

This Provisional PDF corresponds to the article as it appeared upon acceptance. Fully formatted PDF and full text (HTML) versions will be made available soon.

## Generation of comprehensive transposon insertion mutant library for the model archaeon, *Haloferax volcanii*, and its use for gene discovery

*BMC Biology* 2014, **12**:103 doi:10.1186/s12915-014-0103-3

Saija Kiljunen (saija.kiljunen@utu.fi)  
Maria I Pajunen (maria.pajunen@helsinki.fi)  
Kieran Dilks (kierdilks@gmail.com)  
Stefanie Storf (sstorf@yahoo.de)  
Mecky Pohlschroder (pohlschr@sas.upenn.edu)  
Harri Savilahti (harri.savilahti@utu.fi)

Published online: 09 December 2014

**ISSN** 1741-7007

**Article type** Research article

**Submission date** 29 August 2014

**Acceptance date** 26 November 2014

**Article URL** <http://www.biomedcentral.com/1741-7007/12/103>

Like all articles in BMC journals, this peer-reviewed article can be downloaded, printed and distributed freely for any purposes (see copyright notice below).

Articles in BMC journals are listed in PubMed and archived at PubMed Central.

For information about publishing your research in BMC journals or any BioMed Central journal, go to <http://www.biomedcentral.com/info/authors/>

# Generation of comprehensive transposon insertion mutant library for the model archaeon, *Haloferax volcanii*, and its use for gene discovery

Saija Kiljunen<sup>1</sup>  
Email: saija.kiljunen@utu.fi

Maria I Pajunen<sup>1,3</sup>  
Email: maria.pajunen@helsinki.fi

Kieran Dilks<sup>2</sup>  
Email: kierdilks@gmail.com

Stefanie Storf<sup>2</sup>  
Email: sstorf@yahoo.de

Mecky Pohlschroder<sup>2</sup>  
Email: pohlschr@sas.upenn.edu

Harri Savilahti<sup>1\*</sup>  
\* Corresponding author  
Email: harri.savilahti@utu.fi

<sup>1</sup> Division of Genetics and Physiology, Department of Biology, University of Turku, Turku, Finland

<sup>2</sup> Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

<sup>3</sup> Current address: Department of Biosciences, Division of Biochemistry and Biotechnology, University of Helsinki, Helsinki, Finland

## Abstract

### Background

Archaea share fundamental properties with bacteria and eukaryotes. Yet, they also possess unique attributes, which largely remain poorly characterized. *Haloferax volcanii* is an aerobic, moderately halophilic archaeon that can be grown in defined media. It serves as an excellent archaeal model organism to study the molecular mechanisms of biological processes and cellular responses to changes in the environment. Studies on haloarchaea have been impeded by the lack of efficient genetic screens that would facilitate the identification of protein functions and respective metabolic pathways.

## Results

Here, we devised an insertion mutagenesis strategy that combined Mu *in vitro* DNA transposition and homologous-recombination-based gene targeting in *H. volcanii*. We generated an insertion mutant library, in which the clones contained a single genomic insertion. From the library, we isolated pigmentation-defective and auxotrophic mutants, and the respective insertions pinpointed a number of genes previously known to be involved in carotenoid and amino acid biosynthesis pathways, thus validating the performance of the methodologies used. We also identified mutants that had a transposon insertion in a gene encoding a protein of unknown or putative function, demonstrating that novel roles for non-annotated genes could be assigned.

## Conclusions

We have generated, for the first time, a random genomic insertion mutant library for a halophilic archaeon and used it for efficient gene discovery. The library will facilitate the identification of non-essential genes behind any specific biochemical pathway. It represents a significant step towards achieving a more complete understanding of the unique characteristics of halophilic archaea.

