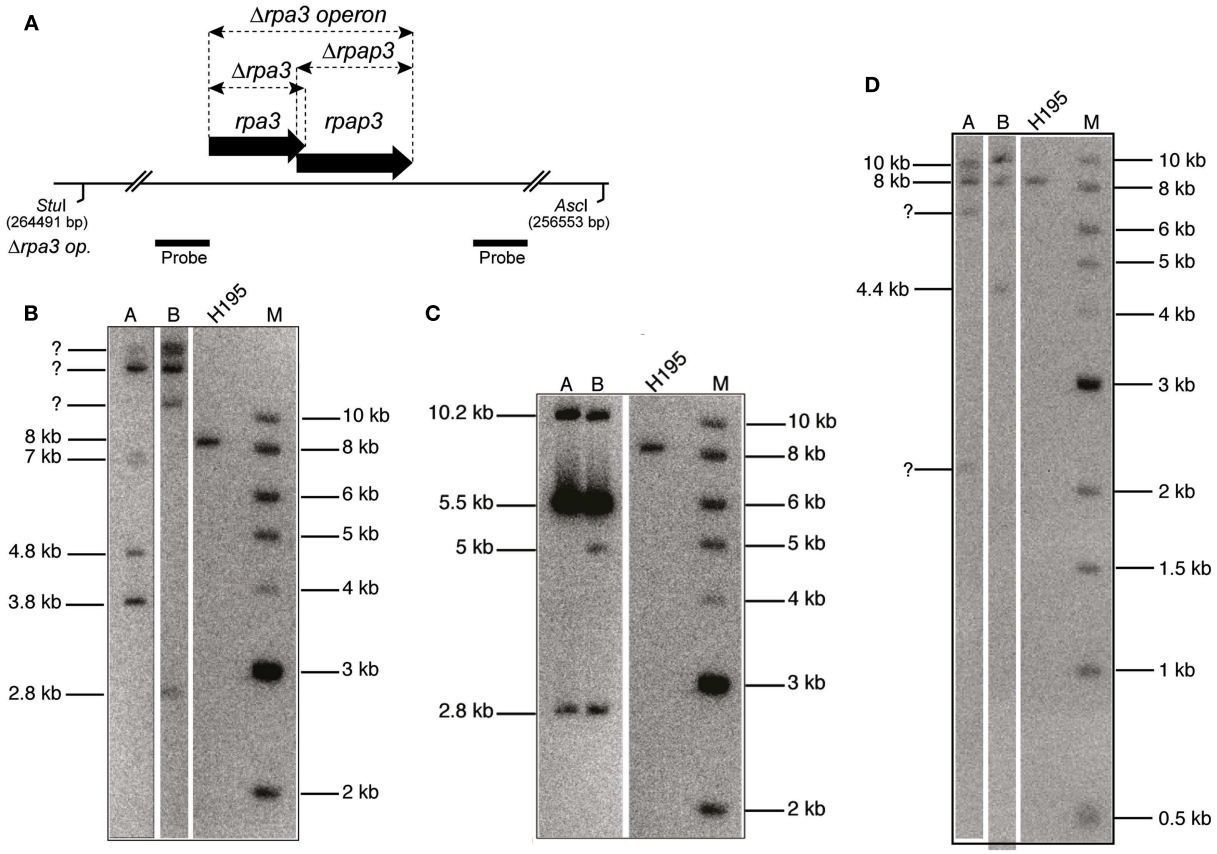


FIGURE A1 | (A) Map of *rpa3* operon indicating location of Δ*rpa3*, Δ*rpap3*, and Δ*rpa3* operon deletions, as well as the *Asc*I and *Stu*I sites and the probe used to verify the deletions. The size of this fragment in the wild-type (H195) is 8 kb. **(B)** Southern blot of genomic DNA cut with *Asc*I and *Stu*I, and probed with flanking regions of *rpa3* operon (*rpa3 op.*), as shown in **(A)**, indicates failure to generate Δ*rpa1* Δ*rpap3* mutant as bands of the expected size for deletion are not seen (3.4 and 4.4 kb). **(C)** Southern blot of genomic DNA cut with *Asc*I and *Stu*I, and probed with flanking regions of *rpa3 op.*, as shown in **(A)**, indicates failure to generate Δ*rpap1* Δ*rpa3* mutant as bands of the expected size for deletion are not seen (2.8 and 5 kb). **(D)** Southern blot of genomic DNA cut with *Asc*I and *Stu*I, and probed with flanking regions of *rpa3 op.*, as shown in **(A)** indicates failure to generate Δ*rpa1* operon Δ*rpa3* operon mutant as bands of the expected size for deletion are not seen (2.8 and 3.4 kb).

frontiers in
MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 02 July 2012
doi: 10.3389/fmicb.2012.00231

Diversity of antisense and other non-coding RNAs in archaea revealed by comparative small RNA sequencing in four *Pyrobaculum* species

David L. Bernick[1], Patrick P. Dennis[2], Lauren M. Lui[1] and Todd M. Lowe[1]*

[1] *Department of Biomolecular Engineering, University of California, Santa Cruz, CA, USA*
[2] *Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA*

Edited by:
Frank T. Robb, University of California, USA

Reviewed by:
Mircea Podar, Oak Ridge National Laboratory, USA
Imke Schroeder, University of California Los Angeles, USA
Matthias Hess, Washington State University, USA
Lanming Chen, Shanghai Ocean University, China

***Correspondence:**
Todd M. Lowe, Department of Biomolecular Engineering, University of California, 1156 High Street, Santa Cruz, CA 95064, USA.
e-mail: lowe@soe.ucsc.edu

A great diversity of small, non-coding RNA (ncRNA) molecules with roles in gene regulation and RNA processing have been intensely studied in eukaryotic and bacterial model organisms, yet our knowledge of possible parallel roles for small RNAs (sRNA) in archaea is limited. We employed RNA-seq to identify novel sRNA across multiple species of the hyperthermophilic genus *Pyrobaculum*, known for unusual RNA gene characteristics. By comparing transcriptional data collected in parallel among four species, we were able to identify conserved RNA genes fitting into known and novel families. Among our findings, we highlight three novel *cis*-antisense sRNAs encoded opposite to key regulatory (ferric uptake regulator), metabolic (triose-phosphate isomerase), and core transcriptional apparatus genes (transcription factor B). We also found a large increase in the number of conserved C/D box sRNA genes over what had been previously recognized; many of these genes are encoded antisense to protein coding genes. The conserved opposition to orthologous genes across the *Pyrobaculum* genus suggests similarities to other *cis*-antisense regulatory systems. Furthermore, the genus-specific nature of these sRNAs indicates they are relatively recent, stable adaptations.

Keywords: antisense small RNA, archaea, transcriptome sequencing, comparative genomics, gene regulation, C/D box small RNA

INTRODUCTION

Archaeal species are known to encode a plethora of small RNA (sRNA) molecules. These sRNAs have a multitude of functions including suppression of messenger RNA (mRNA; Straub et al., 2009), targeting modifications to ribosomal (rRNA) or transfer RNA (tRNA; Omer et al., 2000; Bernick et al., 2012), specifying targets of the CRISPR immune defense system (Barrangou et al., 2007; Hale et al., 2008; Hale et al., 2009), *cis*-antisense regulation of transposase mRNA (Tang et al., 2002; Tang et al., 2005; Jager et al., 2009; Wurtzel et al., 2010), and encoding short proteins less than 30 amino acids in length (Jager et al., 2009).

Only a few previous studies have described sRNA genes in the phylum Crenarchaeota. In the *Sulfolobus* genus, C/D box and H/ACA-box guide sRNAs have been studied, including 18 guide sRNAs in *Sulfolobus acidocaldarius* (Omer et al., 2000), nine in *S. solfataricus* (Zago et al., 2005), and corresponding homologs detected computationally in *S. tokodaii* (Zago et al., 2005). These sRNAs form two distinct classes of guide RNAs: C/D box sRNAs which guide 2′-*O*-methylation of ribose, and H/ACA-box guide RNAs which direct isomerization of uridine to pseudouridine. Eukaryotes also share these two classes of guide RNAs with the same functions, but these homologs are dubbed small nucleolar RNAs (snoRNAs) because of their cellular localization. Recently, we employed high-throughput sequencing to identify ten conserved, novel families of H/ACA-like sRNA within the genus *Pyrobaculum* (Bernick et al., 2012).

Sulfolobus solfataricus has been further characterized using high-throughput sequencing (Wurtzel et al., 2010), revealing 18 CRISPR-associated sRNAs, 13 C/D box sRNAs, 28 *cis*-antisense encoded transposon-associated sRNAs, and 185 sRNA genes encoded antisense to other, non-transposon protein coding genes. It is unclear how many of the latter antisense transcripts are the result of transcriptional noise, overlapping but non-interacting gene products, or biologically relevant products of functional ncRNA genes. The diversity of sRNA genes is just beginning to be studied in depth in other members of the Crenarchaeota.

Genes that produce sRNA antisense to mRNA are known in all three domains of life and many of these sRNA have provided interesting examples of novel regulation. Within bacteria, antisense sRNAs are known and well-studied (Repoila et al., 2003; Aiba, 2007; Vogel, 2009). For example, utilization and uptake of iron in *Escherichia coli* is modulated by the sRNA *RyhB* that acts in concert with the ferric uptake regulator (Fur) protein (Masse et al., 2007). The sRNA is coded in *trans* to its regulatory targets, and the Sm-like protein Hfq is required for its function. In *Pseudomonas aeruginosa*, an analogous regulatory mechanism exists with the *PrrF* regulatory RNA (Wilderman et al., 2004).

In this study, we adapted techniques pioneered by researchers studying microRNA in eukaryotes (Lau et al., 2001; Henderson et al., 2006; Lu et al., 2006), to execute parallel high-throughput pyrosequencing of sRNAs across four *Pyrobaculum* species. This comparative transcriptomic approach enabled us to identify

novel conserved sRNA transcripts among four related hyperthermophiles (*Pyrobaculum aerophilum*, *P. arsenaticum*, *P. calidifontis*, and *P. islandicum*). We provide an overview of the distribution of sRNAs across species, and focus on two major classes: the highly abundant C/D box sRNAs, and sRNAs antisense to three biologically important protein coding genes. We augment our transcriptional analyses further with comparative genomics utilizing two additional *Pyrobaculum* species with sequenced genomes: *P. neutrophilum* (recently renamed from *Thermoproteus neutrophilus*) and *P. oguniense* (NCBI GenBank accession NC_016885.1).

MATERIALS AND METHODS

CULTURE CONDITIONS

Pyrobaculum aerophilum cells were grown anaerobically in media containing 0.5 g/L yeast extract, 1× DSM390 salts, 10 g/L NaCl, 1× DSM 141 trace elements, 0.5 mg/L $Fe(SO_4)_2(NH_4)_2$, pH 6.5, with 10 mM $NaNO_3$. *P. islandicum* and *P. arsenaticum* cells were grown anaerobically in media containing 10 g/L tryptone, 2 g/L yeast extract, 1× DSM390 salts, 1× DSM88 trace elements, and 20 mM $Na_2S_2O_3$. *P. calidifontis* cells were grown aerobically in 1 L flasks using 500 mL media containing 10 g/L tryptone, 2 g/L yeast extract, 1× DSM88 trace metals, 15 mM $Na_2S_2O_3$, pH 6.8, loosely capped with moderate shaking at 125 rpm. Anaerobic cultures were grown in 2 L flasks with 1 L media, prepared under nitrogen with resazurin as a redox indicator at 0.5 mg/L; 0.25 mM Na_2S was added as a reductant. All cultures were grown at 95°C to late log or stationary phase, monitored at OD_{600}.

The 10× DSM390 salts are comprised of (per liter ddH_2O) 1.3 g $(NH_4)_2SO_4$, 2.8 g KH_2PO_4, 2.5 g $MgSO_4{\cdot}7H2O$. The 100× DSM88 trace metal solution is comprised (per liter 0.12 N HCl), 0.9 mM $MnCl_2$, 4.7 mM $Na_2B_4O_7$, 76 μM $ZnSO_4$, 25 μM $CuCl_2$, 12.4 μM $NaMoO_4$, 18 μM $VOSO_4$, 6 μM $CoSO_4$. The 100× DSM141 trace metal solution is comprised of 7.85 mM Nitrolotriacetic acid, 12.2 mM $MgSO_4$, 2.96 mM $MnSO_4$, 17.1 mM NaCl, 0.36 mM $FeSO_4$, 0.63 mM $CoSO_4$, 0.68 mM $CaCl_2$, 0.63 mM $ZnSO_4$, 40 μM $CuSO_4$, 42 μM $KAl(SO_4)_2$, 0.16 mM H_3BO_3, 41 μM Na_2MoO_4, 0.1 mM $NiCl_2$, 1.14 μM Na_2SeO_3.

cDNA LIBRARY PREPARATION

Two preparations were constructed for each of *P. aerophilum*, *P. islandicum*, *P. arsenaticum*, and *P. calidifontis* cultures, yielding a total of eight cDNA libraries. The following protocol was used for each preparation.

Total RNA was extracted from exponential or stationary cultures; 100 μg of each preparation was loaded onto a 15% polyacrylamide gel, and size selected in the range 15–70 nt. The gel was post-stained with SYBR Gold and the tRNA band was used as the upper exclusion point. The lower exclusion point was set at 75% of the region between xylene cyanol (XC) and bromophenol blue (BP) loading dye bands (Ambion protocol). Samples were eluted, EtOH precipitated, and 3′ linker (5′-adenylated, 3′ ddC) was added as described by Lau et al., 2001; IDTDNA, Linker 1). A second gel purification was performed as above, excising the gel fragment above the XC dye band to remove excess 3′ linker. The recovered linked RNAs were reverse transcribed (RT) using Superscript III (Invitrogen) with a DNA primer complementary to Linker 1. Following RT, Exonuclease I (EXO1, Thermo) was added to the RT reaction mixture, and incubated for 30 min to remove excess primer. We utilized standard alkaline lysis treatment with NaOH-EDTA at 80°C for 15 min to remove any residual RNA, as well as to inactivate the reverse transcriptase and the EXO1 ssDNA nuclease. Neutralization and small fragment removal was performed with water-saturated G50 columns (Ambion NucAway). The recovered single stranded cDNA was dried to near completion using a Servo SpeedVac, followed by a second 5′-adenylated linker addition (IDTDNA – Linker 2) to the cDNA using T4 RNA ligase (Ambion).

A 2 μL volume of this reaction was amplified by PCR (20 μL reaction, 16 cycles). This was followed by a second amplification (20 μL reaction, 16 cycles) using 2 μL from the first amplification as template using Roche 454-specific hybrid adapters based on the method described by Hannon[1]. A four-base barcode was included in the 5′ hybrid primer. The final reaction was cleaned using the Zymo clean kit following the manufacturer's protocol.

SEQUENCING AND READ MAPPING

Sequencing was performed using a Roche/454 GS FLX sequencer, and the GS emPCR Kit II (Roche). Sequencing reads described in this work are provided online via the UCSC Archaeal Genome Browser[2] (Chan et al., 2012).

Reads that included barcodes and sequencing linkers were selected from the raw sequencing data and used to identify reads from each of the eight pooled cDNA libraries. Reads were further consolidated, combining identical sequences with associated counts for viewing with the Archaeal Genome Browser. Reads were mapped to the appropriate genome [*P. aerophilum* (NC_003364.1); *P. arsenaticum* (NC_009376.1); *P. calidifontis* (NC_009073.1); *P. islandicum* (NC_008701.1); *P. oguniense* (NC_016885.1); *P. neutrophilum* (*T. neutrophilus*: NC_010525.1)] using BLAT (Kent, 2002), requiring a minimum of 90% identity (-minIdentity), a maximal gap of 3 (-maxIntron) and a minimum score (matches minus mismatches) of 16 (-minScore) using alignment parameters for this size range (-tileSize = 8 -stepSize = 4). Reads that mapped equally well to multiple positions in the genome were excluded from this study. The remaining, uniquely mapped reads were formatted and visualized as BED tracks within the UCSC Archaeal Genome Browser.

Of the 216,538 raw sequencing reads obtained, those that had readable barcodes and could be uniquely mapped to their respective genomes were: 39,294 in *P. calidifontis*, 30,827 in *P. aerophilum*, 31,206 in *P. arsenaticum*, and 42,951 in *P. islandicum*.

NORTHERN ANALYSIS

Northern blots were prepared using ULTRAhyb-Oligo (Ambion) following the manufacturer protocol[3] using Hybond-N+ (GE life sciences) membranes to transfer 10 μg/lane denatured total RNA (45 min, 50°C with glyoxyl loading buffer – Ambion). Size separation was conducted using 23 cm × 25 cm gels (1% agarose) in BPTE running buffer (30 mM bis-Tris, 10 mM PIPES, 1 mM EDTA, pH 6.5). The following DNA oligomers

[1] http://genoseq.ucla.edu/images/a/a9/SmallRNA.pdf
[2] http://archaea.ucsc.edu
[3] http://tools.invitrogen.com/content/sfs/manuals/8663MB.pdf

(Integrated DNA technologies) were used as probes: TFBi-iSense (CCTCCTCTGGAAAGCCCCTCAAGCTCCGA), TFBi-iAnti (TCGGAGCTTGAGGGGCTTTCCAGAGGAGG), PAEsR53 sense (GACCCCGATCGCCGAAAAATGACGAGTGGT).

COMPUTATIONAL PREDICTION OF ORTHOLOGOUS GENE CLUSTERS

Computational prediction of orthologous groups was established by computing reciprocal best BLASTP (Altschul et al., 1990; RBB) protein coding gene-pairs among pairs of four *Pyrobaculum* species. When at least three RBB gene-pairs select the same inter-species gene set (for example A pairs with B, B pairs with C, and C pairs with A), the cluster was considered an orthologous gene cluster.

COMPUTATIONAL PREDICTION OF C/D BOX sRNA HOMOLOG FAMILIES

C/D box sRNA homolog families were constructed from computational predictions with core C/D box features that were supported by transcripts from one or more of the four *Pyrobaculum* species (data from this study). Six *Pyrobaculum* genomes were searched for orthologs using these sRNA candidates as queries to BLASTN (Camacho et al., 2009). The highest scoring candidates were manually curated, then grouped into homologous C/D box sRNA families by multiple alignment.

RESULTS

SMALL RNA POPULATIONS

We prepared eight barcoded sequencing libraries using sRNA fractions (size range 16–70 nt) from anaerobic cultures of *P. aerophilum*, *P. arsenaticum*, *P. islandicum*, and an aerobic culture of *P. calidifontis*. These libraries were prepared using a 5′-independent ligation strategy (Pak and Fire, 2007) which preserves RNA strand orientation, captures both the 5′ and 3′ ends of the sRNA, and does not impose a bias for molecule selection based on 5′-phosphorylation state. Pyrosequencing, followed by selection of uniquely mapped sequence reads, allowed detection of reads associated with both known and novel genomic features (**Figure 1**), including:

(i) snoRNA-like guide RNAs, including known and novel C/D box sRNA and a new class of H/ACA-like sRNA (Bernick et al., 2012),
(ii) RNA sequences encoded *cis*-antisense (asRNA) to known protein coding genes,
(iii) RNA sequences derived from CRISPR arrays, thought to guide the CRISPR-mediated immune response,
(iv) unclassified novel sRNA, and
(v) degradation products of larger RNA including ribosomal RNA, messenger RNA and transfer RNA.