## Keywords

*Haloferax volcanii*, Halophilic archaea, Insertion mutant library, Mu transposition, MuA protein, Gene discovery

## Background

According to the three domains of life paradigm, archaea are distinct life forms and constitute a domain of their own, the other domains being bacteria and eukarya [1,2]. Recent studies suggest that eukaryotes may, in fact, have been originated from ancient archaea [3,4]. Many structural and functional attributes of archaea share a high degree of similarity with the corresponding features present in bacterial or eukaryotic cells. Yet, several aspects of archaeal biochemistry and biology are unique [5]. These include, for example, the ether linkages between glycerol and hydrophobic side chains in archaeal membrane phospholipids [6,7] and the archaeal cell wall, often composed of a single S-layer glycoprotein [8,9]. Moreover, the ability of many archaea to thrive in extreme environments is biologically intriguing and makes archaea a valuable resource for biotechnology applications. Even though a degree of information is available about some unique archaeal features, biochemical pathways and genetic basis behind many of them remain poorly characterized or entirely uncharacterized.

*Haloferax volcanii* is a moderately halophilic, mesophilic, and aerobic archaeon that can be cultured in defined media, and it serves as an excellent model organism for the archaeal domain of life. The genome consists of one major chromosome (2.8 Mb), present in up to thirty copies, and three secondary minichromosomes pHV1 (85 kb), pHV3 (438 kb), and pHV4 (636 kb), with the copy numbers up to 21, 26, and 17, respectively [10,11]. In addition, the characterized wild type strains also harbor a plasmid pHV2 (6.3 kb) [10]. A number of biochemical, genetic, and molecular biology methods have been developed for *H. volcanii*

[12,13], and their successful employment has enabled the identification of key factors involved in several crucial cellular functions, including the components of the two predominant protein transport systems [14,15] and the proteasome [16]. While several types of useable genome manipulation techniques and selectable markers are available for *H. volcanii* [13], its more detailed biological characterization remains critically dependent on the development of a wider selection of advanced methodologies.

Biochemical pathways of archaea can be discovered using comparative genome analyses, where homologous bacterial or eukaryotic genes may indicate a functional archaeal counterpart. However, these types of approaches have only a limited utility in identifying genes behind unique archaeal features. Some archaeal genes may be identified through biochemical means, but the low expression level of many gene products often complicate these analyses. In general, genetics provides the most effective means to discover novel genes, but sophisticated genetics methodologies are yet to be developed for most archaeal species.

A comprehensive collection of mutant clones with randomly distributed genomic alterations provides a valuable resource for studies aimed at delineating molecular mechanisms behind biological functions. Such libraries have proven their immense usefulness, particularly in microbiological research, but only a limited number of archaeal studies have exploited such resources. Chemical and physical mutagens, such as ethyl methanesulfonate (EMS) and ultraviolet light, have been used to introduce random mutations into several archaeal species [17,18]. Although these studies have yielded auxotrophic mutants, mapping of the underlying genetic alterations has been an arduous task. Furthermore, the presence of multiple genome copies within a single cell [11], a common feature of archaea, complicates these types of approaches: mutagenesis produces heterozygous cells with initially only one copy of a mutated allele, and without a selective pressure, newly acquired mutations may be lost or fail to result in detectable phenotypic changes due to the dominance of the wild type alleles [19].

Insertional mutagenesis typically employs selectable markers to tag random genomic loci, and transposition-based methods offer the most straightforward general means for the purpose. Thus far, genome-wide transposon insertion libraries have only been produced for two archaeal species, anaerobic methanogens *Methanosarcina acetivorans* [20] and *Methanococcus maripaludis* [21]. Some progress in transposon mutagenesis has also been made with halophilic *Haloarcula hispanica* [22,23], but comprehensive insertion mutant libraries have not been produced for any haloarchaeal species.