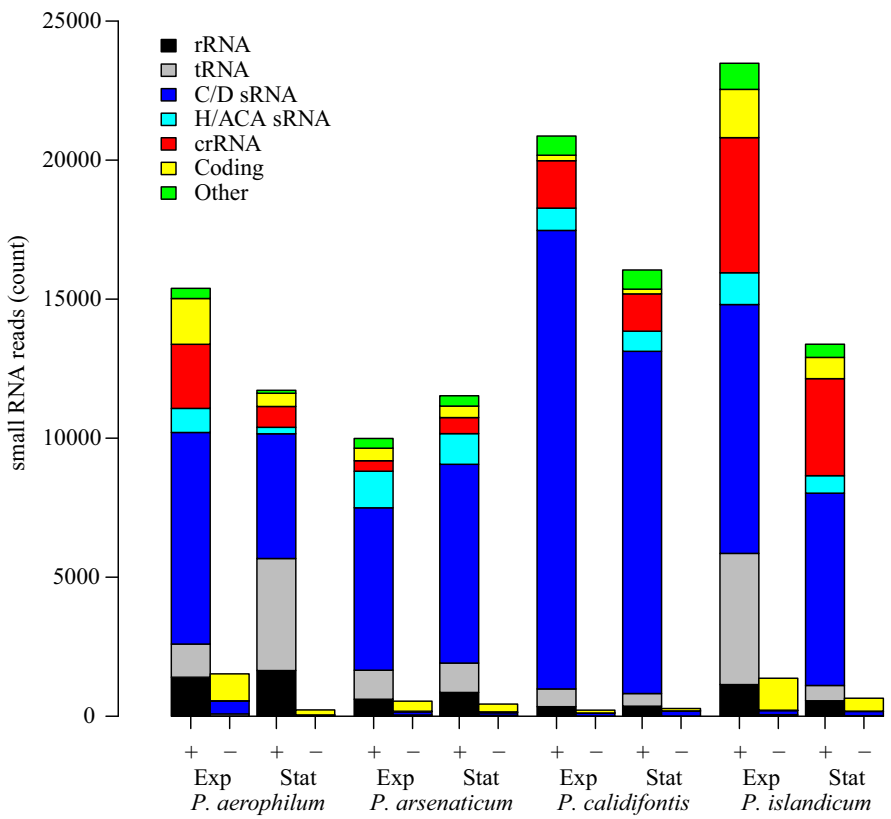


FIGURE 1 | Small RNA transcript abundance in four species of *Pyrobaculum*. Sense oriented reads (+) and antisense-oriented reads (−) shown in barplots for each species. Samples of each species were taken at both exponential (Exp) and stationary phases (Stat). RNA classifications were made based on mapping to genes coding for C/D box sRNA (C/D sRNA), H/ACA-like sRNA, CRISPR arrays (crRNA), fragments of coding regions (coding), ribosomal RNA (rRNA), and transfer RNA (tRNA).

Most antisense-oriented sequencing reads are associated with coding regions (**Figure 1**) in each of the species and growth phases examined. Antisense-oriented reads are frequently the result of convergent expression of a protein coding gene and a snoRNA-like guide RNA (**Tables A1–A4** in Appendix). We find, in some cases, that sequencing reads that appear to be antisense to snoRNA-like RNAs appear to be fragments of novel 3′ untranslated regions (3′ UTRs) of a convergently expressed protein coding region. These antisense-oriented sRNA reads are counted as antisense to the associated snoRNA-like sRNA. We made use of this transcriptional pattern to find novel C/D box sRNA and H/ACA-like sRNA; in these cases, highly abundant antisense reads to coding transcripts often proved to be a hallmark of novel C/D box and H/ACA-like sRNA (**Tables A2** and **A4** in Appendix). In a few remaining cases, we found novel *cis*-encoded antisense reads that were not derived from known classes of sRNA. We note that the proportion of reads belonging to each type of classified RNA is relatively stable across species and conditions (**Figure 1**), with the exception of two conditions in which tRNA fragments were enriched (*P. aerophilum* stationary phase, *P. islandicum* exponential phase). We are further investigating these differences, however the purpose and design of the sequencing portion of this study was aimed at qualitative discovery of novel sRNAs.

C/D BOX sRNA ACCOUNT FOR THE LARGEST FRACTION OF READS IN ALL SPECIES TESTED

In each of the eight small transcriptomes studied (four species sampled at exponential and stationary phase), C/D box sRNA accounted for the largest fraction of reads (**Figure 1**). A previous study (Fitz-Gibbon et al., 2002) has provided computational evidence for 65 C/D box sRNA candidates encoded in the genome of *P. aerophilum*. We now find an additional 23 C/D box sRNA candidates in that genome, representing a 35% increase in family size. By using transcriptional support from the four examined genomes (this study), combined with comparative genomic evidence that includes *P. oguniense* and *P. neutrophilum*, we find at least 74 C/D box sRNA in each *Pyrobaculum* spp. (**Table 1**). Of those genes, 70 appear to be conserved among all six genomes examined (**Figure 2**).

Table 1 | C/D box sRNA genes in each *Pyrobaculum* species based on transcriptional evidence or inferred by homology (*P. oguniense* and *P. neutrophilum*).

Species	C/D box sRNAs
P. aerophilum	88
P. arsenaticum	83
P. calidifontis	88
P. islandicum	84
P. oguniense	83
P. neutrophilum	74

All loci are manually curated.

CONVERGENTLY ORIENTED ncRNA ARE FREQUENTLY FOUND AT THE 3′ TERMINUS OF PROTEIN CODING GENES

It has been noted previously that in the genomes of *S. acidocaldarius* and *S. solfataricus*, C/D box sRNA genes occasionally exhibit antisense overlap to the 3′ end of protein encoding genes (Dennis et al., 2001). In the *Pyrobaculum* clade, we find numerous instances of a convergently oriented C/D box or H/ACA-like guide RNA gene that partially overlap, by a few nucleotides, the 3′ end of a protein-coding gene (**Tables A2** and **A4** in Appendix).

To find conserved, novel *cis*-encoded antisense RNA, we ranked conserved transcript abundance that overlapped orthologous protein coding genes. Among the top 34 predicted ortholog groups of genes with well-annotated function and conserved 3′ antisense transcription (**Table A2** in Appendix), 28 are convergent with C/D box sRNA and three are convergent with H/ACA-like sRNA. Among the top 19 predicted ortholog groups of unknown function with 3′ antisense transcription (**Table A4** in Appendix), 11 are convergent with C/D box sRNA, four are convergent H/ACA-like sRNA, and one is adjacent to a tRNA. Together, 87% of conserved, *cis*-antisense encoded sRNA are snoRNA-like guides, while only 2.6% are tRNA. In *P. aerophilum*, C/D box sRNA genes are nearly twice as abundant (88 compared to 46) as tRNA genes, but the sRNA genes are over 40-fold more likely to have a conserved overlap with the orthologous protein coding region. This may be an indication that these C/D box sRNA play a regulatory role with respect to the associated protein coding genes.

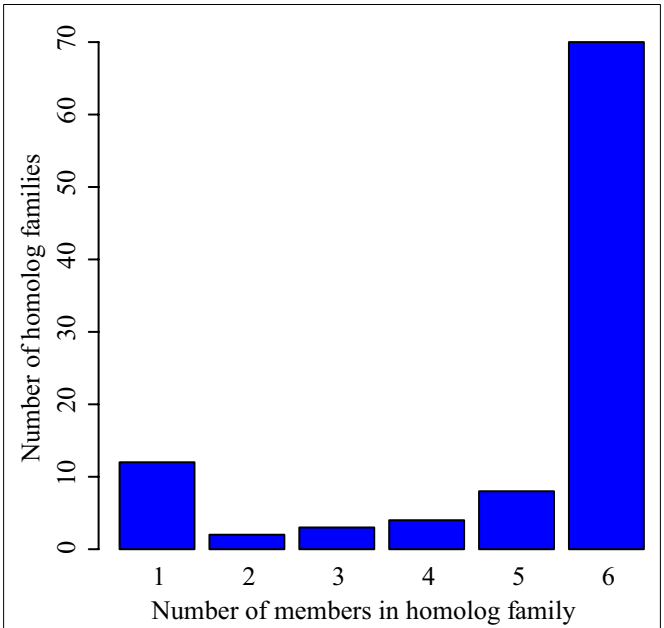


FIGURE 2 | Conservation of C/D box sRNA genes among six *Pyrobaculum* genomes. C/D box sRNA genes were organized by homolog family based on the location targeted by the encoded guide regions. These homolog families were then compared among the six studied species to verify conservation. While each individual species encodes more than 74 C/D box sRNA genes, 70 of those are conserved among each of the six studied *Pyrobaculum* spp. (group 6).

A notable example of a convergent ncRNA occurs at the 3′ terminus of the *electron transport flavoprotein* (*etf*) operon, where a C/D box sRNA, PAEsR53, overlaps the terminal gene (PAE0721 in *P. aerophilum*) in this four-gene operon. Like other operons within the *Pyrobaculum* genus, multiple promoters appear to drive expression of the *etf* operon (**Figure 3**). For this operon, an upstream promoter generates a 3400-nt-long full length *etfDH-ferredoxin-etfB-etfA* transcript. Two predicted internal promoters appear to generate respectively, the *ferredoxin-etfB-etfA* ∼2250 nt transcript, and the *etfA*-only 1040 nt transcript.

The *P. aerophilum* sRNA sequencing data revealed a strong abundance of sequences mapping to PAEsR53, as well as sequences of the same general size and location, mapping to the opposite strand (the UTR of the *etf* operon). Northern hybridization was performed to determine the origin of these "anti-PAEsR53" reads. **Figure 3** shows that these reads likely originate from the overlapping 3′ UTR of the *etf* operon, suggesting a possible interaction of the C/D box machinery with the *etf* mRNA. Predicted orthologs of this C/D box sRNA (PAEsR53) are syntenic with *etfA* in all *Pyrobaculum* species studied, overlapping the 3′ end of *etfA* orthologs by ∼12 bases. The overlap positions the D box guide sequence of PAEsR53 over the *etfA* stop codon in all *Pyrobaculum* species. If the guide RNA interacts through complementarity with the *etfA* mRNA, it could enable a 2′-*O*-methyl modification of the central "A" nucleotide within the conserved TAA stop codon in all four species.

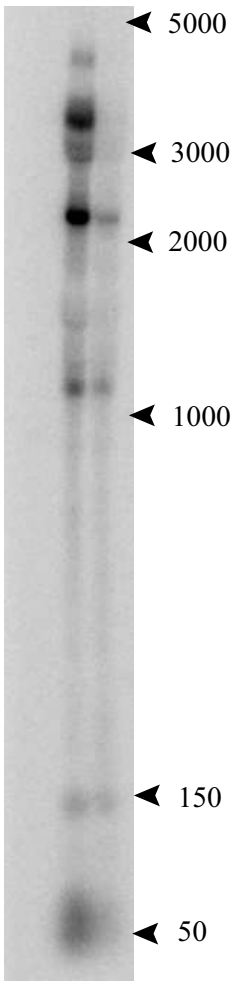


FIGURE 3 | Northern analysis of the 3′ UTR of the electron transport flavoprotein (*etf*) operon. *P. aerophilum* total RNA, exponential phase (left lane) and stationary phase (right lane). The probe was designed to anneal beyond the stop codon of the terminal gene in the *etf* operon, in the region of the convergently oriented PAEsR53 C/D box sRNA. Multiple bands at 3400, 2250, and 1040 nt are consistent with the *etf* operon and suboperon transcripts. The band near 50 nt, consistent with the RNA sequencing data, shows an apparent antisense transcript to PAEsR53 (sense relative to the 3′ UTR of the *etf* operon).

THE TRANSCRIPTION INITIATION FACTOR B GENES, *tfb*1 AND *tfb*2

The genomes of *Pyrobaculum* species contain a pair of paralogous genes that encode alternate forms of transcription initiation factor B (TFB). This factor is required for the initiation of basal level transcription at archaeal promoters (Santangelo et al., 2007).

In every sequenced *Pyrobaculum* species, TFB1 (PAE1645 and orthologs) contains a short N-terminal extension (22 amino acids in *P. aerophilum*) that is not present in the TFB2 proteins (PAE3329 and orthologs). Sequencing data reveals the presence of an abundant sRNA (*asR1*) encoded on the antisense strand that overlaps the 5′ end of *tfb1* (**Figure 4A**) in all four *Pyrobaculum* species examined (**Table A1** in Appendix). *Tfb1* also appears to have two promoters separated by 17–18 nt, such that the upstream promoter (P_u) is positioned to drive expression of full length *tfb1*, while the downstream promoter (P_d) generates transcripts that would lack a start codon near the start of the transcript.

In *P. aerophilum*, *asR1* sRNA is about 59 nt in length (**Table 2**; **Figure 4**), with a well-defined 5′ end that overlaps the extension region of the *tfb1* gene. The 3′ end of *asR1* is located just upstream of the *tfb1* translation initiation codon, precisely at the predicted start of transcription consistent with the P_u promoter. Importantly, there is an additional set of *asR1* sRNA reads of 41 nt in length, starting at the same 5′ position but terminating early, at the 5′ end of tfb1 transcripts consistent with the alternate P_d promoter. Mirroring the two variants of the antisense *asR1* transcript, deep sequencing revealed a large number of short sense strand sequencing reads, consistent with fragments representing the 5′ end of *tfb1* transcripts generated by P_u and P_d, spanning 50 and 32 nt in length respectively.

Northern analysis of total RNA from *P. aerophilum* confirmed the presence of a population of sense oriented transcripts of about 1000 nt in length, consistent with full length mRNA and another transcript population consistent with the sense oriented sRNAs described above (**Figure 5A**). When the antisense sRNA is probed, a population of short transcripts near 50 nt is detected (**Figure 5B**). The full length sense transcripts appear to be relatively constant in abundance across growth phase and culture conditions, consistent with data from a prior microarray study using the same RNA samples (Cozen et al., 2009). The correlated abundance of sense and antisense sRNA (**Figures 5C–E**) suggests that these sense::antisense pairs are associated, potentially as a double-stranded RNA. The elevated abundance of these pairs relative to

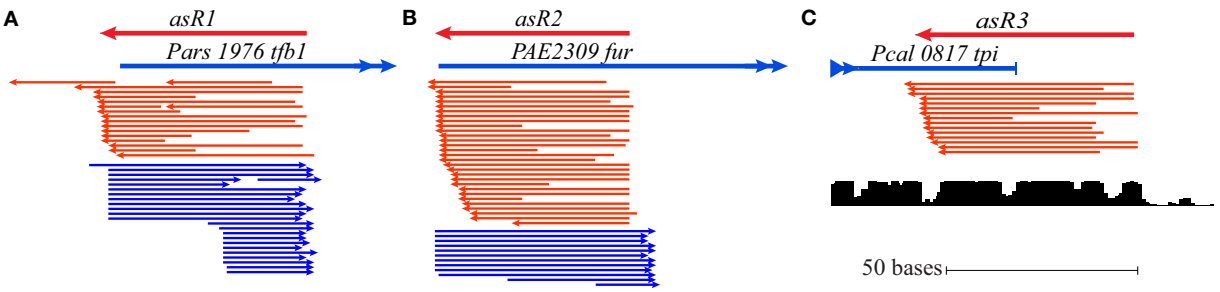


FIGURE 4 | *Cis*-antisense transcription occurring at three conserved loci in *Pyrobaculum* spp. The antisense sRNA genes (indicated as *asR1*, *asR2*, and *asR3*) are defined over the region of antisense-oriented sequencing reads (shown in red) with the associated gene (sense orientation) shown in blue. **(A)** Sequencing reads map to the 5′ terminus of transcription initiation factor B in both sense and antisense orientations (Pars_1976 is shown from *Pyrobaculum arsenaticum*). Sense oriented transcripts (blue) are present in two populations, consistent with the two conserved promoters, P_u and P_d. **(B)** The antisense sRNA (*asR2*) gene is defined over the region of antisense-oriented reads that map to the 5′ terminus of the ferric uptake regulator gene (*fur*; PAE2309 from *Pyrobaculum aerophilum* is shown). **(C)** Antisense-oriented reads map to the 3′ end of the triose-phosphate isomerase (*tpi*) locus (Pcal_0817 from *Pyrobaculum calidifontis*). Sequence conservation [lower graphic **(C)**] extends beyond the *tpi* stop codon (UCSC Archaeal Browser; Chan et al., 2012).

Table 2 | Terminal cis-antisense encoded sRNA in *Pyrobaculum* species.

sRNA		Len	Sequence
asR1	Pae	59	5′-AACTCGGAGCTTGAGGGGCTTTCCAGAGGAGGGGGGATTTGAGACCGACATAGCGTGTT
	Par	79	5′-TATGCGGAGCTTTAGGGGCTTGCCGGAAGAAGGTAGGCTTGTACTCGACATAGCGTGTTTATAAGCTTTTCTAGCGTAT
	Pca	33	5′-..TACGGAATTTTAAGGGCTTGCCGGGCGGGGTAG
	Pis	63	5′-.ATACGTAGCTTAAGGGGTTTCCCAGAAGACGTCGGACTTGACGACGACATAGCGAGTTTATAA
asR2	Pae	60	5′-GGGAGTCACTCTGTACCCCCTCTCCTTCAACGCTTGTACTAACTGGGCTGACTCCATCGT
	Pca	54	5′-......GACGCGGTATCCCTTCTCCTTTAGCGTGGCGACGAGCTGTGCCGTCTCCATAAT
asR3	Pae	65	5′-ACCCCCGAATTGGGGGCAAATGAGCGGCGGACACTTAAGGCGGCCCCGCCGCGAGCGGTTTCGCC
	Par	58	5′-.CCCCCGGA.CGGGGGCGAATGAGCGGCGGGCACCTGTGGCGGCTCCGCCGCGACTACT
	Pca	63	5′-.ACCCCGGA.TGGGGGCGGATGAGCGGCAGACACCTAAGGCGGCCCTGCCGCGACCAAGGGCTT
	Pis	59	5′-GACCCCCTGCTGGGGGCATATGAGCGGCGGGCACCTAAGGCGGCTCCGCCGCGACTGTA

Position of start codon (on coding strand) shown underlined for asR1 and asR2 (CAT). Position of stop codon (on coding strand) underlined (CTA, TTA) for asR3. Pae (P. aerophilum); Par (P. arsenaticum); Pca (P. calidifontis); Pis (P. islandicum); len (length of sRNA approximated from sequence read population).

the mRNA (**Figure 5A**) suggests that the sRNA pairs are stabilized within a dsRNA complex. The role of *asR1* with respect to *tfb1* transcripts is unclear, though the modulation of sRNA (both sense and antisense) while *tfb1* mRNA remains at constant and low abundance is reminiscent of negative feedback control.

The presence of complementary sense and antisense transcripts has been observed in a previous RNA sequencing study (Tang et al., 2005). Those authors suggested that the presence of an antisense transcript might enhance the stability of the mRNA target. As exemplified with *tfb1*, the presence of *cis*-antisense transcripts in our data are often accompanied by the presence of complementary sense strand fragments of similar size. This observation suggests that formation of a dsRNA duplex between the antisense sRNA and the 5′ region of the mRNA target may trigger destabilization of the mRNA; or alternatively, that base pairing between the antisense sRNA and the 5′ end of the nascent mRNA early in elongation may trigger premature transcription termination. For either mechanism, the result appears to be a constant level of *tfb1* mRNA under a variety of different culture conditions and growth phases.

THE FERRIC UPTAKE REGULATOR GENE (*fur*)

In a number of bacteria, the ferric uptake regulator FUR, is a transcriptional regulator of genes encoding proteins involved in iron homeostasis and protection from the toxic effects of iron under aerobic conditions. Some bacteria also encode a FUR-associated sRNA, for example *ryhB*; its synthesis is negatively regulated by FUR. The *ryhB* sRNA functions as a negative regulator of genes whose transcription is indirectly activated by FUR. The mechanism of *ryhB* sRNA negative regulation involves base pairing followed by selective degradation of the targeted mRNA (Andrews et al., 2003).