All transposon-mediated mutagenesis approaches employ an efficient transpositional recombination machinery, which catalyzes the joining of the transposon DNA into target DNA, and the most useful transposition systems integrate into a nearly random selection of target sites [24]. The machinery is normally introduced into target cells by the use of plasmid vehicles [25] or via electroporation of pre-assembled DNA transposition complexes, transpososomes [26,27]. The plasmid vehicle strategy requires a functional transposition system, either native or of heterologous origin. However, no native transposon has so far been recruited for mutagenesis in *H. volcanii*, and heterologous systems may not be functional due to the high (1-2 M) intracellular potassium chloride concentration in haloarchaea. The transpososome delivery strategy is critically dependent on electrotransformation under salt-free conditions, and as *H. volcanii* cells are sensitive to a low salt concentration, the utilization of such a strategy may be difficult or impossible. However, alternative approaches may be used, if the organism harbors a homologous recombination

machinery sufficiently operational for efficient gene targeting. In principle, insertion mutant libraries can be generated by the combined use of *in vitro* transposition mutagenesis and simultaneous targeting of a mixture of genomic fragments, as shown with yeast and fungi [28,29].

Previous studies have shown that the *H. volcanii* homologous recombination machinery functions for gene targeting [30-32], and that the MuA-catalyzed *in vitro* DNA transposition reaction can be used to introduce randomly distributed insertions into any target DNA [33]. Encouraged by these achievements, we developed here a strategy to generate an insertion mutant library for *H. volcanii* and generated a broad collection of clones, each individually tagged with transposon DNA. We validated the quality of the library for gene discovery by isolating pigmentation-defective and auxotrophic mutants. The library represents the first comprehensive insertion mutant collection generated in haloarchaea. As the applied strategy is characteristically species non-specific, it should readily be transferable to other archaeal species.

## Results

### Plasmid library with *H. volcanii* inserts tagged with transposon DNA

Initially, we wanted to generate a plasmid library, which would contain transposon-tagged *H. volcanii* genomic DNA segments as inserts, altogether covering the entire genome as overlapping fragments (Figure 1). For the tagging we made a shuttle transposon, TrpA-cat-Mu, which includes two marker genes, *cat* and *trpA*, allowing selection in *Escherichia coli* and *H. volcanii*, respectively (Figure 1A). Initially, *H. volcanii* genomic DNA was partially digested, in separate reactions, with three restriction enzymes (HpaII, AciI, TaqI), each recognizing a 4-bp sequence and generating a protruding 5'-GC. TrpA-cat-Mu was then integrated into the generated fragments using *in vitro* transposition reaction, and the fragments were pooled. Following chromatographic size-selection, 4-6 kb fragments were cloned into a unique ClaI site of pSuperscript SK+, ultimately yielding a plasmid library with 28,000 independently generated member clones (Figure 1B). As *H. volcanii* harbors approx. 4,130 genes [10], on the average, each transposon-tagged gene should be present roughly seven times in this plasmid library.

---

**Figure 1 Transposon TrpA-cat-Mu and overview of the transposon mutagenesis strategy.** (A) The transposon was made from Cat-Mu [33] by the addition of a gene cassette containing *trpA* and *Pfdx*. Orange rectangles indicate 50 bp of Mu R-end DNA. The *trpA* gene (pink arrow) under the control of *Pfdx* promoter (red arrow) was used for the selection in *H. volcanii*; the *cat* gene (green arrow) under the *cat* promoter (blue arrow) was used for the selection in *E. coli*. The transposon was released from its carrier plasmid by BglII digestion. (B) *H. volcanii* H53 DNA was partially digested with HpaII, AciI, and TaqI and used as a target in an *in vitro* Mu transposition reaction with TrpA-cat-Mu as a donor DNA. From transposition products, 4-6 kb fragments were isolated and cloned into the ClaI site of pBlueSript SK+ to yield a plasmid library. Inserts were released by digesting with XhoI + HindIII or KpnI + EcoRV, gel-purified, and used to transform *H. volcanii* H295 cells to generate a transposon insertion mutant library.

---

To validate the plasmid library, we isolated plasmid DNA from three randomly selected clones. The *H. volcanii* DNA segment present in each of these plasmids (pSKT10, pSKT11

and pSKT12) was identified by sequencing. Plasmid pSKT10 harbored transposon-tagged DNA from the minor chromosome pHV3, whereas pSKT11 and pSKT12 contained tagged DNA from the main chromosome (Table 1). These data and the applied cloning strategy with strong biochemical and biological selections imply that each clone in the library includes transposon-tagged *H. volcanii* DNA.