A homolog of the *fur* gene is conserved in the genomes of all known *Pyrobaculum* species. Embedded in each of the associated genes and located about 75 nt downstream from the 5′ start codon is an antisense-oriented, promoter-like sequence. In the two studied facultative aerobes (*P. aerophilum* and *P. calidifontis*), we detected a novel 54 nt-long *cis*-antisense transcript (**Table A1** in Appendix), designated as *asR2*, with precise transcription initiation consistent with the noted antisense promoter-like sequence. The 3′ end of the *asR2* transcript (**Table 2**; **Figure 4B**) transcript

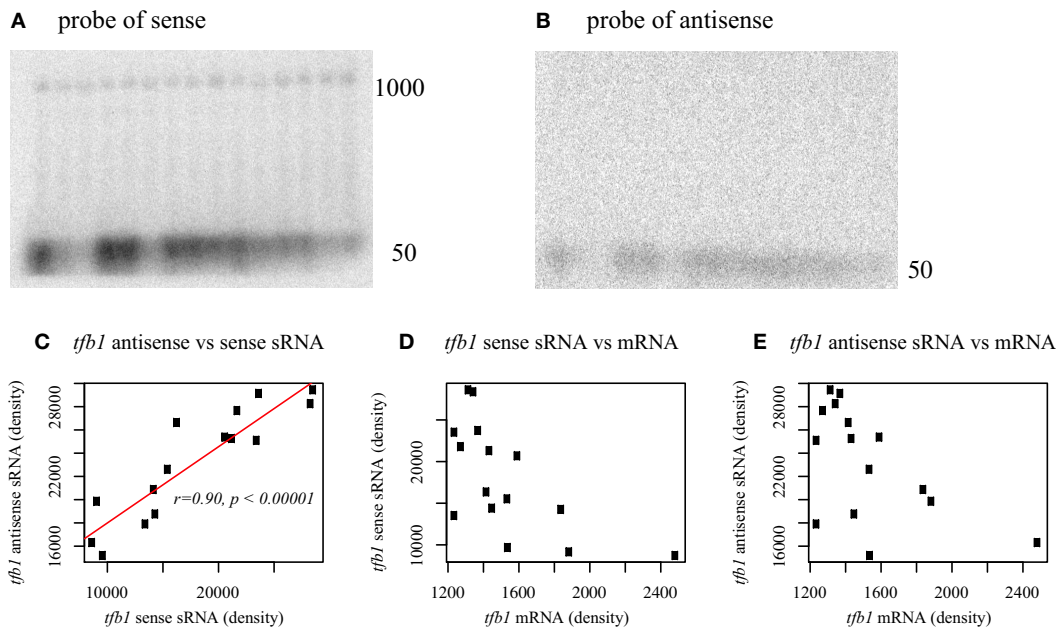


FIGURE 5 | Northern blot analysis of tfb1 from *Pyrobaculum aerophilum* using probes to the sense strand of *tfb* (A) and antisense strand (B). Sense transcripts of *tfb1* occur at both full length (∼1000 nt) and at 50 nt. Antisense transcripts occur at ∼50 nt. Lanes 1–15 (upper panels, left to right); total RNA across five respiratory growth conditions in three time series. Lanes 1–3 stationary phase, Lanes 4–6 growth with O_2, 7–9 growth with NO_3, 10–12 growth with As(V), 13–15 Fe(III). Each set of three lanes extracted from a time series for all five respiratory conditions at $T = (2.5, 4.5, 7.5\,h)$ with indicated terminal electron acceptor. Sense and antisense sRNA transcript abundance, inferred from band density, is positively correlated across growth conditions **(C)**, while no significant correlation is found between full length *tfb1* mRNA and either sRNA population **(D,E)**. Full length *tfb1* transcripts [1000 nt] remain nearly constant under all conditions tested **(A)**. Band density established using imagej (http://rsb.info.nih.gov/ij/).

terminates just upstream of the *fur* translation start codon. Both the *asR2* transcript and a complementary RNA fragment apparently derived from the 5′ end of *fur* mRNA, were present at high levels in anaerobically grown *P. aerophilum* and at modest levels in *P. calidifontis*. In the strict anaerobes (*P. islandicum*, *P. arsenaticum*), it appears that sequencing depth was insufficient to resolve any antisense-sense pairs under the limited set of growth conditions; however, we note that the predicted promoter for *asR2* in the facultative aerobes is equally well-conserved across all *Pyrobaculum* species.

THE TRIOSE-PHOSPHATE ISOMERASE (*tpi*) GENE

The *tpi* gene encodes triose-phosphate isomerase, an enzyme that is central to the modified Embden–Meyerhoff glycolytic pathway in *Pyrobaculum* species (Reher et al., 2007). We detected a 65-nt-long antisense transcript *asR3* (**Table A2** in Appendix) that overlaps the 3′ end of the *tpi* gene (**Figure 4C**) in all four of the species examined. Upon further examination of the 3′ terminal portion of *tpi*, we also detected a conserved sequence and associated secondary structure that is present in all sequenced *Pyrobaculum* spp. (**Figure 6**), which we term the tpi-element. In *P. aerophilum*, *P. islandicum*, and *P. calidifontis*, the tpi-element includes the stop codon of *tpi*, while the entire element is encoded immediately downstream of the *tpi* stop codon in the remaining *Pyrobaculum* spp.

A dsRNA formed by an interaction of *asR3* with the tpi-element could potentially compete against the mRNA intramolecular structure, and thus modulate function of the highly conserved tpi-element. Alternatively, *asR3* might itself be the active element of the pair, and in that case, presence of free *tpi* transcript might act as a repressor of *asR3*. In this model, *asR3* may have other *trans* targets in the genome and play a more general role in coordination of glycolysis in *Pyrobaculum* species.

DISCUSSION

Comparative transcriptomics has revealed compelling, conserved cases of novel *cis*-encoded transcripts that are antisense to core protein coding genes involved in transcription initiation and metabolism. We have considered these most obviously as potential regulators of their opposite strand partners, but they might also have broader regulatory roles.

We found that 28 of the top 34 cases of conserved 3′ antisense expression among orthologous *Pyrobaculum* proteins of known function coincide with convergent C/D box guide RNAs. This finding suggests that guide directed 2′-*O*-methylation of the mRNA in the region or downstream of the stop codon might be an unrecognized component of mRNA metabolism and gene regulation. It has been shown that pseudouridine modification of a stop codon can suppress termination of translation (Karijolich and Yu, 2011), but there are currently no studies of the possible implications of 2′-*O*-methyl modification on mRNA translation or stability. Alternatively, the presence of abundant mRNA fragments at the 3′ end may indicate that a sense-antisense interaction between the C/D box sRNA and mRNA terminus results in truncation of the mRNA

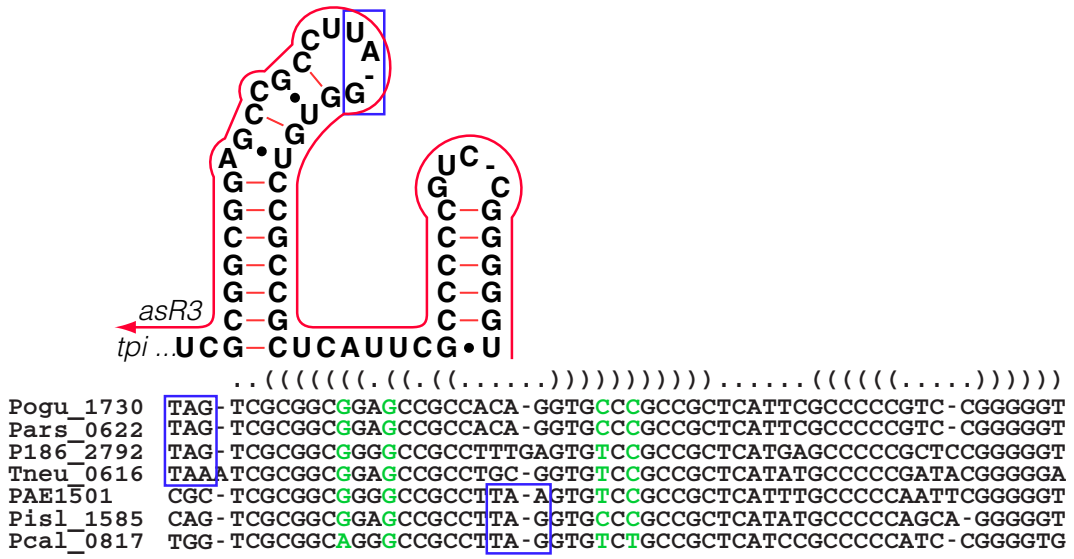


FIGURE 6 | The *Pyrobaculum* tpi-element and the associated antisense element, *asR3*. In all sequenced *Pyrobaculum* species, a highly conserved primary sequence forms a predicted secondary structure element (upper panel) at the 3′ end of the triose-phosphate isomerase gene (*tpi*). The depicted secondary structure contains the stop codon (blue box) in *P. aerophilum* (PAE1501), *P. islandicum* (Pisl_1585), and *P. calidifontis* (Pcal_0817). In *P. oguniense* (Pogu_1730), *P. arsenaticum* (Pars_0622), *Pyrobaculum* sp.1860 (P186_2792), and *P. neutrophilum* (Tneu_0616), the stop codon is immediately upstream of the tpi-element. *asR3* (red line) is encoded on the opposite strand, and has potential to compete/interfere with the tpi-element secondary structure. The genomic alignment of the 3′ portion of *tpi* and 3′ UTR is shown with the consensus secondary structure (nested parentheses). Base paired columns with one or more substitutions that maintain secondary structure are highlighted (green).

by an unknown mechanism, leading to mRNA destabilization and degradation.

The coordinated regulatory program implemented by Fur and its regulatory sRNA *ryhB* in some bacteria, provides a mechanism that yields both repression of some genes and activation (de-repression) of others. This program balances the needs of iron storage and utilization while protecting from iron-induced toxicity under oxic conditions. These dual roles are mediated by the inverse expression patterns of Fur and *ryhB*. Fur negatively regulates *ryhB*, which negatively regulates downstream genes. This circuit allows Fur to derepress (activate) those downstream genes. In published studies, active transcription in one direction can negatively regulate expression of the *cis*-encoded antisense partner (Lapidot and Pilpel, 2006), thus creating exclusive access to the shared genomic region. Likewise in *Pyrobaculum*, the *cis*-antisense transcription observed may yield the same type of inverse expression pattern. In this view, if the *cis*-antisense gene product is capable of repressing transcription or translation of targets in *trans*, then positive expression of *Pyrobaculum fur*, *tfb1*, or *tpi* may act through their corresponding antisense partner to activate (derepress) additional members of the associated regulon. Identification and verification of targets in *trans* is difficult in species that are not amenable to genetic manipulation such as *Pyrobaculum*, although future studies will explore computationally predicted targets.

The presence of *asR1*, a *cis*-encoded antisense RNA found within *tfb1* but not *tfb2* is of special interest when we consider these possible models of action for the *cis*-encoded antisense RNA. A specific TFB and TATA binding protein (TBP) pair in the archaeal halophile *Halobacterium* sp. NRC-1 has been shown to activate transcription under heat shock conditions (Coker and DasSarma, 2007). The observations that there are two instances of *tfb* in all *Pyrobaculum* genomes, and that only one harbors an antisense gene, suggest that *tfb1* might be essential only under particular conditions and/or initiate transcription for a subset of *Pyrobaculum* genes. Under this model, *tfb1* transcription might be held at low levels by the presence of *asR1* and possibly a dsRNA-binding complex. Under the alternative view, the *cis*-encoded *asR1* might facilitate activation of a *trans*-encoded regulon via de-repression. In the former view, the mechanism(s) that regulate sRNA transcription, stability, and mRNA interaction are central, while in the latter model, the sRNA is a downstream effector molecule of the independently regulated top-strand mRNA partner. In either case, resolving the molecular details of the sRNA's interaction with *tfb1* are needed to better understand this potential high-level mechanism for broad gene regulation in *Pyrobaculum*.

The tpi-element and its associated antisense partner, *asR3*, may provide a novel regulatory circuit acting from the 3′ UTR of *tpi*. The structure of the tpi-element (**Figures 4C** and **6**) contains the stop codon in some species while in other species the conserved structure is just downstream of the *tpi* stop codon. Possibilities for the function of the tpi-element include early transcription termination or translation termination. In either case, the tpi-element could be acting as a novel 3′ UTR riboswitch by binding a small molecule, or alternatively may be involved in a protein-binding event. Just as described above, the *cis*-antisense element *asR3*, encoded opposite the tpi-element, may act as a repressor of tpi-element function, or may have a role in *trans* with other genes in the *tpi* regulon.

In this study, we have described 74 or more expressed C/D box sRNA in each of four transcriptomes, most of which are conserved among multiple *Pyrobaculum* species. We have shown evidence that an unexpectedly large number of these sRNA overlap protein coding genes. Three novel sRNAs *asR1*, *asR2*, and *asR3* overlap genes involved in core transcription, iron regulation and core metabolism. Sequencing data have revealed the presence of sRNA originating from both strands, and these transcripts can be supported by promoter analysis, and verified by northern analyses. By contrast, less than 1% of transcripts mapped to CRISPR arrays show any evidence of dual strand transcripts (**Figure 1**). We suggest that the presence of dual-stranded transcript reads is an indication of an interaction of an sRNA with a convergently oriented mRNA, potentially mediated by one or more unknown dsRNA-binding complexes.

Future RNA-seq studies employing deeper sequencing technologies, alternative growth conditions, and other archaeal species will likely uncover many more cases of candidate regulatory antisense RNA. This work suggests multiple new research directions and will require complementary methodologies to better understand the complexity of sRNA function in Archaea. Given the conserved patterns of *cis*-antisense RNA transcripts now apparent, we anticipate rapid progress from follow-up studies that will demonstrate new modes of gene regulation homologous or analogous to those found in bacteria and eukaryotes.

ACKNOWLEDGMENTS

We are grateful to members of the Joint Genome Institute for making 454 sequencing possible (P. Richardson and J. Bristow for providing resources, and E. Lindquist and N. Zvenigorodsky for sample preparation and analysis). We thank Aaron Cozen for his generous procedural guidance and for the use of RNA blots used in the study. This work was supported by National Science Foundation Grant EF-082277055 (Todd M. Lowe and David L. Bernick); the Graduate Research and Education in Adaptive Bio-Technology (GREAT) Training Program sponsored by the University of California Bio-technology Research and Education Program (David L. Bernick); and by the National Science Foundation while Patrick P. Dennis was working at the Foundation. The opinions, findings, and conclusion expressed in this publications are ours and do not necessarily reflect the views of the National Science Foundation.

AUTHOR CONTRIBUTIONS

David L. Bernick designed and performed the experimental and computational analyses, and wrote the manuscript. Lauren M. Lui analyzed the C/D box sRNA sequencing data. Patrick P. Dennis provided assistance with the manuscript, collaborative review, and structure determination of C/D box sRNA. Todd M. Lowe provided scientific direction, contributed to interpretation of results, and edited the manuscript.

REFERENCES

Aiba, H. (2007). Mechanism of RNA silencing by Hfq-binding small RNAs. *Curr. Opin. Microbiol.* 10, 134–139.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Andrews, S. C., Robinson, A. K., and Rodriguez-Quinones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiol. Rev.* 27, 215–237.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.

Bernick, D. L., Dennis, P. P., Hochsmann, M., and Lowe, T. M. (2012). Discovery of *Pyrobaculum* small RNA families with atypical pseudouridine guide RNA features. *RNA* 18, 402–411.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421

Chan, P. P., Holmes, A. D., Smith, A. M., Tran, D., and Lowe, T. M. (2012). The UCSC Archaeal Genome Browser: 2012 update. *Nucleic Acids Res.* 40, D646–D652.

Coker, J. A., and DasSarma, S. (2007). Genetic and transcriptomic analysis of transcription factor genes in the model halophilic archaeon: coordinate action of TbpD and TfbA. *BMC Genet.* 8, 61. doi:10.1186/1471-2156-8-61

Cozen, A. E., Weirauch, M. T., Pollard, K. S., Bernick, D. L., Stuart, J. M., and Lowe, T. M. (2009). Transcriptional map of respiratory versatility in the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *J. Bacteriol.* 191, 782–794.

Dennis, P. P., Omer, A., and Lowe, T. (2001). A guided tour: small RNA function in archaea. *Mol. Microbiol.* 40, 509–519.

Fitz-Gibbon, S. T., Ladner, H., Kim, U. J., Stetter, K. O., Simon, M. I., and Miller, J. H. (2002). Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 984–989.

Hale, C., Kleppe, K., Terns, R. M., and Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14, 2572–2579.

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945–956.

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., and Jacobsen, S. E. (2006). Dissecting *Arabidopsis thaliana* DICER function in small RNA processing, gene silencing and DNA methylation patterning. *Nat. Genet.* 38, 721–725.

Jager, D., Sharma, C. M., Thomsen, J., Ehlers, C., Vogel, J., and Schmitz, R. A. (2009). Deep sequencing analysis of the Methanosarcina mazei Go1 transcriptome in response to nitrogen availability. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21878–21882.

Karijolich, J., and Yu, Y. T. (2011). Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* 474, 395–398.

Kent, W. J. (2002). BLAT – the BLAST-like alignment tool. *Genome Res.* 12, 656–664.

Lapidot, M., and Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.* 7, 1216–1222.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862.

Lu, C., Kulkarni, K., Souret, F. F., Muthuvalliappan, R., Tej, S. S., Poethig, R. S., Henderson, I. R., Jacobsen, S. E., Wang, W., Green, P. J., and Meyers, B. C. (2006). MicroRNAs and other small RNAs enriched in the *Arabidopsis* RNA-dependent RNA polymerase-2 mutant. *Genome Res.* 16, 1276–1288.

Masse, E., Salvail, H., Desnoyers, G., and Arguin, M. (2007). Small RNAs controlling iron metabolism. *Curr. Opin. Microbiol.* 10, 140–145.

Omer, A. D., Lowe, T. M., Russell, A. G., Ebhardt, H., Eddy, S. R., and Dennis, P. P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science* 288, 517–522.

Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315, 241–244.

Reher, M., Gebhard, S., and Schonheit, P. (2007). Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN), key enzymes of the respective modified Embden-Meyerhof pathways in the hyperthermophilic crenarchaeota *Pyrobaculum aerophilum* and Aeropyrum pernix. *FEMS Microbiol. Lett.* 273, 196–205.

Repoila, F., Majdalani, N., and Gottesman, S. (2003). Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol. Microbiol.* 48, 855–861.

Santangelo, T. J., Cubonova, L., James, C. L., and Reeve, J. N. (2007). TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription in vitro. *J. Mol. Biol.* 367, 344–357.

Straub, J., Brenneis, M., Jellen-Ritter, A., Heyer, R., Soppa, J., and Marchfelder, A. (2009). Small RNAs in haloarchaea: identification, differential expression and biological function. *RNA Biol.* 6, 281–292.

Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Hutenhofer, A. (2002). Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7536–7541.

Tang, T. H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J. P., and Huttenhofer, A. (2005). Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* 55, 469–481.

Vogel, J. (2009). A rough guide to the non-coding RNA world of *Salmonella*. *Mol. Microbiol.* 71, 1–11.

Wilderman, P. J., Sowa, N. A., Fitzgerald, D. J., Fitzgerald, P. C., Gottesman, S., Ochsner, U. A., and Vasil, M. L. (2004). Identification of tandem duplicate regulatory small RNAs in *Pseudomonas aeruginosa* involved in iron homeostasis. *Proc. Natl. Acad. Sci. U.S.A.* 101, 9792–9797.

Wurtzel, O., Sapra, R., Chen, F., Zhu, Y., Simmons, B. A., and Sorek, R. (2010). A single-base resolution map of an archaeal transcriptome. *Genome Res.* 20, 133–141.

Zago, M. A., Dennis, P. P., and Omer, A. D. (2005). The expanding world of small RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Mol. Microbiol.* 55, 1812–1828.

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

Received: 28 April 2012; accepted: 06 June 2012; published online: 02 July 2012.

Citation: Bernick DL, Dennis PP, Lui LM and Lowe TM (2012) Diversity of antisense and other non-coding RNAs in archaea revealed by comparative small RNA sequencing in four Pyrobaculum species. Front. Microbio. ***3****:231. doi: 10.3389/fmicb.2012.00231*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

APPENDIX

Table A1 | Orthologous genes with 5′ sequencing reads. Orthologous groups are shown in each row where the locus tag number (e.g., 1645 for gene PAE1645) is followed by counts of (antisense, sense) reads. Groups are ranked by the total number of reads found within groupings formed by the number of species in a group with antisense sequencing reads. Read counts are accumulated by considering the largest region covered by at least one read in an overlapping region along a given strand, and assigning the read count to that region. Footnoted gene IDs have associated snoRNA-like sRNA (C/D box or H/ACA-like) – a, antisense oriented; s, sense oriented.

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Transcription initiation factor IIB	1645 (225,409)	1976 (12,16)	0584 (1,1)	1667 (8,39)
DNA-cytosine methyltransferase	1659 (4,0)	1839 (1,0)	0576	1675 (2,0)
Rhomboid family protein	1099 (3,0)	0267 (1,0)	0686	1249 (2,0)
Ferric uptake regulator, Fur family	2309 (40,11)	1526	1653 (1,0)	1023
30S ribosomal protein S12P	0670	2326 (2,0)	2096	0698 (7,1)
Cobalamin adenosyltransferase	1715	0782	0623[a] (86,4)	1701 (0,3)
Thiol:disulfide interchange protein	3152	1672	1794	0523 (32,0)
30S ribosomal protein S11P	3179 (15,2)	1654 (0,2)	1813	0540
NAD-dependent deacetylase	3500	1959[a] (12,3)	1963	0793
NADH dehydrogenase subunit A	3520 (9,0)	1954 (0,1)	1983	0847
30S ribosomal protein S3P	1779	0769	0553	1729 (9,0)
Translation initiation factor IF–1A	1072 (7,0)	0278	0681	1256
Putative transcriptional regulator, GntR family	2315 (0,10)	1532 (4,2)	1659 (0,2)	1028
Valyl-tRNA synthetase	2297 (4,0)	1497 (0,1)	1649	1019 (0,1)
Putative signal-transduction protein with CBS domains	2961 (4,0)	1332 (0,1)	1143	0364
Major facilitator superfamily MFS_1	1550[s] (3,5)	0660[s] (0,2)	0530[s] (0,2)	
Elongation factor 1, beta/beta′/delta chain	0695 (3,1)	2345	2114	0684
Egghead-like protein	0042 (3,2)	1076	2043 (0,1)	0056
V-type ATP synthase subunit B	1146	0237 (3,0)	0698	1264
Conserved protein (possible ATP binding)	0793 (3,11)	0044	2138	1084
Putative transcriptional regulator, ModE family	0813 (2,0)	0057	0023	1100
50S ribosomal protein L18e	0672	2328 (2,0)	2098	0696 (0,1)
ABC transporter related	1393 (2,0)	0445	1879	1525
Peptidase M50	1702	2238 (2,1)	0618	1696
Cation diffusion facilitator family transporter	0568 (2,0)	2239	1215	0125 (0,2)
paREP10	1480	0613 (2,0)	0811 (0,1)	1575
Exosome complex RNA-binding protein Rrp42	2206 (2,1)	1938	0932	0835
Inner-membrane translocator	3412 (2,0)	1174	1046	0977
NADH-ubiquinone oxidoreductase subunit		2274	2047	0329 (1,0)
CopG domain protein DNA-binding domain protein	2357 (1,0)	1561	1689	0622
Inner-membrane translocator	3348 (1,0)	1760	0444	0590
Amino acid-binding ACT domain protein	2296	1510	1648	1018 (1,1)
Hydrogen sulfite reductase	2596 (1,0)	1213	1457	
DNA-directed RNA polymerase subunit P	2258 (0,1)	1825	1624 (1,0)	0899
DNA polymerase, beta domain protein	1893	0821 (1,0)	1502	
Phosphate ABC transporter, inner membrane subunit PstC	1396	0443	1881	1527 (1,0)
Nicotinamide-nucleotide adenylyltransferase	1438	0405 (1,0)	0794	1561
Peptidase S8 and S53, subtilisin, kexin, sedolisin	1983	2056		1805 (1,0)
Glu/Leu/Phe/Val dehydrogenase,C terminal	3438	1871	1031	0980 (1,0)
D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding	3320 (1,1)	1736	1741	0566
Electron transfer flavoprotein, alpha subunit	0721	2372 (1,0)	2132	0645 (0,1)
Sua5/YciO/YrdC/YwlC family protein	2978 (1,0)	1345	1129	0378 (0,1)
AAA ATPase	3527	1626 (1,0)	1978	0145
MazG nucleotide pyrophosphohydrolase	1159 (1,0)	0222	0722	
30S ribosomal protein S8P	2098	2009 (0,2)	0176	1865 (1,0)
Transcriptional regulator, XRE family	0783	0037 (1,0)	2145	1076

(Continued)

Table A1 | Continued

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Acyl-CoA dehydrogenase domain protein	2070	2103	0199	1853 (1,0)
2-dehydropantoate 2-reductase	3409 (1,0)	2003	0383	1363
FHA domain containing protein	0816	0060 (1,0)	0026	1103
PaREP1 domain containing protein	3235		0464 (0,3)	1514 (1,0)
30S ribosomal protein S19e	3043 (1,1)	1790	0988	0440 (0,39)
CutA1 divalent ion tolerance protein	2325	1539	1667	1044 (1,0)
Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	2075 (1,1)	2019	0203	1857
Inner-membrane translocator	2083 (1,0)	1504	0826	0317 (0,2)
30S ribosomal protein S7P	0733 (1,1)	0001	0006	0655 (0,3)
Ribosomal protein L11	3104 (1,0)	1602	1832	0464
Metallophosphoesterase	3211	1639 (1,0)	0239	1924
Acetolactate synthase, large subunit, biosynthetic type	3300	1724 (1,1)	1753	0554 (0,2)
NAD+ synthetase	1219 (0,1)	0310 (1,0)	0793	1302
30S ribosomal protein S3Ae	3472 (1,0)	1852	1182	0771 (0,1)
Band7 protein	0750	0015	2166 (1,0)	1055
TGS domain protein	1649	1844	0581	1670 (1,0)
MoaD family protein	0727 (0,1)	2368	2136 (1,0)	0649
Putative circadian clock protein, KaiC	0729 (1,1)	2366 (0,2)	0010	0651
Tryptophanyl-tRNA synthetase	3091 (1,0)	1612	1822	0454
Aldehyde ferredoxin oxidoreductase	0622 (1,4)	2285	2057	0738
Inner-membrane translocator	3350	1761	0445 (0,1)	0591 (1,0)
Tyrosyl-tRNA synthetase	0630	2290 (1,0)	2062	0733
NADH-quinone oxidoreductase, B subunit	2928	1001	1957	0336 (1,0)
Prephenate dehydratase	0893[s] (0,51)	0111	0075	1150 (1,0)

Table A2 | Orthologous genes with 3′ sequencing reads. Orthologous groups, read counts, and footnotes displayed are as described in **Table A1**.

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Electron transfer flavoprotein, alpha subunit	0721[a] (381,54)	2372[a] (258,14)	2132[a] (2145,21)	0645[a] (22,6)
DNA-directed RNA polymerase, M/15 kDa subunit	3480[a] (153,0)	1847 (1,0)	1177[a] (94,0)	0776[a] (31,0)
NAD-dependent deacetylase	3500[a] (83,2)	1959[a] (6,0)	1963 (1,0)	0793[a] (26,1)
SMC domain protein	2280[a] (2,7)	1811[a] (9,0)	1637[a] (7,1)	0884[a] (10,0)
Triosephosphate isomerase	1501 (2,1)	0622 (1,0)	0817 (13,0)	1585 (1,0)
Metallophosphoesterase	2243[a] (287,0)	1913	0956[a] (26,0)	0802[a] (63,0)
Resolvase, N-terminal domain	3513[a] (250,0)	1963[a] (66,0)	1967	0797[a] (27,0)
Succinate dehydrogenase subunit D	0719[a] (94,10)	2361[a] (129,0)	2130	0668[a] (46,0)
HhH-GPD family protein	0880[a] (23,2)	0101[a] (10,3)	0066	1140[a] (235,0)
Elongation factor EF-2	0332[a] (183,0)	2139	0213[a] (15,0)	1957[a] (3,0)
Twin-arginine translocation protein, TatA/E family subunit	1546[a] (32,16)	0666	0534[a] (53,1)	1615[a] (7,2)
Aldo/keto reductase	2929[a] (66,14)	1002[a] (1233,1)	0966	
Putative agmatinase	2260	1823	1626[a] (13,2)	0897[a] (611,2)
MazG nucleotide pyrophosphohydrolase	1159 (133,1)	0222 (352,2)	0722	
Ferric uptake regulator, Fur family	2309[a] (128,0)	1526[a] (141,1)	1653	1023
Seryl-tRNA synthetase	3158[a] (50,0)	1667	1802	0528[a] (39,36)
Uridylate kinase	3159	1665	1804 (20,0)	0530 (39,0)
Purine and other phosphorylases, family 1	1476	0610	0814[a] (24,0)	1572 (23,0)
Isoleucyl-trna synthetase	1617	1993[a] (5,0)	0601[a] (2,1)	1650
Transcriptional regulator, Fis family	3027[s] (4,10)	1779[a] (3,0)	0999[s] (0,5)	0429[s] (0,47)
GCN5-related N-acetyltransferase	3246	1807	1556[a] (4,0)	0488 (2,0)
Putative circadian clock protein, kaic	0729	2366[a] (5,0)	0010	0651 (1,0)
Conserved protein (RNA polymerase related?)	1975	2051 (1,0)	1587	1800 (1,0)
Lysine exporter protein (LYSE/YGGA)	2077	2018	0708[a] (5260,0)	1858
Translation initiation factor IF-2 subunit gamma	0064	1171	0242[a] (162,0)	0078
Alpha-l-glutamate ligases, rimk family	1818	0723	0506[a] (116,0)	1747
Oxidoreductase, molybdopterin binding	0389	0833	1263	1366[a] (84,0)
Ribosomal protein L25/L23	1972	2048	1585[a] (66,8)	1798
Proliferating-cell nuclear antigen-like protein	0720	2362	2131[a] (36,0)	0667
3-dehydroquinate synthase	1685	1827	0566[a] (25,0)	1689
DNA polymerase, beta domain protein region	1153	1067	0856 (23,0)	
Haloacid dehalogenase domain protein hydrolase	1785	0739	0554[a] (20,0)	1734
Mn2+-dependent serine/threonine protein kinase	2192	1948	0924	0825 (12,0)
Radical SAM domain protein	2153 (0,1)	0818	1068	0189[a] (9,0)
DNA polymerase I	2180	0798	1087	0816 (6,0)
Ribonuclease HII	1216	0312	0780	1305 (5,0)
Bifunctional GMP synthase/glutamine amidotransferase protein	3369	1772	1723	0600 (4,0)
Band 7 protein	0750	0015	2166	1055 (4,0)
Alpha-l-glutamate ligases, RimK family	0645 (4,0)	2302	2074	0721
Ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent	3155 (4,0)	1670	1797	0525
Thermosome	3273 (0,3)	1704 (0,1)	1771	0501 (3,5)
Metallophosphoesterase	1087 (3,0)	0270	1512	1254
Peptidase M24	2025	2086	1010	1836 (3,0)
Pyruvate/ketoisovalerate oxidoreductase, gamma subunit	3279	1708	1767	0497 (3,0)
DNA polymerase, beta domain protein region	0045	1137 (3,0)	0385	
Creatininase	2983	1350	1124	0383 (2,0)
Acetyl-CoA acetyltransferase	1220	0309	0781	1301 (2,2)
Indole-3-glycerol-phosphate synthase	0570 (2,0)	2240	1213	0124 (0,2)
Regulatory protein, ArsR	0731	2364	0008 (2,0)	0653
PaREP1 domain containing protein	0002	1095	1373[a] (2,0)	
Putative signal-transduction protein with CBS domains	3588		1394	0254 (2,0)

(Continued)

Table A2 | Continued

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Exosome complex exonuclease Rrp41	2207 (2,1)	1937	0933	0836
ABC transporter related	3413 (2,0)	1175	1045	0976
Uroporphyrinogen III synthase HEM4	0589	2250	1712	0116 (2,0)
Potassium transport membrane protein, conjectural	2422 (2,0)	1446	0314	0883
Undecaprenyl diphosphate synthase	2942 (2,0)	1319	1157	0348
Nucleotidyl transferase	0837	0080	0043	1119 (2,0)
Carbon starvation protein CstA	1423 (2,0)	0894	0860	
Carboxypeptidase Taq	0885 (1,2)	0104	0069 (0,1)	1143
Leucyl-tRNA synthetase	1107	0260	0691	1246 (1,0)
HEPN domain protein	1894	0820 (1,0)	1501	
DNA-directed RNA polymerase subunit E, RpoE2	3563 (1,0)	2230	1991	0921
5-carboxymethyl-2-hydroxymuconate Δ-isomerase	2688	0535	1503 (1,0)	
ATPase	1789	0736		1446 (1,0)
Oligosaccharyl transferase, STT3	3030 (1,0)	1781	0997	0431
paREP7	0906		0492	0185 (1,0)
Haloacid dehalogenase domain protein hydrolase	2017^{s} (0,12)	2080^{s} (0,15)	1016^{s} (1,4)	1830
Egghead-like protein	0042	1076	2043	0056 (1,0)
Putative transcriptional regulator, CopG family	1443 (0,2)	0399	0796	1563 (1,1)
Asparaginyl-tRNA synthetase	2973 (1,0)	1342	1133	0375
Succinate dehydrogenase iron-sulfur subunit	0717	2359	2128	0670 (1,0)
Peptidase T2, asparaginase 2	3083 (1,0)	1892	0970	0908
Radical SAM domain protein	0596	2255	1716	0113 (1,0)
30S ribosomal protein S25e	2188 (0,1)	0790 (0,1)	1079	0808 (1,0)
Ribosomal-protein-alanine acetyltransferase	2246 (1,0)		0958	1001
PilT protein domain protein	3561 (1,0)	1614	1989	0923
Peptidase S8 and S53, subtilisin, kexin, sedolisin	0712	2355	2124 (1,0)	0674
Nitrilase/cyanide hydratase and apolipoprotein N-acyltransferase	2075^{s} (0,49)	2019	0203	1857^{s} (1,8)
Beta-lactamase domain protein	2160	0810	1074	0803 (1,0)
Xanthine dehydrogenase accessory factor	2669	0253	1324 (1,0)	
ABC transporter related	3269	1702 (1,0)	1774	0503
tRNA CCA-pyrophosphorylase	3325	1740	1737	0570 (1,0)
Starch synthase	3429	1878 (1,0)	1038	0968
Dual specificity protein phosphatase	1536	0675 (1,0)	0541	1603
Putative endoribonuclease L-PSP	3003^{s} (1,62)	1258^{s} (0,210)	1096^{s} (0,81)	0414^{s} (0,65)
Sulfite reductase, dissimilatory-type beta subunit	2597	1212 (1,0)	1456	
Methyltransferase small	0261	2199	0236	0747 (1,0)
Putative transcriptional regulator, AsnC family	1507 (1,0)	0627 (0,1)	0822	1590
Methyltransferase type 11	1165	0216 (1,0)	1364	1338
Serine/threonine protein kinase	0815	0059 (1,0)	0025	1102
Transcriptional regulator, PadR-like family	0013	1087 (1,0)		0038
Inner-membrane translocator	3350 (1,0)	1761	0445	0591
Geranylgeranyl reductase	2989	1355	1119	0388 (1,0)
Extracellular solute-binding protein, family 5	2391	1494	0422	0602 (1,0)
2-methylcitrate synthase/citrate synthase II	1689 (1,0)	2234	0563	1692
30S ribosomal protein S6e	1505 (0,1)	0626 (1,0)	0821	1589

Table A3 | Hypothetical genes with 5′ sequencing reads. Orthologous groups, read counts, and footnotes displayed are as described in **Table A1**.

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Hypothetical protein	3282 (3,3)	1710	1765	0495 (11,6)
Hypothetical protein	0301	2159[a] (673,0)	0225	2002
Hypothetical protein	3499	1958	1962	0792[a] (26,1)
Hypothetical protein	0432 (6,0)		0474	0140
Hypothetical protein	1798	0175 (4,0)	0509	1742
Protein of unknown function DUF107	0749	0014	2167	1056 (4,0)
Hypothetical protein	1503 (3,0)	0624	0819	1587
Hypothetical protein	2934	1279 (3,0)	1165	0340 (0,1)
Hypothetical protein	1517	0632 (3,0)	0877	
Hypothetical protein	3546	1625 (3,0)	1977	0933
Hypothetical protein	0433 (3,6)		0479	0139
Hypothetical protein	3051[s] (0,284)	1797	0981[s] (0,152)	0447[s] (3,129)
Hypothetical protein	0838 (2,2)	0081	0044	1120
Hypothetical protein	1710 (0,3)	0785 (2,3)	0620	1698
Hypothetical protein	0728	2367 (2,0)	2137	0650
Hypothetical protein	1147	0229	0706	1284 (2,0)
Hypothetical protein	1522	0636	0874	1594 (2,0)
Hypothetical protein	2941 (2,0)	1318	1158	0347
Hypothetical protein	2338	1549	1677	0634 (2,0)
Hypothetical protein	2822	0279	1187	1388 (2,1)
Hypothetical protein	1943	2025 (2,0)	1297	
Hypothetical protein	2416 (2,0)	1479	0319	0879 (0,1)
Hypothetical protein	0746 (2,0)	0012	2168	1006
Hypothetical protein	1069 (0,13)	0281	0680	1257 (2,0)
Hypothetical protein	3081 (2,0)	1891	0969	0907 (0,1)
Hypothetical protein	0800	0050	0016 (0,1)	1092 (2,4)
Protein of unknown function DUF77	1158	0223 (2,0)	0711	1327
Hypothetical protein	3135 (0,2)	1683	1783	0512 (2,0)
Hypothetical protein	1683	1828 (1,0)	0567 (0,1)	1688
Hypothetical protein	0789	0040 (1,0)	2142	1079
Protein of unknown function DUF72	2078 (1,0)	2017	0205	1859
Hypothetical protein	1641 (0,3)	1979 (1,1)	0587	1664 (0,1)
Protein of unknown function DUF54	2213	1931	0938	0842 (1,0)
Protein of unknown function DUF437	0638 (1,0)	2296	2068	0727
Hypothetical protein	3550	1622	1974	0930 (1,0)
Hypothetical protein	3467	1857	1188	0993 (1,0)
Hypothetical protein	1318	0471		1365 (1,1)
Hypothetical protein	3556 (1,0)	1619	1969	0927
Hypothetical protein	2598	1211 (1,0)	1455	
Hypothetical protein	1643 (1,0)	1977	0585	1666
Hypothetical protein	2824	0884 (1,0)	0867	1387
Hypothetical protein	2322	1537	1664	1037 (1,0)
Hypothetical protein	2177	0799	1088	0817 (1,0)
Hypothetical protein	1449 (0,1)	0601 (0,1)	0800 (1,0)	1567 (0,1)
Protein of unknown function DUF52	0818	0062 (1,0)	0028	1105
Hypothetical protein	1173	0212	0743	1320 (1,0)
Hypothetical protein	3004[s] (1,62)	1259	1095	0415[s] (0,65)
Protein of unknown function DUF72	3079 (0,1)	1889 (1,3)	0967	0905
Hypothetical protein	2268 (1,0)	1820	1629	0894
Hypothetical protein	3324	1739	1738 (0,1)	0569 (1,0)

(Continued)

Table A3 | Continued

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Protein of unknown function UPF0027	0998	0172	0141 (1,0)	1219
Hypothetical protein	2210	1934 (1,0)	0941	0839
Hypothetical protein	1797	0174 (1,1)	0508	1741
Hypothetical protein	0718	2360	2129	0669 (1,0)
Hypothetical protein	1448	0600	0799	1566 (1,0)
Hypothetical protein	1613	2156	0603	1648 (1,10)
Hypothetical protein	1018	0184	0656 (0,1)	1203 (1,0)
Hypothetical protein	2429	1441	0309	1892 (1,0)
Hypothetical protein	3148	1674	1792	0521 (1,0)
Hypothetical protein	1676	1833		1682 (1,0)
Hypothetical protein	2403 (1,0)	1470	0327	0871 (0,3)

Table A4 | Hypothetical genes with 3′ sequencing reads. Orthologous groups, read counts, and footnotes displayed are as described in **Table A1**.