**Table 1 Plasmids used in the work**

Plasmid	Relevant feature <sup>#</sup>	Reference/source
pBlueScript SK+	<i>E. coli</i> cloning vector; Amp <sup>r</sup>	Stratagene
pUC19	<i>E. coli</i> cloning vector; Amp <sup>r</sup>	[34]
pTA231	<i>E. coli</i> – <i>H. volcanii</i> shuttle vector; Amp <sup>r</sup> , Trp <sup>+</sup>	[35]
pTA351	<i>E. coli</i> – <i>H. volcanii</i> shuttle vector; Amp <sup>r</sup> , Trp <sup>+</sup>	[36]
pMPH20	Carrier plasmid of transposon TrpA-cat-Mu; transposon released by BglIII digestion. Amp <sup>r</sup> , Cam <sup>r</sup> , Trp <sup>+</sup>	This work
pSKT10	H295 pHV3 bases 228, 564 – 230, 899* cloned in pBlueScript SK+. TrpA-cat-Mu inserted between HVO_B0199 and HVO_B0200. Amp <sup>r</sup> , Cam <sup>r</sup> , Trp <sup>+</sup>	This work
pSKT11	H295 chromosomal DNA bases 385, 714 – 390, 107* cloned in pBlueScript SK+. TrpA-cat-Mu inserted in HVO_0434. Amp <sup>r</sup> , Cam <sup>r</sup> , Trp <sup>+</sup>	This work
pSKT12	H295 chromosomal DNA bases 1, 530, 448 – 1, 534, 708* cloned in pBlueScript SK+. TrpA-cat-Mu inserted in HVO_1650. Amp <sup>r</sup> , Cam <sup>r</sup> , Trp <sup>+</sup>	This work

<sup>#</sup>Amp: ampicillin, Cam: chloramphenicol, \*Numbering according to *H. volcanii* D2 pHV3 (NC\_013964.1) and genomic sequence (NC\_013967.1), respectively.

## Validation of gene targeting procedure

Our initial targeting attempts with circular plasmid library DNA were problematic, yielding *H. volcanii* clones with no genomic transposon insertions but harboring replicative plasmids (data not shown, see Discussion). Therefore, we decided to use linear DNA fragments for targeting *en masse* (Figure 1B). Prior to the actual experiments, we wanted to validate the targeting procedure using known individual fragments. Accordingly, we isolated and gel-purified the inserts from the abovementioned plasmids, pSKT10, pSKT11, and pSKT12, using XhoI + HindIII double digestion. These fragments, named iSKT10, iSKT11, and iSKT12, were then used to transform *H. volcanii* H295 (Table 2) for gene targeting. This strain was chosen, as it portrays an elevated level of homologous recombination activity [37]. As dam methylation potentially influenced the outcome, we used both methylated and unmethylated DNA. The replicative plasmid pTA231 served as a control for the transformation of circular DNA (Table 3). While dam-methylation reduced the transformation efficiency of circular DNA by tenfold, the methylation status, surprisingly, did not influence the number of produced colonies in transformations with linear DNA. The efficiency of transformation varied among fragments, being 10<sup>4</sup>-10<sup>5</sup> CFU (colony-forming units)/μg DNA with iSKT11 and iSKT12 but only 10<sup>2</sup> CFU/μg DNA with iSKT10. These results likely directly reflect the differences in the length of the homology regions flanking the transposon DNA in these fragments (Table 3). The novel finding that methylation status had no effect on the transformation efficiency of linear fragments demonstrated that there is no need to passage the DNA to be transformed via a Dam<sup>-</sup> bacterial strain. Accordingly, methylated DNA isolated from *E. coli* DH5α was used for further experiments.