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Protein of unknown function DUF6, transmembrane	1545[s] (16,32)	0667[s] (2,20)	0535[s] (1,53)	1614[s] (2,7)
Hypothetical protein	1519[a] (550,1)	0634[a] (197,0)	0875[a] (202,0)	1596
Hypothetical protein	0577[a] (6,0)	2243[a] (18,7)	1195[a] (2,0)	0121
Hypothetical protein	1836 (0,2)	0710 (1,0)	0502 (5,0)	1752 (2,0)
Hypothetical protein	3249 (4,0)	1805	1855	0485 (2148,0)
Hypothetical protein	1234 (349,0)	0295 (1,0)	1511	0244
Protein of unknown function DUF1614	2020	2082 (15,0)	1014	1832 (58,0)
Hypothetical protein	1687[a] (48,4)	2232[a] (17,3)	0565	1690
Hypothetical protein	3138[a] (15,23)	1680[a] (27,2)	1786	0515
Hypothetical protein	0889	0108	0073 (4,0)	1147[a] (28,3)
Protein of unknown function DUF192	2955[a] (15,1)	1329	1147[a] (10,4)	0358
Hypothetical protein	3550	1622	1974 (1,0)	0930[s] (2,44)
Hypothetical protein	3005[a] (177,0)	1260	1094	0416
Hypothetical protein	3630	1030	2015	0003 (74,0)
Hypothetical protein	3245[a] (63,1)	1808		0487
Hypothetical protein	2069	2102	0198[a] (44,0)	1852
Hypothetical protein	3468	1856	1186	0994 (31,0)
Hypothetical protein	0730	2365 (18,0)	0009	0652
Hypothetical protein	3497	1956	1958 (12,6)	0790
Hypothetical protein	0936	0136 (8,0)	0105	1172
Hypothetical protein	3135 (0,1)	1683	1783[s] (8,6)	0512[s] (0,116)
Hypothetical protein	3156	1669	1798	0526 (6,0)
Hypothetical protein	0748	0013[s] (6,6)		1058[s] (0,1982)
Protein of unknown function DUF62	3627	2209[s] (0,5)	2013	0002 (6,0)
Hypothetical protein	1177	0209	0746[a] (5,16)	1317
Hypothetical protein	3189	1645	1819	0545 (4,0)
Hypothetical protein	3295 (4,0)	1719	1758	0550
Hypothetical protein	2549	0845 (3,0)	1351	0257
Hypothetical protein	1173 (3,0)	0212	0743	1320
Protein of unknown function UPF0027	0998	0172 (3,0)	0141	1219
Hypothetical protein	2504 (3,0)	1412		1927
Hypothetical protein	2326	1541	1669	1042 (3,0)
Hypothetical protein	1549[s] (3,5)	0661[s] (0,2)	0531[s] (0,5)	1618[s] (0,4)
Hypothetical protein		0611[s] (2,15)	0813[s] (0,24)	1573 (0,23)
Protein of unknown function DUF64	0371		1533	1367 (2,0)
Hypothetical protein	2285 (2,0)	1583	1640	1010
Hypothetical protein	1307	0473		1351[s] (2,14)
Hypothetical protein	1816	0724 (2,0)		0422 (0,1)
Hypothetical protein	3251 (2,0)	1803	1853	0483
Hypothetical protein	1497	0619	0805	1582 (2,0)
Hypothetical protein	1895	0678 (2,0)	0646	
Protein of unknown function DUF224, cysteine-rich region domain protein	1762	0754	0545	1721 (2,0)
Hypothetical protein	3568	2225	1996 (2,0)	0916
Hypothetical protein	1998	2066	1030	1817 (2,0)
Protein of unknown function DUF115	2328	1542	1670	1041 (2,0)
Hypothetical protein	2337	1548	1676	0635 (2,0)
Protein of unknown function DUF100	0944 (1,0)	0141	0111	1181
Protein of unknown function DUF340, membrane	1479	0612	0812	1574 (1,0)
Hypothetical protein	3304 (1,0)	1727	1750	0557
Hypothetical protein	0882 (1,0)	0102	0067	1141
Hypothetical protein	1130	0243 (1,0)	0697	1241

(Continued)

Table A4 | Continued

Product	*P. aerophilum*	*P. arsenaticum*	*P. calidifontis*	*P. islandicum*
Hypothetical protein	1449	0601	0800	1567 (1,0)
Hypothetical protein	2190	1946	0927	0827 (1,0)
Hypothetical protein	0927 (1,0)	0131	0083	1161
Hypothetical protein	0708	2353	2122	0676 (1,0)
Hypothetical protein	2606	1232 (1,0)	1431	
Hypothetical protein	2187	0791 (1,0)	1080	0809 (0,1)
Hypothetical protein	0239	1512 (1,0)	0690	
Protein of unknown function DUF1028	3380	1006 (1,0)	0160	
Hypothetical protein	2154 (1,0)	0817	1069	0942
Hypothetical protein	3161	1666	1803	0529 (1,0)
Hypothetical protein	2058	2311	0401	1850 (1,0)
Hypothetical protein	0840	0083 (1,0)	0046	1122

frontiers in MICROBIOLOGY

ORIGINAL RESEARCH ARTICLE
published: 13 July 2012
doi: 10.3389/fmicb.2012.00251

Comparative genomic and transcriptional analyses of CRISPR systems across the genus *Pyrobaculum*

David L. Bernick[1], Courtney L. Cox[1], Patrick P. Dennis[2] and Todd M. Lowe[1]*

[1] Department of Biomolecular Engineering, University of California, Santa Cruz, CA, USA
[2] Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

Edited by:
Zvi Kelman, University of Maryland, USA

Reviewed by:
Uri Gophna, Tel Aviv University, Israel
Qunxin She, University of Copenhagen, Denmark

***Correspondence:**
Todd M. Lowe, Department of Biomolecular Engineering, University of California, 1156 High St, Santa Cruz, CA 95064, USA.
e-mail: lowe@soe.ucsc.edu

Within the domain Archaea, the CRISPR immune system appears to be nearly ubiquitous based on computational genome analyses. Initial studies in bacteria demonstrated that the CRISPR system targets invading plasmid and viral DNA. Recent experiments in the model archaeon *Pyrococcus furiosus* have uncovered a novel RNA-targeting variant of the CRISPR system. Because our understanding of CRISPR system evolution in other archaea is limited, we have taken a comparative genomic and transcriptomic view of the CRISPR arrays across six diverse species within the crenarchaeal genus *Pyrobaculum*. We present transcriptional data from each of four species in the genus (*P. aerophilum*, *P. islandicum*, *P. calidifontis*, *P. arsenaticum*), analyzing mature CRISPR-associated small RNA abundance from over 20 arrays. Within the genus, there is remarkable conservation of CRISPR array structure, as well as unique features that are have not been studied in other archaeal systems. These unique features include: a nearly invariant CRISPR promoter, conservation of direct repeat families, the 5′ polarity of CRISPR-associated small RNA abundance, and a novel CRISPR-specific association with homologues of *nurA* and *herA*. These analyses provide a genus-level evolutionary perspective on archaeal CRISPR systems, broadening our understanding beyond existing non-comparative model systems.

Keywords: *Pyrobaculum*, CRISPR, sRNA, crRNA, repeat, RNAseq

INTRODUCTION

CRISPR immunity systems, like the vertebrate adaptive immune system (Boehm, 2011), include mechanisms to adapt to new pathogens, surveillance methods for detecting previously encountered pathogens, and means to inactivate those pathogens. In the case of the CRISPR system, the targeted molecule is a nucleic acid sequence, and the sequence of events moves from adaptation, where the invading nucleic acid sequence is recognized and acquired, to expression, where the CRISPR-specific small RNA recognition molecules (crRNA) are transcribed, processed and loaded by the CAScade complex of CRISPR-specific proteins (Brouns et al., 2008; Jore et al., 2011). The third phase, interference, is initiated upon detection of a targeted nucleic acid sequence and results in specific inactivation of the recognized nucleic acid from the invading "pathogen." DNA of viral or plasmid origin has been shown to be the target of CRISPR defense in bacteria (Barrangou et al., 2007; Marraffini and Sontheimer, 2008) and the archaeon *Sulfolobus solfataricus* (Manica et al., 2011). RNA sequences are targeted in the CRISPR system present in *Pyrococcus furiosus* (Hale et al., 2009, 2012), opening the possibility of endogenous targeting of messenger RNA sequence.

Most archaeal and many bacterial genomes contain one or more loci that encode the CRISPR system. Each CRISPR locus consists of an array of short DNA sequences, and frequently includes a cluster of CRISPR-associated (CAS) protein coding genes (Haft et al., 2005). The DNA arrays are composed of a leader sequence, followed by a set of 24–47 nucleotide (nt) direct repeats (DR) that form the delimiting punctuation of the array. The sequences between DR, termed spacers, are found to be 26–72 nt in length and encode small RNAs that are the stored immune memory for the system. The transcriptional promoter for the array is likely to be encoded within the leader sequence (Haft et al., 2005; Lillestol et al., 2009; Horvath and Barrangou, 2010). In *Escherichia coli*, the specific promoters for the array and associated CAS genes have been identified (Pul et al., 2010).

CRISPR arrays are dynamic structures, some containing only a single sequence while others may be quite large; for example, crispr4 in *Metallosphaera sedula* is over 10,000 nt in length and contains over 160 spacer sequences (Grissa et al., 2007). The genomes of most strains of *Methanococcus maripaludis* contain only one CRISPR array locus whereas the genome of strain S2 has no CRISPR array present. In contrast, the genomes of Methanocaldococcus strains encode between seven and 20 individual CRISPR arrays. In *Sulfolobus*, recent work has shown that selective pressure can be introduced *in vivo*, which results in deletion of genomic loci containing all or part of the CRISPR/CAS system (Gudbergsdottir et al., 2011).

Individual spacer elements in CRISPR arrays are acquired in the adaptation phase, during exposure to an invading genetic element. Evidence from surviving, phage-challenged cells shows an addition of one or more spacer sequences at the leader-proximal end of the array. These new spacer sequences are identical to phage sequence, can be from either phage genome strand, and confer immunity to survivor progeny (Barrangou et al., 2007). During this spacer acquisition phase, the target sequence is integrated into the array, likely through the action of CAS1, CAS2,

and possibly other CAS proteins (for example, CSN2 in the *Streptococcus thermophilus* Type II system). This adaptation process only requires a single direct repeat in the array (Yosef et al., 2012). It is unclear if the acquired DNA spacer is derived directly from invading DNA, or if the DNA spacer is a copy produced during the adaptation process.

The mechanism of immunity is still incompletely understood, but immunity is dependent on CAS genes (Barrangou et al., 2007; Brouns et al., 2008), usually located near one or more CRISPR arrays. Early studies showed that four CAS genes (*cas1–4*) were frequently associated with CRISPR arrays (Jansen et al., 2002; Haft et al., 2005). A role in CRISPR adaptation (acquisition of new spacers) has been proposed for *cas1* and *cas2* (Wiedenheft et al., 2012). Potentially, CAS4 is also involved during the acquisition phase; this hypothesis is based on the frequent *cas4* genomic proximity to *cas1* (Makarova et al., 2011).

The CAS genes have recently been reclassified into three main families based on gene content and mode of action of the associated system (Makarova et al., 2011). In Type I, II, and III-A CRISPR systems (Makarova et al., 2011), the target of the CRISPR immunity system is invading DNA (Marraffini and Sontheimer, 2008). In contrast, Type III-B systems target RNA instead of DNA (Hale et al., 2009, 2012). Type I systems have been studied in both bacteria and archaea, and have recently yielded low-resolution structures of the multimeric CAScade complex in both *E. coli* (Jore et al., 2011; Wiedenheft et al., 2011) and in the archaeon *Sulfolobus solfataricus* (Lintner et al., 2011). In Type I systems, the CAScade complex is required for maturing of CRISPR RNA (crRNA) that guide protective immunity during subsequent invasion by foreign DNA elements. This crRNA-enabled complex is also responsible for surveillance and eventual interference by recruiting additional CAS proteins (Wiedenheft et al., 2011). The primary transcript of the CRISPR array, pre-crRNA, is cleaved within the DR to generate the individual crRNA segments. In the Sulfolobus variant of CAScade, CAS6 is responsible for cleavage of pre-crRNA, while in *E. coli* this role is carried out by CAS6e, also known as CasE (Brouns et al., 2008). The short RNA segments that are released from pre-crRNA processing retain an 8 nt 5′ "handle" sequence from the upstream DR as part of the mature crRNA (Brouns et al., 2008). Processing of pre-crRNA transcripts in *Sulfolobus* has been reported to proceed from the 3′ distal end toward the 5′ leader sequence (Lillestol et al., 2009). It is unclear how this 3–5′ directionality is established, given the site-specific endonucleolytic nature of CAS6 (Carte et al., 2008).

The Type III-B RNA-targeting CRISPR systems have been investigated in *Pyrococcus furiosus* (Hale et al., 2009, 2012) and in *Sulfolobus solfataricus* (Zhang et al., 2012). These systems include the cmr family of CAS genes along with the nearly ubiquitous *cas1*, *cas2*, and *cas6*. The *cmr* complex is composed of the protein products of *cmr1*, *cas10*, and *cmr3–cmr6*, plus the *cas6*-derived crRNA. In *Sulfolobus*, an additional *cmr* component, *cmr7*, joins the complex.

All CRISPR systems examined to date load crRNAs with 5′ OH ends, although the crRNA length and mature state of the 3′ end varies by CRISPR type and by species. We have therefore utilized a cloning strategy that is independent of 5′ end chemistry and partially independent of 3′ end chemistry.

In this study, we show linkage of CAS protein types with families of CRISPR arrays, conservation of CRISPR array elements across the genus, a novel *nurA-csm6-herA* gene cluster associated with *Pyrobaculum* CRISPR arrays, and provide transcriptional support for polarity in crRNA abundance.

METHODS

CULTURE CONDITIONS

P. aerophilum cells were grown anaerobically in media containing 0.5 g/L yeast extract, 1X DSM390 salts, 10 g/L NaCl, 1X DSM 141 trace elements, 0.5 mg/L $Fe(SO_4)_2(NH_4)_2$, pH 6.5, with 10 mM $NaNO_3$. *P. islandicum* and *P. arsenaticum* cells were grown anaerobically in media containing 10 g/L tryptone, 2 g/L yeast extract, 1X DSM390 salts, 1X DSM88 trace elements, and 20 mM $Na_2S_2O_3$. *P. calidifontis* cells were grown aerobically in 1L flasks using 500 ml media containing 10 g/L tryptone, 2 g/L yeast extract, 1X DSM88 trace metals, 15 mM $Na_2S_2O_3$, pH 6.8, loosely capped with moderate shaking at 125 rpm. Anaerobic cultures were grown in 2L flasks with 1L media, prepared under nitrogen with resazurin as a redox indicator at 0.5 mg/L; 0.25 mM Na_2S was added as a reductant. All cultures were grown at 95C to late log or stationary phase, monitored at OD_{600}.

The 10X DSM390 salts are comprised of (per liter ddH_2O) 1.3 g $(NH_4)_2SO_4$, 2.8 g KH_2PO_4, 2.5 g $MgSO_4{\cdot}7H2O$. The 100X DSM88 trace metal solution is comprised (per liter 0.12N HCl), 0.9 mM $MnCl_2$, 4.7 mM Na_2B4O_7, 76 μM $ZnSO_4$, 25 μM $CuCl_2$, 12.4 μM $NaMoO_4$, 18 μM $VOSO_4$, 6 μM $CoSO_4$. The 100X DSM141 trace metal solution is comprised of 7.85 mM Nitrolotriacetic acid, 12.2 mM $MgSO_4$, 2.96 mM $MnSO_4$, 17.1 mM NaCl, 0.36 mM $FeSO_4$, 0.63 mM $CoSO_4$, 0.68 mM $CaCl_2$, 0.63 mM $ZnSO_4$, 40 μM $CuSO_4$, 42 μM $KAl(SO_4)_2$, 0.16 mM H_3BO_3, 41 μM Na_2MoO_4, 0.1 mM $NiCl_2$, 1.14 μM Na_2SeO_3.

cDNA LIBRARY PREPARATION

The cDNA libraries were prepared using small RNA fractions collected from cells grown to stationary and exponential phase, using methods previously described (Bernick et al., 2012), with brief details given in Results. These two preparations were constructed for each of *P. aerophilum, P. islandicum, P. arsenaticum,* and *P. calidifontis* cultures, yielding a total of eight cDNA libraries.

The 3′ end chemistries of crRNA have been reported as either 2–3′ cyclic phosphate (Hale et al., 2012; Jore et al., 2011), or as 3′ OH (Hatoum-Aslan et al., 2011; Zhang et al., 2012). Under the acidic conditions (pH 5) used in RNA preparation in this study, we expect an equilibrium population of 3′ OH terminated RNA to exist under either scenario, providing a cloning method that is semi-independent of 3′ end chemistry.

SEQUENCING AND READ MAPPING

Sequencing was performed using a Roche/454 GS FLX sequencer, and the GS emPCR Kit II (Roche). Sequencing reads in support of this work are provided online via the UCSC Archaeal Genome Browser (http://archaea.ucsc.edu) (Chan et al., 2012).

Reads that included barcodes and sequencing linkers were selected from the raw sequencing data and used to identify reads from each of the eight pooled cDNA libraries. Reads were

further consolidated, combining identical sequences with associated counts for viewing with the Archaeal Genome Browser. Reads were mapped to the appropriate genome [*P. aerophilum* (NC_003364.1); *P. arsenaticum* (NC_009376.1); *P. calidifontis* (NC_009073.1); *P. islandicum* (NC_008701.1); *P. oguniense* (NC_016885.1); *P. neutrophilum* (*T. neutrophilus*: NC_010525.1)] using BLAT (Kent, 2002), requiring a minimum of 90% identity (-minIdentity), a maximal gap of 3 (-maxIntron) and a minimum score (matches minus mismatches) of 16 (-minScore) using alignment parameters for this size range (-tileSize = 8-stepSize = 4). Reads that mapped equally well to multiple positions in the genome were excluded from this study. The remaining, uniquely mapped reads were formatted and visualized as BED tracks within the UCSC Archaeal Genome Browser.

COMPUTATIONAL PREDICTION OF ORTHOLOGOUS GENE CLUSTERS

Computational prediction of orthologous groups was established by computing reciprocal best BLASTP (Altschul et al., 1990) (RBB) protein coding gene-pairs among pairs of four *Pyrobaculum* species. When at least three RBB gene-pairs select the same inter-species gene set (for example A pairs with B, B pairs with C, and C pairs with A), the cluster is considered an orthologous gene cluster.

CRISPR ARRAY MAPPING

Arrays were predicted using CRISPRfinder (Grissa et al., 2007). Arrays were merged in some cases-based on sequencing data evidence.

RESULTS

CRISPR/CAS PROTEIN FAMILIES

Three distinct types of CAS gene clusters exist within the six *Pyrobaculum* species examined (**Figure 1** and **Table A1**) (Makarova et al., 2011). In most *Pyrobaculum* species, the Type I system is present, organized in submodules. Typically we find a submodule that includes: *cas1*, *cas2*, *cas4*, and a *cas4* variant herein referred to as *cas4′*, previously described as *csa1* (Haft et al., 2005) (submodule abbreviation *cas4′-1-2-4*). A second submodule is found nearby, comprising *cas6*, *cas7*, *cas5*, *cas3′*, *cas3″*, and *cas8a2* (abbreviated *cas6-7-5-3′-3″-8a2*) (**Figure 1**). With the exception of *P. islandicum*, each species in the genus has these submodules or close variants, and one or more submodules may be duplicated. In some cases, terminal members of the submodule may be relocated, such as *cas6* in *P. calidifontis* or *P. neutrophilum*. Type I subtypes are defined by the presence of specific genes: *cas8a1* or *cas8a2* (subtype I-A); *cas8b* (subtype I-B); *cas8c* (subtype I-C); *cas10d* (subtype I-D); *cse1* (subtype I-E); and *csy1* (subtype I-F) (Makarova et al., 2011). *P. aerophilum*, *P. oguniense*, and *P. neutrophilum* contain *cas8a2*, so fall within the definitive Type I-A subtype. *P. arsenaticum* and *P. calidifontis* do not appear to contain any recognized signature genes, so the subtype remains indeterminate. Notably, the Type I system is completely absent from *P. islandicum*.

A second CAS group, the Type III-B family of RNA-targeting CAS genes, is present in four *Pyrobaculum* species but not in *P. aerophilum* or *P. islandicum*. Again, this second family is present as submodules, with *cmr4*, *cmr5*, *cmr1*, and *cmr6* (*cmr4-5-1-6*) adjacent but on the opposite strand of the *cmr3-cas10* submodule. One or both of these submodules include *csx1*, and are currently classified as members of Type III-U (unclassified Type III). We find that *csx1* also appears in the Type I modules, so this suggests a broader role for *csx1* among *Pyrobaculum* CAS modules.

The third kind of module found in the genus, Type III-A (*csm*), appears to be complete in *P. aerophilum*, and is the only apparent CAS family found in *P. islandicum*. Previously, Makarova suggested that CRISPR adaptation for Type III families may require use of *cas1* and *cas2* in *trans* from a resident Type I family member (Makarova et al., 2011). However, this option is unavailable in *P. islandicum*, suggesting that adaptation for *Pyrobaculum* Type III systems may not require *cas1-cas2*. Possibly, an undescribed enzyme fulfills this role, or *P. islandicum* may have lost the ability to further adapt its CRISPR arrays.

Curiously, *csm6* is absent from *P. islandicum*, but is present in every other species examined in this study. This is notable because *csm6* would be expected to be part of the Type III-A system in *P. islandicum*, and would not be expected in species that do not encode a complete Type III-A module. Both *P. oguniense* and *P. neutrophilum* encode a portion of the Type III-A module (*csm3-csm5-cas10-csx1*) but both species are missing *csm2* and *csm4*. Where *csm6* is present, it is located next to a conserved paralog of *nurA* and *herA*; these genes are near a CRISPR array in species of *Pyrobaculum*, *Thermoproteus*, and *Vulcanisaeta*, suggesting that this arrangement is widespread among the Thermoproteales.

The *nurA-herA* protein complex is comprised of a 5–3′ DNA exonuclease (*nurA*) and a bidirectional helicase (*herA*) with probable involvement in homologous recombination (HR) (Constantinesco et al., 2004). HR processing requires a 3′ single stranded DNA (ssDNA) resection of chromosomal ends resulting from a double-strand break, and in thermophilic archaea, that resection is carried out by the helicase-nuclease complex of HerA-NurA (Blackwood et al., 2012). In most *Pyrobaculum* spp., there are three or more paralogs of this gene-pair, one of which is clustered with *csm6* and near a CRISPR array (**Figure 1**). Computationally predicted orthologs of the CRISPR-associated *nurA-herA* genes (RBB) show that this pair has been retained throughout the *Pyrobaculum* genus and more broadly among the Thermoproteales (**Figure 2** and **Table A2**). In *P. islandicum*, however, the CRISPR-associated *nurA-herA* pair and *csm6* are absent. We propose that the *nurA-csm6-herA* complex may be associated with adaptation in *Pyrobaculum* species. Three possibilities arise from this proposal: (1) adaptation in *P. islandicum* may have been lost; (2) adaptation in *P. islandicum* may occur using an alternative mechanism, possibly one of the *nurA-herA* paralogs; or (3) the *nurA-csm6-herA* trio may only be required in Type I CRISPR systems (Yosef et al., 2012).

CRISPR ARRAYS

We have characterized three distinct families of CRISPR arrays present among six sequenced *Pyrobaculum* genomes (**Table 1**). These three families are defined by the sequences central to

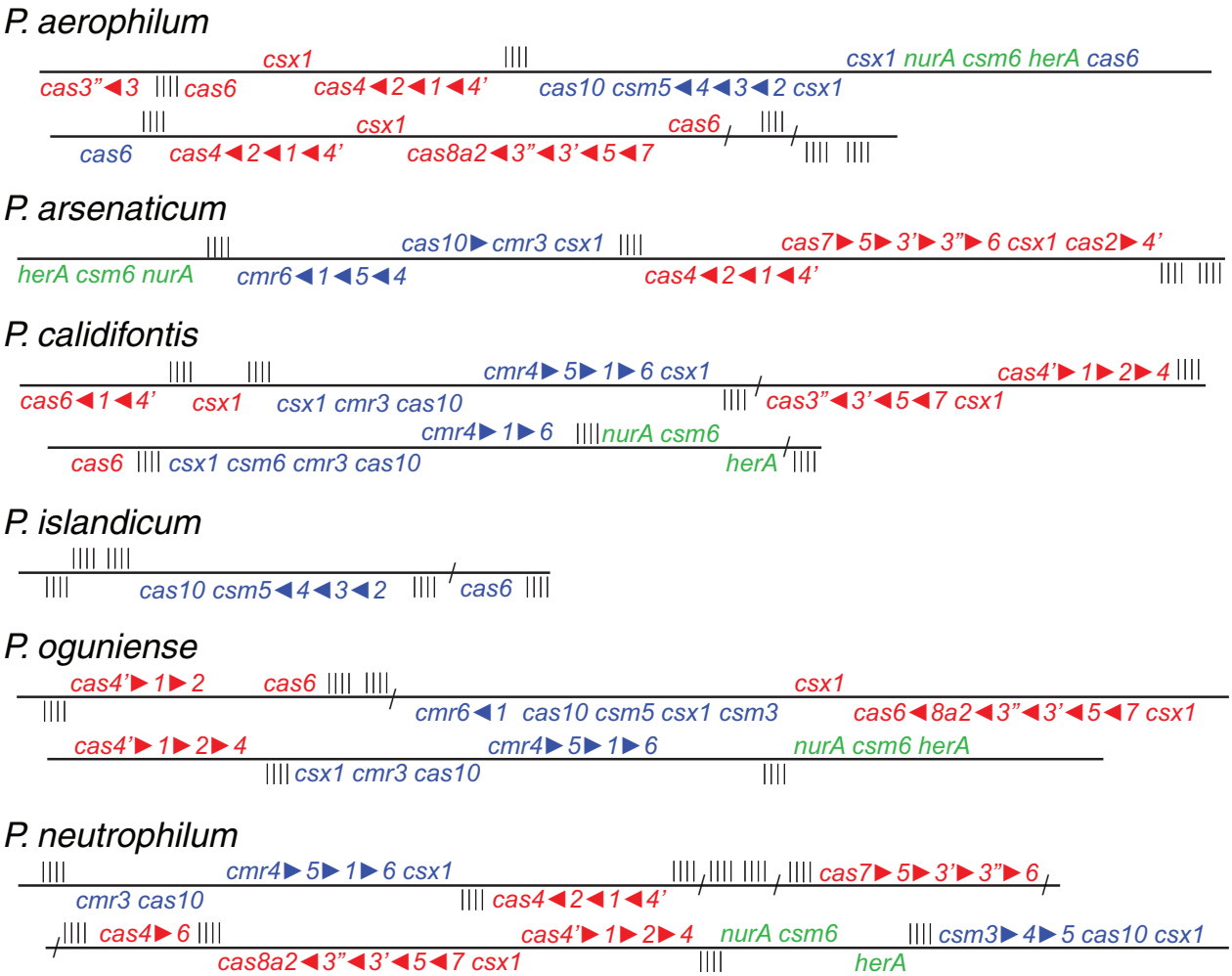


FIGURE 1 | Genomic arrangement of CRISPR modules within *Pyrobaculum* species. Colors indicate Type I CAS modules (red), Type III CAS modules (blue) and CRISPR-associated *nurA-csm6-herA* clusters (green). CRISPR DNA arrays represented as vertical bars. Arrangements of multiple genes in the same CAS family are indicated using filled triangles (*cas1-cas2* is indicated as *cas1*▶*2*). Genomic distances greater than 10 Kb are indicated using diagonal slashes ("/"). Gene strand is indicated relative to the solid black line for positive (above) and negative (below) orientations. Most *Pyrobaculum* species encode both Type I and III CAS modules; *P. islandicum* encodes only a Type III CAS module.

the DR and typically contain an A-rich core of 3–5 nt. These central motifs are flanked by short reverse complement (RC)-palindromes. The DR is terminated by an 8 nt-long sequence that becomes the 5′ handle of the mature crRNAs (Brouns et al., 2008). The various *Pyrobaculum* species encode between four and seven CRISPR arrays within their respective genomes. Except for *P. islandicum*, all species contain one or more representatives of family I and at least one additional representative from family III.

A single array may include multiple families of DR sequences, as found in *crispr1* of *P. oguniense* and *crispr5* of *P. neutrophilum*. In these unusual cases, the DRs are clustered; for example in the *P. neutrophilum* case, the type I DR array begins with 11 repeats using the "AAGTT" core, followed by a set of four repeats mixing "AAAAA" with "AAAGA" cores, and terminating with three "AAAGA" core repeats. In *P. oguniense*, *crispr1* has eight repeats with a 5′ motif of "GTCAAA" and five repeats with a 5′ motif of "CCAGAA." In both cases where DR mixing was observed, the array type (based on CAS proteins) is maintained (**Table 1**). Previous studies in *E. coli* have shown that new DRs are added to an array during adaptation, by copying the first DR in the array (leader-proximal) (Yosef et al., 2012). We note that non-mixed arrays exist in *P. neutrophilum* whose leader-proximal repeats include the "AAAGA" and "AAGTT" cores. Potentially, DR mixing may come about through HR (duplication) events, or possibly by copying a leader-proximal DR from another array during adaptation.

A 5′ promoter-like sequence (AAAAACTTAAAAA) is ultra-conserved with only three single nt polymorphisms among all 37 CRISPR arrays in the six *Pyrobaculum* species studied. The same promoter-like element is also associated with some tRNA genes in these genomes. The sequence variation in the corresponding

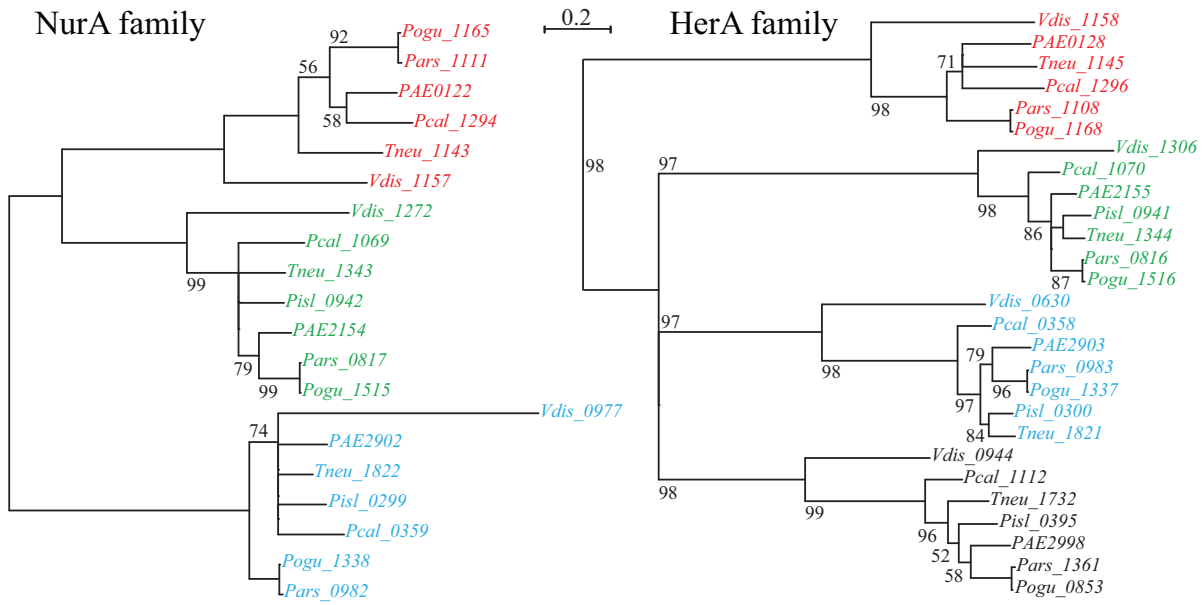


FIGURE 2 | NurA and HerA gene families in *Pyrobaculum* species. *Pyrobaculum* species encode at least three *nurA-herA* gene-pairs, shown above to display parallel phylogenetic structure (left and right panels), with genomically adjacent genes matched by color. Within each gene clade (red, green or blue), the relationship of genes follows the expected species tree, indicating robust ortholog groups. The CRISPR-associated *nurA-herA* pairs (shown in red above) maintain the paired relationship throughout the Thermoproteales. An unmatched paralogous group of *herA* genes appears in black. The *Vulcanisaeta distributa nurA-herA* genes are provided as outgroups of the *Pyrobaculum* genus; note that only the CRISPR-associated *nurA-herA* pair (Vdis_1157-Vdis_1158) is collocated in *V. distributa*, further supporting the tight functional association. Bootstrap percentages are included in black (100% unless otherwise noted). Gene name prefixes: *Vdis—Vulcanisaeta distributa*; *PAE—Pyrobaculum aerophilum*; *Pars—P. arsenaticum*; *Pisl—P. islandicum*; *Pcal—P. calidifontis*; *Pogu—P. oguniense*; *Tneu—P. neutrophilum*. Multiple alignment of amino-acid sequences was performed using Clustal-Omega (v 1.04), maximum-likelihood trees were computed using Tree-puzzle (v 5.02), and the final tree was visualized using newicktops (v 1.0), available through the Pasteur Institute's Mobyle portal (http://mobyle.pasteur.fr/).

Table 1 | *Pyrobaculum* direct repeat (DR) families.

Type	5′ motif	p	core	p′		5′ crRNA handle	*P. aerophilum*	*P. arsenaticum*	*P. islandicum*	*P. neutrophilum*	*P. calidifontis*	*P. oguniense*
I	GAAT	CTC	AAAAA	GAG	G	ATTGAAAG	1	3				2
	GAAT	CTC	AAGAA	GAG	G	ATTGAAAG					4	
	GAAT	CTC	AAAGA	GAG	G	ATTGAAAG				2		
	GAAT	CTC	AAGTT	GAG	G	ATTGAAAG				2*		
	GATT	CTC	AGATA	GAG	A	TTTGAAGG				1		
III-B	GAGAAT	CCCC	AAA	GGGG		GTAGAAAC					3	
III-A	CCAGAA	ATC	AAAA	GAT	A	GTTGAAAC	4					1
III	CCAGAA	ATC	AAAA	GAT	A	GTAGAAAC			5	5		
III-B	GTCAAA	ATC	AAAA	GAT	A	GTTGAAAC		1				1*

Alignment of direct repeats across known Pyrobaculum species. Pyrobaculum DR sequences include a variable length 5′ motif, two short inverted repeats (p and p′) surrounding an A-rich core region, followed by one or zero nucleotides, and ending in what will become the 5′ handle of processed crRNA. Identical motifs are shown in gray below first instance. Numbers in species columns refer to number of CRISPR arrays harboring DRs of that type. Asterisk (∗) indicates DR mixing has occurred in one of the CRISPR arrays in this species. The associated CAS type is inferred by adjacency to an array using that DR family.

promoter elements for other genes is commonly much more diverse. This finding suggests that the invariant CRISPR promoter sequence is maintained either through strong purifying selection or through frequent gene-conversion (Liao, 2000).

CRISPR/CAS protein families appear to be associated with arrays of a given sequence family. This association is upheld to the CAS type, but does not extend to the subtype. For example, in *P. islandicum*, the only CAS family present is Type III-A (**Figure 1**)

and the five encoded arrays in that species use a single DR type (**Table 1**). This same DR is also found in *P. neutrophilum* next to a Type III-B CAS cluster. In a second example, the mixed *crispr1* in *P. oguniense* is made up of DRs associated with Type III-A CAS clusters as found in *P. aerophilum*, and Type III-B CAS clusters, as found in *P. arsenaticum*. Both of these examples demonstrate the association of CAS types (not subtypes) with CRISPR array families in the *Pyrobaculum* genus.

Pre-crRNA transcripts are subjected to endonucleolytic processing to yield individual crRNA sequences, which we detect within small-RNA libraries. Deep sequencing from four *Pyrobaculum* species yielded thousands of sequencing reads, representing between 3% (*P. arsenaticum*) and 20% (*P. islandicum*) of the total sequencing reads in the 20–70 nt size range (**Table 2**).

The abundance of individual crRNAs appears to be related to their position within the array (**Figure 3**). Abundance is generally highest when the spacer is located in the leader-proximal (5′) portion of the array, and decays distally (3′) (**Figure 4**), as seen in *Pyrococcus* (Hale et al., 2012). This pattern is evident in most *Pyrobaculum* arrays that contain more than five spacers. We also see significant variation in crRNA abundance against this decaying background pattern as described for *Sulfolobus* species (Zhang et al., 2012).

Table 2 | CRISPR crRNA abundance (counts) in *Pyrobaculum* species from each CRISPR array, measured during exponential (expo) and stationary (stat) growth phases.

Species	CRISPR id	Type	Size	expo	stat	total
P. aerophilum	*crispr1*	III	13	361	146	507
	crispr2	III	17	342	91	433
	crispr3	I	80	1298	417	1715
	crispr5			degenerate array		
	crispr7/6	III	11	305	101	406
	sum			2306	755	3061
	Total RNA			17,785	13,042	30,827
	crispr%			13.0%	5.8%	9.9%
P. arsenaticum	*crispr2*	I	34	178	339	517
	crispr3	I	84	183	230	413
	crispr4			degenerate array		
	crispr5	III		degenerate array		
	crispr6	I	6	5	10	15
	sum			366	579	945
	Total RNA			14,854	16,352	31,206
	crispr%			2.5%	3.5%	3.0%
P. islandicum	*crispr1*	III	17	691	455	1146
	crispr2	III	14	635	349	984
	crispr3	III	2	627	586	1213
	crispr4	III	3	594	416	1010
	crispr5	III	34	2363	1661	4024
	sum			4910	3467	8377
	Total RNA			28,128	14,823	42,951
	crispr%			17.5%	23.4%	19.5%
P. calidifontis	*crispr1*	III	2	545	340	885
	crispr2	III	3	302	226	528
	crispr3	III	2	156	150	306
	crispr4	I	8	180	85	265
	crispr5	I	35	233	270	503
	crispr6	I	36	274	248	522
	crispr7	I	2	12	13	25
	sum			1702	1332	3034
	Total RNA			22,102	17,192	39,294
	crispr%			7.7%	7.7%	7.7%

The majority of terminal positions of sequencing reads found in *Pyrobaculum* species include an 8-base portion of the upstream DR at the 5′ end (**Figure A1**); this corresponds to the 5′ handle (Brouns et al., 2008) (**Figure 3**). We also see a minority population of sequencing reads that include a 5-base portion of the upstream DR (**Figure A1**), though these are not present in *P. islandicum*.

We tested two models for 3′ maturation considering an upstream DR ruler-mechanism as seen in *Staphylococcus* species (Hatoum-Aslan et al., 2011), and a wrap-around model involving the downstream DR, as described for *Pyrococcus furiosus* (Wang et al., 2011). Because spacer sizes are not uniform in these species, we examined 3′ processing by testing distributions of 3′ end positions as measured from either the upstream DR or the downstream DR, under the assumption that spacer size variation would provide added noise to the incorrect model. Under the ruler-mechanism model, the 3′ distribution of end positions in *P. aerophilum*, *P. arsenaticum,* and *P. calidifontis* includes majority peaks at positions 40–41, and a minority peak at position 32 in *P. aerophilum* (**Figure A2**). Under the downstream DR based wrap-around model (**Figure A3**), *P. aerophilum* has a reduced peak at −25 (corresponding to position 40 in the ruler-mechanism model) and the minority peak is absent (seen previously at position 32). We consider this evidence as consistent with a ruler-mechanism for *P. aerophilum* CRISPR systems. In the remaining species, this analysis was inconclusive.

We find limited evidence for bidirectional CRISPR transcription as reported in *Sulfolobus* (Lillestol et al., 2009). Across all four of the *Pyrobaculum* species in the selected 16–70 nt size range, we see less than 1% of 15,417 CRISPR reads that map to the reverse strand of the array. Where those antisense reads are present, they appear to originate within the spacers and terminate at poly-T motifs within the DR regions. With the limited number of reverse reads seen in this size range, it appears that transcription from the opposite strand is either not processed down to the size range studied, or that reverse transcripts are much less abundant in *Pyrobaculum*. Potentially, this negative finding could be the result of the ubiquitous poly-A sequence present in every DR studied in this genus (**Table 1**). We anticipate that the poly-A sequence could mimic a poly-T terminator on the reverse strand, and thereby prevent significant reverse strand transcription.