**Table 2 Microbial strains used in the work**

Strain	Genotype	Use	Reference/source
<i>H. volcanii</i> H53	<i>pyrE2 trpA</i>	Genomic DNA as target in <i>in vitro</i> mutagenesis	[35]
<i>H. volcanii</i> H295	<i>pyrE2 bgaHa-Kp trpA mre11 rad50</i>	Parental strain for library construction	[37]
<i>E. coli</i> DH10B	F <sup>-</sup> <i>endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 80lacZΔM15 araD139 Δ(ara-leu)7697 mcrA (mrr-hsdRM-mrcBC)</i>	Cloning host	[38]
<i>E. coli</i> MC1061	F <sup>-</sup> <i>araD139 Δ(ara-leu)7696 galE15 galK16 Δ lacX74 rpsL (Str<sup>r</sup>) hsdR2 (r<sub>k</sub><sup>-</sup> m<sub>k</sub><sup>+</sup>)mcrA mcrB1</i>	Cloning host	[39]
<i>E. coli</i> DH5α	F <sup>-</sup> <i>endA1 supE44 thi-1 recA1 relA1 gyrA96 deoR nupG 80lacZΔM15 Δ(lacZYA-argF)U169 hsdR17 (r<sub>k</sub><sup>-</sup> m<sub>k</sub><sup>+</sup>)</i>	Isolation of methylated (Dam <sup>+</sup> ) DNA	Invitrogen
<i>E. coli</i> GM2929	F <sup>-</sup> <i>ara-14 leuB6 thi-1 fhuA31 lacY1 tsx-78 galK2 galT22 glnV44 hisG4 rpsL136 (Str<sup>r</sup>) xyl-5 mtl-1 dam13::Tn9 (Cam<sup>r</sup>) dcm-6 mcrB1 hsdR2 (r<sub>k</sub><sup>-</sup> m<sub>k</sub><sup>+</sup>) mcrA recF143</i>	Isolation of unmethylated (Dam <sup>-</sup> ) DNA	New England Biolabs

**Table 3 Efficiency of H295 transformation<sup>#</sup>**

DNA	Right flank (bp)	Left flank (bp)	Dam <sup>+</sup> *	Dam <sup>-</sup> *
H <sub>2</sub> O	n.a. <sup>n</sup>	n.a. <sup>n</sup>	0	0
pTA231	n.a. <sup>n</sup>	n.a. <sup>n</sup>	1.3 x 10 <sup>4</sup>	1.6 x 10 <sup>5</sup>
iSKT10	2170	166	250	100
iSKT11	3074	1319	1.2 x 10 <sup>5</sup>	1.1 x 10 <sup>5</sup>
iSKT12	1326	2930	9.8 x 10 <sup>4</sup>	2.3 x 10 <sup>4</sup>

<sup>#</sup>Results from one representative experiment are shown.

\*Efficiency in CFU/μg DNA.

n.a. = not applicable.

PCR was then used to verify proper targeting. Initially, genomic DNA was isolated from four clones obtained from each of the three transformations (H295/iSKT10, H295/iSKT11, H295/iSKT12), and DNA was amplified from the corresponding loci (Figure 2). Properly targeted clones were expected to generate a PCR fragment with a 2.2 kb size increase (TrpA-cat-Mu transposon length) when compared to the corresponding fragment from the parental strain H295. Diagnosing successful targeting, the clones from transformations H295/iSKT11 and H295/iSKT12 produced prominent 6.9 kb and 6.7 kb fragments, respectively (Figure 2AB). However, most clones also portrayed faint parental-size fragments, likely indicating the presence of a small number of non-mutated chromosome copies in the studied clonal cell populations. The results with clones from H295/iSKT10 transformation were less uniform (Figure 2C), as a prominent diagnostic 4.8 kb fragment was produced by one of the clones. The other three clones portrayed a major parental fragment of 2.6 kb, although a faint diagnostic fragment was also visible with two of them. Thus, it seemed that the distribution of mutated versus parental minichromosomes was not uniform among the clones studied. This finding was somewhat surprising, as all of these clones grew well without tryptophan, indicating the obligate presence of the *trpA* selection marker. We hypothesized that correctly targeted minichromosomes nevertheless were present in all of these clones studied, but with a very low copy number when compared with the copy number of the wild type pHV3. To test this hypothesis, we devised a PCR strategy that amplified DNA only from a properly targeted minichromosome (Figure 2D). All of the four clones studied exhibited a diagnostic 2.3 kb fragment, indicating proper targeting. Taken together, the data from these experiments showed that linear selection-marker-containing genomic fragments can be targeted in a single step into specific loci, not only in the main chromosome but also in the minichromosome. The results also showed that a mixture of mutated and non-mutated chromosomes was

present in cells even after the selection regime applied, and this was particularly evident with the minichromosome pHV3. We suspect these findings directly reflect the multicopy nature of *H. volcanii* chromosomes and processes involved in the segregation and homogenization of its genome.