DISCUSSION

Within CRISPR arrays, we see an overabundance of reads emanating from the 5′ proximal portion in larger arrays, where transcription from these arrays is likely initiated from a single promoter. The polarity is not perfect given that the abundance of some distal spacers is greater in comparison to more proximal spacer positions. Clearly, there are a number of mechanisms

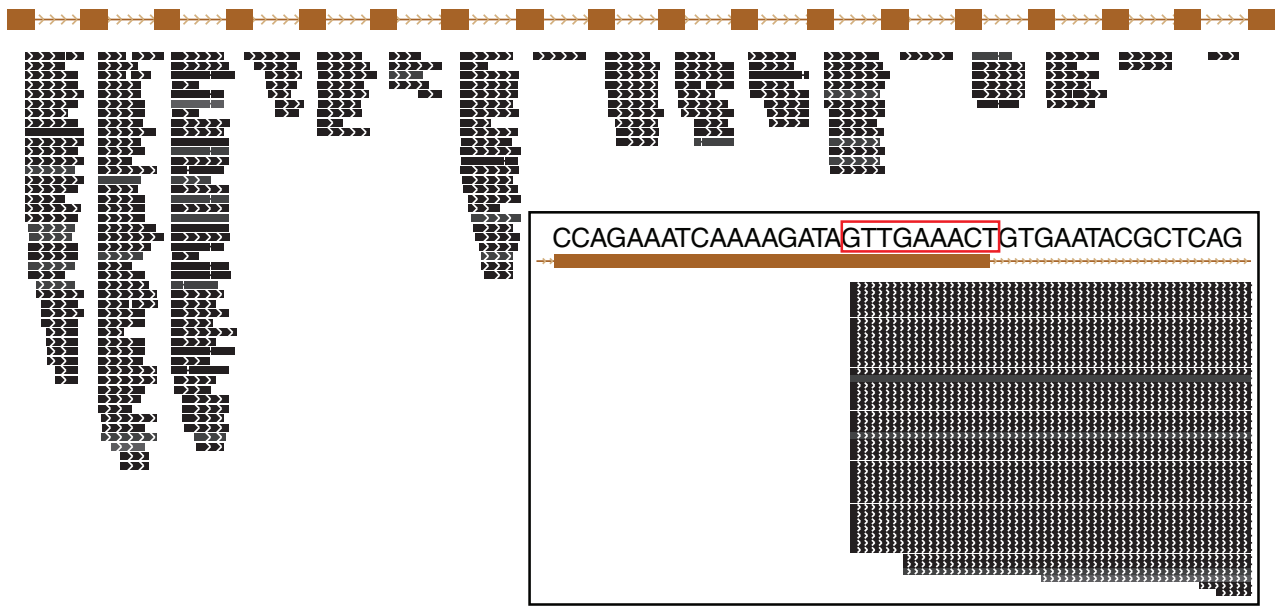


FIGURE 3 | Small RNA expressed from a typical CRISPR array in *Pyrobaculum aerophilum*. The *crispr2* array is shown, depicting the near-identical repeat regions (brown rectangles) with intervening spacer sequences. Mature crRNA reads (black bars) that map uniquely to individual spacers within CRISPR arrays are shown with strand indicated by interior arrowheads. The 5′ ends of crRNA are sharply terminated and includes an 8-base sequence (5′ handle boxed in red) derived from the upstream DR (inset panel). Images generated from the UCSC Archaeal browser (Chan et al., 2012). CRISPR annotation derived from CRISPRFinder (Grissa et al., 2007).

or phenomena that could contribute to crRNA abundance across the array, including: (1) simple stochastic termination of the pre-crRNA transcript, (2) differential efficiency in the endonucleolytic processing of individual crRNAs, (3) transcriptional polarity within the array, (4) differential stability of individual crRNAs, (5) selective recovery and amplification of certain crRNA sequences during library preparation, and (6) recently evolved changes in spacer content (gain or loss or rearrangements) between the reference genome strain and the cultured strains used in our RNA-seq experiments.

It is unknown which or how many of the six possibilities are most relevant, although our data do not equally favor all. If we consider a model of passive, stochastic termination of the primary transcript, we could explain the 5′ polarity but fail to account for the intermediate crRNA variation. Alternatively, a model where individual spacers are matured (excised) from pre-crRNA with varying efficiency might explain the variation in spacer abundance, but the 5′ polarity would be more difficult to accommodate. Instead, we tend toward a model that relies on coupling of pre-crRNA transcription with processing of the transcript, which might explain both polarity and the intermediate variation; for example, if transcription is aborted under conditions of limiting processing capability. We note that some bacterial systems make use of *rho*-mediated termination, coupling transcription and translation in a manner that aborts transcription under conditions of limiting polysomes; this process yields an abundance polarity favoring genes that are near the 5′ end of an operon transcript. Recently, operon polarity has been described in the archaeon *Thermococcus kodakaraensis* (Santangelo et al., 2008). In a polarity model that couples CRISPR pre-crRNA transcription with crRNA processing, we hypothesize that given a limitation in processing by the CRISPR CAScade complex (or *cmr*-processing complex), the pre-crRNA transcript might be prematurely aborted, yielding an abundance of 5′ crRNA. Compelling evidence exists for incremental, endonucleolytic processing of the primary transcript in other species (Brouns et al., 2008; Hale et al., 2008). Under this 5′ polarity model, we would expect to see both polarity as well as a degree of variation in individual spacer abundance, which seems to match our data the closest. This model is necessarily incompatible with 3–5′ directional processing that has been suggested previously (Lillestol et al., 2006).

Within the *Pyrobaculum* genus, one of the conserved *nurA-herA* clusters of syntenic orthologs is always found next to a CRISPR array (**Figure 2**). This cluster includes *csm6*, a gene classified with the Type III-A CRISPR/CAS family. In every case observed, *nurA-csm6* appear to be co-transcribed, in some cases with *herA*. The studied function of *nurA-herA* involves preparation of dsDNA ends as part of HR repair. If these genes participate in CRISPR processing, we suggest that they may be part of new spacer acquisition. That process requires the creation of a new DR and the integration of a novel spacer sequence into an existing array. Generally, this process yields an array with perfect copies, suggesting that the source of the novel DR sequence is an existing array element. In this model, a *nurA-herA* protein complex could provide the HR activity required to repair the array incision.

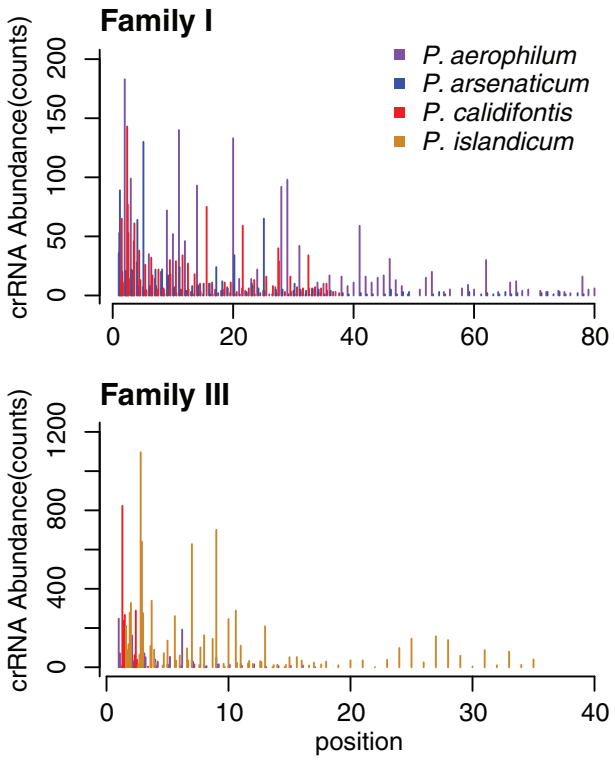


FIGURE 4 | CRISPR small RNA abundance density in *Pyrobaculum* CRISPR arrays by family. Abundance is depicted on the vertical axes, in each species at every spacer position in the array. The horizontal axes plot crRNA genetic origin in terms of spacer number within CRISPR arrays, starting at the 5′ end of arrays. The abundance profile appears higher at the 5′ portion CRISPR arrays, with considerable variation deeper (3′) in the array. Multiple arrays of the same type within a species are plotted as adjacent bars by spacer position.

The phylogeny of the *nurA-herA* orthologous pairs suggests that they have been inherited vertically (**Figure 2**). Furthermore, a parsimonious interpretation of these gene trees indicates that the CRISPR-specific pair predates the divergence of *Pyrobaculum* species, and is well-represented across the Thermoproteaceae. The DR sequences that are in use throughout the *Pyrobaculum* are also remarkably conserved, with only three major sequence variants found, corresponding to the CAS proteins that make use of these structures. The structural conservation of the CAS operons is consistent across the *Pyrobaculum* clade, though not quite as invariant as seen in other archaeal or bacterial models. Finally, we find an ultra-conserved *Pyrobaculum*-specific promoter-like sequence across every CRISPR array examined. Taken together, we infer that the CRISPR system is endemic in the *Pyrobaculum* clade, and is unlikely to have been horizontally acquired through independent events for each of its members.

Cas6 is presumed to be responsible for cleavage of pre-crRNA, and through its association with the Cas complex is likely responsible for the association of Cas protein Types with CRISPR array families. Cas6 is believed to be responsible for recognition and cleavage of pre-crRNA (Hale et al., 2008). In Type I complexes, CAScade (Brouns et al., 2008) and aCAScade (Lintner et al., 2011), Cas6 is a co-purifying member of the complex. In Type III systems where Cas6 does not appear to be part of the Cas complex, specific proteins that are members of the complex are required for maturation of crRNA (Hatoum-Aslan et al., 2011). Furthermore, the binding of Cas6 in *Pseudomonas aeruginosa* has been shown to be quite specific (Sternberg et al., 2012), and in *S. solfataricus*, there are five distinct Cas6 proteins possibly specialized for specific repeats (Zhang et al., 2012). Taken together, we suggest that Cas6 mediates the association between Cas protein families and CRISPR array families in *Pyrobaculum* species. This mediation may be by direct participation in the Cas complex (Type I systems), or through an indirect association as suggested for Type III systems.

Our transcriptional data clearly show that the *P. islandicum* Type III-A system is capable of generating mature crRNA from each of its five arrays. This Type III-A system is operating without *cas1, cas2, or csm6*. In *Pyrococcus abyssi*, the Type I-A system generates crRNA (Phok et al., 2011) and is also missing *cas1* and *cas2*. Possibly one or both of these systems has an alternative enzymatic method for incorporating novel spacers without CAS1, or one or both of these systems may be incapable of CRISPR adaptation. The missing *csm6* in *P. islandicum* is equally surprising given that it has been considered essential in the Type III-A (csm) system, the only system present in this species. Establishing if *P. islandicum* is still capable of CRISPR adaptation could be a first step in identifying an alternative mechanism for spacer incorporation.

The classification system authored by Makarova (Makarova et al., 2011) has been instrumental in coordinating diverse efforts across the field of CRISPR research. As we examine new phylogenetic clades in detail, we have both a convenient mechanism for classifying our findings as well as adding variations brought into focus by new groups. In light of our new analyses, the consolidation of *csa1* (described herein as *cas4′*) with *cas4* may not be justified, as this would suggest many *Pyrobaculum* submodule examples with two copies of *cas4* (*cas4′-cas1-cas2-cas4*). Alternatively, we suggest that the functions of *cas4* and *cas4′* (*csa1*) are distinct in *Pyrobaculum* and should be uniquely classified. Furthermore, we find *csm6* (previously named APE2256) deeply associated with a CRISPR-associated *nurA-herA* pair, and not apparently part of the Type III-A module where it is currently classified. Finally, we observe that the *csx1* classification (part of Type III-U) given to the numerous *Pyrobaculum* genes encoding a DXTHG domain (or MJ1666-like protein) may not be optimal; in *Pyrobaculum*, these genes appear to be found among Type I and III systems. Clearly, the unique comparative perspective afforded by *Pyrobaculum* provides numerous opportunities for future discovery.

AUTHOR CONTRIBUTIONS

David L. Bernick designed and performed the experimental and computational analyses, and wrote the manuscript. Courtney L. Cox provided the analysis of the CRISPR promoter sequence conservation. Patrick P. Dennis provided assistance with the manuscript and collaborative review. Todd M. Lowe provided scientific advising, suggested analyses, and edited the manuscript.

ACKNOWLEDGMENTS

We are grateful to members of the Joint Genome Institute for making 454 sequencing possible (P. Richardson and J. Bristow for providing resources, and E. Lindquist and N. Zvenigorodsky for sample preparation and analysis). This work was supported by National Science Foundation Grant EF-082277055 (Todd M. Lowe and David L. Bernick); the Graduate Research and Education in Adaptive Bio-Technology (GREAT) Training Program sponsored by the University of California Biotechnology Research and Education Program (David L. Bernick); and by the National Science Foundation while Patrick P. Dennis was working at the Foundation. The opinions, findings and conclusions expressed in this publication do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.

Bernick, D. L., Dennis, P. P., Lui, L. M., and Lowe, T. M. (2012). Diversity of antisense and other non-coding RNAs in Archaea revealed by comparative small RNA sequencing in four Pyrobaculum species. *Front. Microbio.* 3:231. doi: 10.3389/fmicb.2012.00231

Blackwood, J. K., Rzechorzek, N. J., Abrams, A. S., Maman, J. D., Pellegrini, L., and Robinson, N. P. (2012). Structural and functional insights into DNA-end processing by the archaeal HerA helicase-NurA nuclease complex. *Nucleic Acids Res.* 40, 3183–3196.

Boehm, T. (2011). Design principles of adaptive immune systems. *Nat. Rev. Immunol.* 11, 307–317.

Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., and Van Der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964.

Carte, J., Wang, R., Li, H., Terns, R. M., and Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 22, 3489–3496.

Chan, P. P., Holmes, A. D., Smith, A. M., Tran, D., and Lowe, T. M. (2012). The UCSC archaeal genome browser: 2012 update. *Nucleic Acids Res.* 40, D646–D652.

Chenchik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov, S., and Siebert, P. D. (1996). Full-length cDNA cloning and determination of mRNA 5′ and 3′ ends by amplification of adaptor-ligated cDNA. *Biotechniques* 21, 526–534.

Constantinesco, F., Forterre, P., Koonin, E. V., Aravind, L., and Elie, C. (2004). A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea. *Nucleic Acids Res.* 32, 1439–1447.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8, 172.

Gudbergsdottir, S., Deng, L., Chen, Z., Jensen, J. V., Jensen, L. R., She, Q., and Garrett, R. A. (2011). Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* 79, 35–49.

Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* 1:e60. doi: 10.1371/journal.pcbi.0010060

Hale, C., Kleppe, K., Terns, R. M., and Terns, M. P. (2008). Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*. *RNA* 14, 2572–2579.

Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V. 3rd, Graveley, B. R., Terns, R. M., and Terns, M. P. (2012). Essential features and rational design of CRISPR RNAs that function with the Cas RAMP module complex to cleave RNAs. *Mol. Cell* 45, 292–302.

Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139, 945–956.

Hatoum-Aslan, A., Maniv, I., and Marraffini, L. A. (2011). Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci. U.S.A.* 108, 21218–21222.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170.

Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575.

Kent, W. J. (2002). BLAT–the BLAST-like alignment tool. *Genome Res.* 12, 656–664.

Jore, M. M., Lundgren, M., Van Duijn, E., Bultema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J., Van Der Oost, J., and Brouns, S. J. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18, 529–536.

Liao, D. (2000). Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *J. Mol. Evol.* 51, 305–317.

Lillestol, R. K., Redder, P., Garrett, R. A., and Brugger, K. (2006). A putative viral defence mechanism in archaeal cells. *Archaea* 2, 59–72.

Lillestol, R. K., Shah, S. A., Brugger, K., Redder, P., Phan, H., Christiansen, J., and Garrett, R. A. (2009). CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol. Microbiol.* 72, 259–272.

Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copie, V., Young, M. J., White, M. F., and Lawrence, C. M. (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* 286, 21643–21656.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., Van Der Oost, J., and Koonin, E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477.

Manica, A., Zebec, Z., Teichmann, D., and Schleper, C. (2011). *In vivo* activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol. Microbiol.* 80, 481–491.

Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322, 1843–1845.

Phok, K., Moisan, A., Rinaldi, D., Brucato, N., Carpousis, A. J., Gaspin, C., and Clouet-D'orval, B. (2011). Identification of CRISPR and riboswitch related RNAs among novel noncoding RNAs of the euryarchaeon Pyrococcus abyssi. *BMC Genomics* 12, 312.

Pul, U., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., and Wagner, R. (2010). Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.* 75, 1495–1512.

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E. L., Eddy, S. R., Bateman, A., and Finn, R. D. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301.

Santangelo, T. J., Cubonova, L., Matsumi, R., Atomi, H., Imanaka, T., and Reeve, J. N. (2008). Polarity in archaeal operon transcription in *Thermococcus kodakaraensis*. *J. Bacteriol.* 190, 2244–2248.

Sternberg, S. H., Haurwitz, R. E., and Doudna, J. A. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* 18, 661–672.

Wang, R., Preamplume, G., Terns, M. P., Terns, R. M., and Li, H. (2011). Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* 19, 257–264.

Wiedenheft, B., Lander, G. C., Zhou, K., Jore, M. M., Brouns, S. J., Van Der Oost, J., Doudna, J. A., and Nogales, E. (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477, 486–489.

Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331–338.

Yosef, I., Goren, M. G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* 40, 5569–5576.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S. V., Naismith, J. H., Spagnolo, L., and White, M. F. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell* 45, 303–313.

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

Received: 30 April 2012; paper pending published: 11 May 2012; accepted: 24 June 2012; published online: 13 July 2012.

Citation: Bernick DL, Cox CL, Dennis PP and Lowe TM (2012) Comparative genomic and transcriptional analyses of CRISPR systems across the genus Pyrobaculum. Front. Microbio. ***3****:251. doi: 10.3389/fmicb.2012.00251*

This article was submitted to Frontiers in Evolutionary and Genomic Microbiology, a specialty of Frontiers in Microbiology.

APPENDIX

Table A1 | Gene reannotations in *Pyrobaculum species*.

Locus	Function	Strand		
Pyrobaculum aerophilum				
PAE0067	*cas3″*	–		
PAE0068	*cas3*	–		
Crispr1		–	39812	40776
PAE0075	*cas6*	–		
PAE0077	*csx1*	+		
PAE0079	*cas4*	–		
PAE0080	*cas2*	–		
PAE0081	*cas1*	–		
PAE0082	*cas4′*	–		
Crispr2		+	45503	46687
PAE0109	*cas10*	–		
PAE0111	*csm5*	–		
PAE0112	*csm4*	–		
PAE0114	*csm3*	–		
PAE0115	*csm2*	–		
PAE0117	*csx1*	–		
PAE0119	*csx1*	+		
PAE0122	*nura*	+		
PAE0124	*csm6*	+		
PAE0126	*csm6*	+		
PAE0128	*hera*	+		
PAE0131	*cas6*	+		
PAE0181	*cas6*	–		
Crispr3		+	95531	101005
PAE0198	*cas4*	–		
PAE0199	*cas2*	–		
PAE0200	*cas1*	–		
PAE0201	*cas4′*	–		
PAE0202	*csx1*	–		
PAE0205	*cas8a2*	–		
PAE0207	*cas3″*	–		
PAE0208	*cas3*	–		
PAE0209	*cas5*	–		
PAE0210	*cas7*	–		
PAE0212	*cas6*	+		
Crispr4	+		268866	269081
Crispr5	–		591745	592220
Crispr6/7	–		1898722	1899654
Pyrobaculum arsenaticum				
Pars_1108	*herA*	–		
Pars_1109/10	*csm6*	–		
Pars_1111	*nurA*	–		
Crispr2		+	999187	1001495
Pars_1114	*cmr6*	–		
Pars_1115	*cmr1*	–		
Pars_1116	*cmr5*	–		
Pars_1117	*cmr4*	–		
Pars_1118	*cas10*	+		
Pars_1119	*cmr3*	+		
Pars_1120	*csx1*	+		
Crispr3		+	1012951	1018930
Pars_1121	*cas4*	–		
Pars_1122	*cas2*	–		
Pars_1123	*cas1*	–		
Pars_1124	*cas4′*	+		
Pars_1127	*cas7*	+		
Pars_1128	*cas5*	+		
Pars_1130	*cas3*	+		
Pars_1131	*cas3″*	+		
Pars_1133	*cas6*	+		
Pars_1134	*csx1*	–		
Pars_1145	*cas2*	–		
Pars_1147	*cas4′*	–		
Crispr5		+	1039190	1039289
Crispr6		–	1307876	1308104
Pyrobaculum calidifontis				
Pcal_0261	*cas6*	–		
Pcal_0263	*cas1*	–		
Pcal_0265	*cas4′*	–		
Crispr1		–	260542	260703
Pcal_0266	*csx1*	–		
Crispr2		+	264904	265204
Pcal_0270	*csx1*	–		
Pcal_0271	*cmr3*	–		
Pcal_0272	*cas10*	–		
Pcal_0273	*cmr4*	+		
Pcal_0274	*cmr5*	+		
Pcal_0275	*cmr1*	+		
Pcal_0276	*cmr6*	+		
Pcal_0277	*csx1*	+		
Crispr3		–	277746	277908
Pcal_1267	*cas3″*	–		
Pcal_1268	*cas3*	–		
Pcal_1270	*cas5*	–		
Pcal_1271	*cas7*	–		
Pcal_1273	*csx1*	–		
Pcal_1274	*cas4′*	+		
Pcal_1275	*cas1*	+		
Pcal_1276	*cas2*	+		
Pcal_1277	*cas4*	+		
Crispr4		–	1185256	1185816
Pcal_1278	*cas6*	–		
Crispr5		–	1188156	1190531
Pcal_1280	*csx1*	–		
Pcal_1281	*csm6*	–		
Pcal_1283	*cmr3*	–		
Pcal_1284	*cas10*	–		
Pcal_1285	*cmr4*	+		
Pcal_1286	*cmr1*	+		
Pcal_1287	*cmr6*	+		
Crispr6		+	1203351	1205855
Pcal_1294	*nurA*	+		

Table A1 | Continued

Locus	Function	Strand		
Pcal_1295	*csm6*	+		
Pcal_1296	*herA*	−		
Crispr7		−	1669194	1669346
Pyrobaculum islandicum				
Crispr1		−	34	1216
Crispr2		+	38866	39842
Crispr3		+	1404032	1404192
Pisl_1541	*cas10*	−		
Pisl_1542	*csm5*	−		
Pisl_1543	*csm4*	−		
Pisl_1544	*csm3*	−		
Pisl_1545	*csm2*	−		
Crispr4		−	1413797	1414026
Pisl_1932	*cas6*	−		
Crispr5		−	1756971	1759456
Pyrobaculum oguniense				
Crispr1			937975	938897
Pogu_1100	*cas4′*	+		
Pogu_1101	*cas1*	+		
Pogu_1102	*cas2*	+		
Pogu_1106	*cas6*	+		
Crispr2			945613	946605
Crispr3			952361	953217
Pogu_1118	*cmr6*	−		
Pogu_1119	*cmr1*	−		
Pogu_1125	*cas10*	−		
Pogu_1126	*csm5*	−		
Pogu_1127	*csx1*	−		
Pogu_1128	*csm3*	−		
Pogu_1135	*csx1*	+		
Pogu_1138	*cas6*	−		
Pogu_1143	*cas8a2*	−		
Pogu_1144	*cas3″*	−		
Pogu_1145	*cas3*	−		
Pogu_1146	*cas5*	−		
Pogu_1147	*cas7*	−		
Pogu_1149	*csx1*	−		
Pogu_1150	*cas4′*	+		
Pogu_1151	*cas1*	+		
Pogu_1152	*cas2*	+		
Pogu_1153	*cas4*	+		
Crispr4			986121	987397
Pogu_1154	*csx1*	−		
Pogu_1155	*cmr3*	−		
Pogu_1156	*cas10*	−		
Pogu_1157	*cmr4*	+		
Pogu_1158	*cmr5*	+		
Pogu_1159	*cmr1*	+		
Pogu_1160	*cmr6*	+		
Crispr5			999562	1002179
Pogu_1165	*nurA*	+		
Pogu_1166/7	*csm6*	+		
Pogu_1168	*herA*	+		
Pyrobaculum neutrophilum				
Crispr1		+	511830	513709
Tneu_0562	*cmr3*	−		
Tneu_0563	*cas10*	−		
Tneu_0564	*cmr4*	+		
Tneu_0565	*cmr5*	+		
Tneu_0566	*cmr1*	+		
Tneu_0567	*cmr6*	+		
Tneu_0572	*csx1*	+		
Crispr2		−	526375	526738
Tneu_0576	*cas4*	−		
Tneu_0577	*cas2*	−		
Tneu_0578	*cas1*	−		
Tneu_0579	*cas4′*	−		
Crispr3		+	530828	531454
Crispr4		+	849844	851759
Crispr5		+	856227	857471
Crispr6		+	883097	885730
Tneu_0994	*cas7*	+		
Tneu_0995	*cas5*	+		
Tneu_0997	*cas3*	+		
Tneu_0998	*cas3″*	+		
Tneu_0999	*cas6*	+		
Crispr7		+	994025	995068
Tneu_1114	*cas4*	+		
Tneu_1128	*cas6*	+		
Crispr8		+	1017598	1019142
Tneu_1132	*cas8a2*	−		
Tneu_1133	*cas3″*	−		
Tneu_1134	*cas3*	−		
Tneu_1135	*cas5*	−		
Tneu_1136	*cas7*	−		
Tneu_1138	*csx1*	−		
Tneu_1139	*cas4′*	+		
Tneu_1140	*cas1*	+		
Tneu_1141	*cas2*	+		
Tneu_1142	*cas4*	+		
Crispr9		−	1030559	1032170
Tneu_1143	*nurA*	+		
Tneu_1144	*csm6*	+		
Tneu_1145	*herA*	−		
Crispr10		+	1035988	1038486
Tneu_1149	*csm3*	+		
Tneu_1150	*csm4*	+		
Tney_1151	*csm5*	+		
Tneu_1152	*cas10*	+		
Tneu_1154	*csx1*	+		

(*Continued*)

Table A2 | NurA and HerA paralogs in *Pyrobaculum* species.

NurAFamily	Pfam Evalue	HerA Family	Blastp Evalue
PAE0122	2.2E-16	PAE0128	7.0E-06
Pcal_1294	7.9E-11	Pcal_1296	4.0E-05
PisI	NA	PisI_	NA
Pars_1111	5.3E-12	Pars_1108	4.0E-05
Pogu_1165	7.3E-11	Pogu_1168	1.0E-05
Tneu_1143	1.4E-09	Tneu_1145	2.0E-05
Vdis_1157	5.0E-07	Vdis_1158	3.0E-05
PAE2154	3.3E-43	PAE2155	9.0E-05
Pcal_1069	1.1E-28	Pcal_1070	1.0E-05
PisI_0942	2.9E-33	PisI_0941	8.0E-06
Pars_0817	2.5E-31	Pars_0816	3.0E-05
Pogu_1515	2.4E-31	Pogu_1516	6.0E-05
Tneu_1343	1.5E-31	Tneu_1344	4.0E-05
Vdis_1272	1.2E-18	Vdis_1306	5.0E-12
PAE2902	1.6E-65	PAE2903	5.0E-40
Pcal_0359	6.1E-47	Pcal_0358	3.0E-38
PisI_0299	1.7E-43	PisI_0300	1.0E-37
Pars_0982	1.2E-45	Pars_0983	4.0E-40
Pogu_1338	2.9E-46	Pogu_1337	5.0E-40
Tneu_1822	1.9E-50	Tneu_1821	2.0E-38
Vdis_0977	5.0E-21	Vdis_0630	1.0E-34
		PAE2998	9.0E-23
		Pcal_1112	4.0E-22
		PisI_0395	1.0E-24
		Pars_1361	4.0E-22
		Pogu_0853	7.0E-22
		Tneu_1732	5.0E-21
		Vdis_0944	3.0E-19

E-values for NurA family paralogs are established using Pfam 26.0 (November 2011) (Punta et al., 2012); E-values for HerA family paralogs are established with Blastp (Altschul et al., 1990), using Sulfalobus solfataricus HerA (SSO2251) as the query and the specific species as the target (wordsize 2, Blosum45 score matrix, Gap existence 13, Gap extension 3). The CRISPR-associated NurA-HerA paralogs are shown in red, and the putative ortholog of NurA-HerA involved in homologous recombination is shown in blue.

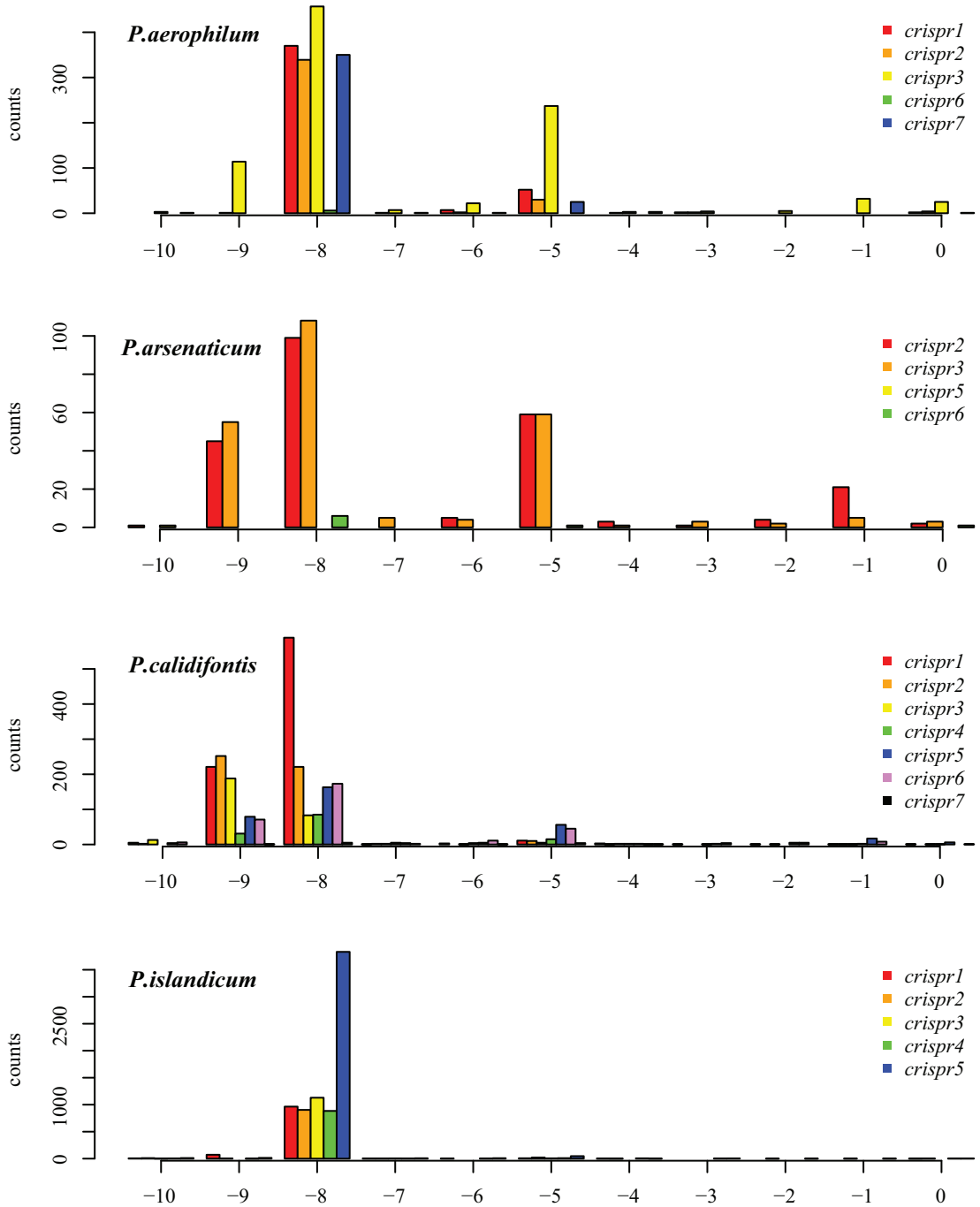


FIGURE A1 | Distribution of mapped 5′ ends of crRNA associated reads within CRISPR arrays in *P. aerophilum, P. arsenaticum, P. calidifontis,* and *P. islandicum*. The majority of transcription sequencing reads begin at position −8 (relative to the beginning of the associated spacer (position 0). This finding implies that most crRNA associated sequencing reads include the 8 nucleotide 5′ handle sequence. A minority population of transcription reads begins at position −5. A third population of sequencing reads begin at position −9; these may be an artifact of the terminal transferase activity of MMLV derived reverse transcriptases. This activity most often yields a terminal cytosine residue to the 3′ end of the cDNA, yielding an implied "G" to the 5′ end of the sequencing read Chenchik et al. (1996).

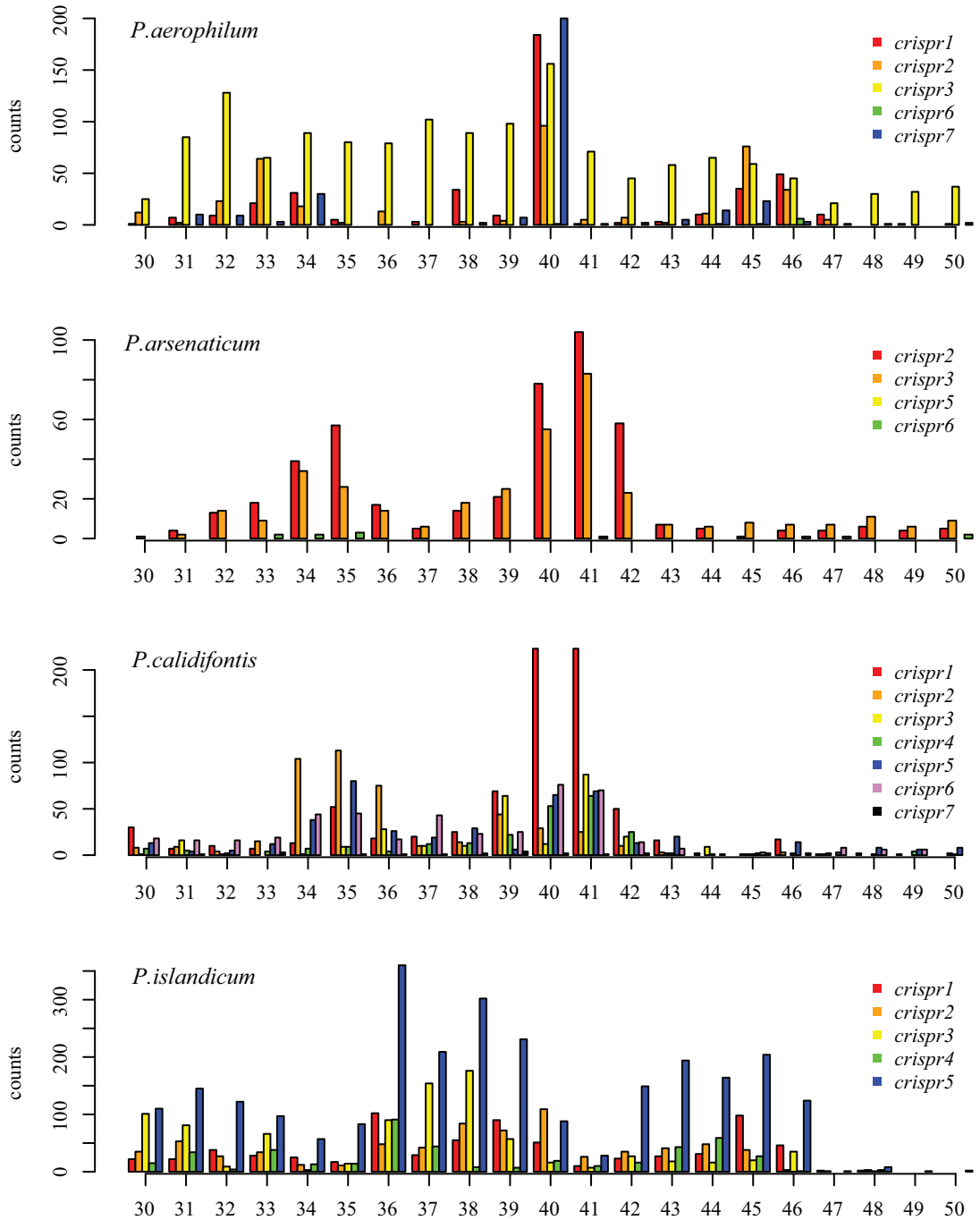


FIGURE A2 | Distribution of mapped 3′ ends of crRNA associated sequencing reads, relative to the 5′ end of the associated spacer. This model proposes that cleavage of the 3′ end of crRNA associated reads utilizes a ruler-mechanism measured from the upstream DR. In *P. aerophilum*, *P. arsenaticum*, and *P. calidifontis*, the majority population of crRNA associated sequencing reads have a 3′ end centered around positions 40–41. A second minority population has a 3′ end centered around positions 32–35. In *P. islandicum*, the majority 3′ end is centered at positions 36–38, and a second minority 3′ end is centered around positions 42–46.

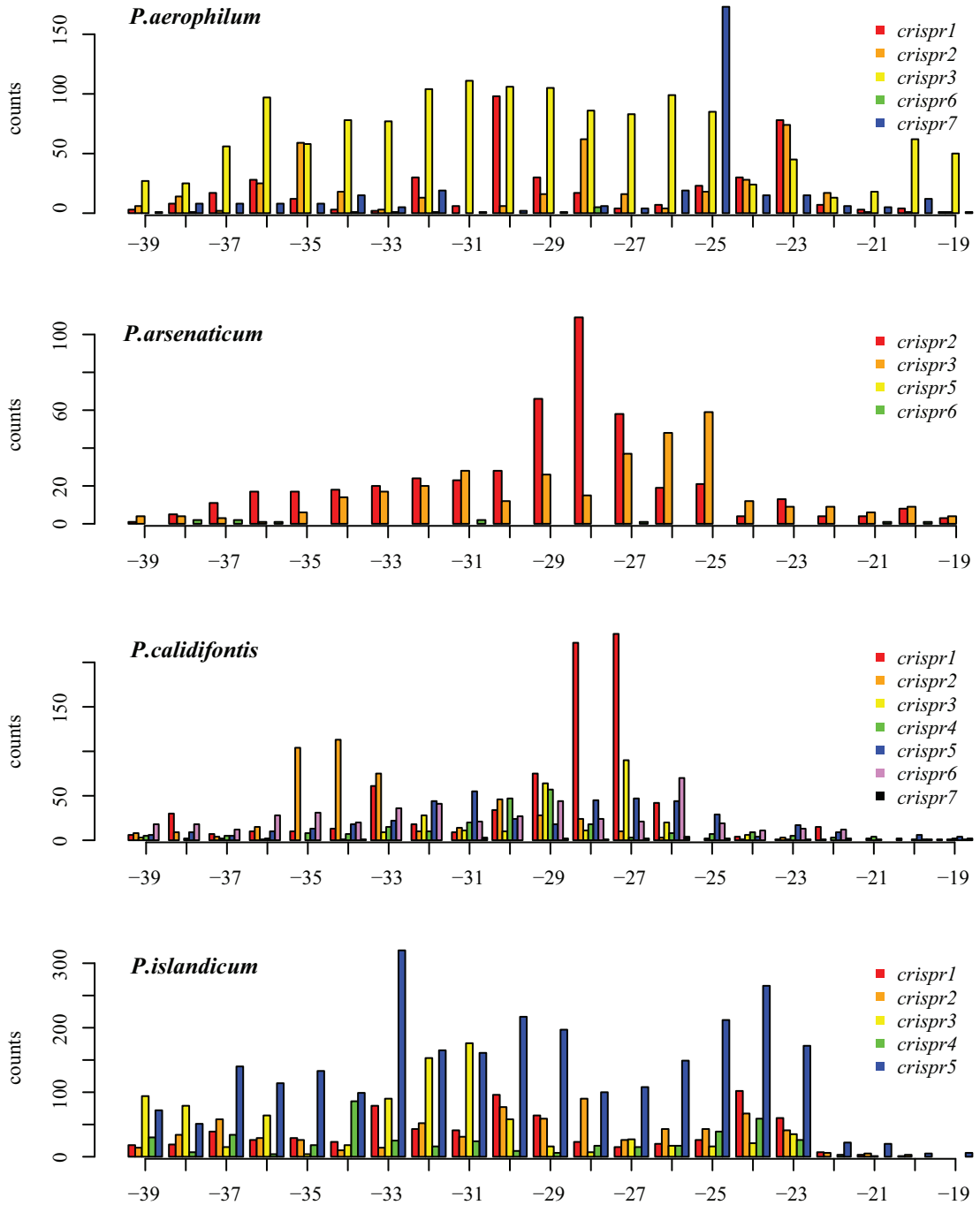


FIGURE A3 | Distribution of mapped 3′ ends of crRNA associated sequencing reads, relative to the downstream Direct repeat. (position 0 is the start of the downstream spacer). This alternative model proposes that the downstream DR establishes the 3′ cut site. In *P. aerophilum*, the major population of 3′ ends shown for crispr 1–3 (**Figure A2**, position 40) is much more diffuse when measured in relation to the down stream direct repeat; this suggests that the 3′ cleavage of crRNA better modeled using the upstream DR as reference rather than the alternative, downstream DR reference. A single spacer region dominates abundance of crRNA in *P. aerophilum*.crispr7; this abundance provides the peak at −25 (corresponding to position 40 in **Figure A2**). In the remaining species, an attempt to distinguish between models of the underlying 3′ cleavage position was inconclusive.