---

**Figure 2 PCR analysis of transposon insertions.** Linear fragments containing transposon TrpA-cat-Mu and flanking *H. volcanii* DNA were released from plasmid vectors by restriction digestion and transformed into H295 cells. Four clones from three transformations were isolated, their DNA was extracted, and PCR was used to detect the presence of the inserted transposon. (A) H295/iSKT11 clones were analyzed with primers HSP746 and HSP747. (B) H295/iSKT12 clones were analyzed with primers HSP744 and HSP745. (C) H295/iSKT10 clones were analyzed with primers HSP742 and HSP743 as well as (D) HSP742 and HSP750. White rectangles indicate the 2.2 kb transposon. Flanking DNA from the integrated fragments is shown with purple. The primer recognition sites are shown with arrows. H295 indicates control PCR with DNA from the parental strain, Ctrl indicates template-free control reactions. 1 kb DNA ladder (Invitrogen) was used as a size standard.

---

### ***H. volcanii* mutant libraries with genomic insertions**

To generate *H. volcanii* insertion mutants, a 4 - 6 kb fragment mixture was released from the plasmid library using restriction enzymes, and the mixture was used to transform H295 cells for chromosomal targeting *en masse*, in which each fragment would target its counterpart genomic locus (Figure 1B). In order to avoid a bias caused by restriction enzyme recognition site distribution, two enzyme pairs were used. Fragments released with XhoI + HindIII and KpnI + EcoRV double digestions both produced transformants with the efficiency of  $\sim 2 \times 10^4$  CFU/ $\mu$ g DNA. For these two sets of fragments, we performed two transformations each, in total yielding  $\sim 4 \times 10^4$  (XhoI + HindIII) and  $\sim 3 \times 10^4$  (KpnI + EcoRV) colonies. To detect the presence and copy number of the transposon DNA, 12 of these clones were then studied by Southern hybridization using a transposon-specific probe (Figure 3). A single fragment was detected with all of the clones, indicating that at least the majority of the library clones harbored the transposon as a single copy. These data are fully consistent with proper targeting via homologous recombination.

---

**Figure 3 Southern analysis of transposon insertion mutants.** Genomic DNA (2.5  $\mu$ g) was digested with EcoRI + KpnI and probed with a transposon-specific probe. Purified TrpA-cat-Mu transposon (15.5 ng) and parental H295 DNA were used as positive and negative controls, respectively. Lanes 1-2 represent pigmentation mutants, lanes 3-4 auxotrophic mutants, and lanes 5-12 randomly selected uncharacterized mutants. Faint bands in lanes 11-12 are due to a smaller amount of DNA in these two samples.

---

### **Pigmentation-defective mutants**

*H. volcanii* synthesizes a distinct set of carotenoids [40], resulting in pink colonies and providing a convenient phenotype to validate a mutant library for gene discovery. Accordingly, we screened the above two libraries for mutants deficient in color formation (Additional file 1). A total of  $3.4 \times 10^4$  colonies were visually inspected, and 18 whitish clones were discovered. To identify the genes affected, we determined by sequencing the transposon integration site in 15 of these clones. In each of them, the transposon was located in the *crtB* gene (HVO\_2524). *CrtB* encodes phytoene synthase, which catalyzes the first specific step in the carotenoid synthesis pathway [41,42], logically explaining the observed

phenotype and verifying that the library represents a potent tool for gene identification. Four different insertion sites in *crtB* were identified (Table 4), the number falling within the expectancy range of seven hits per gene on the average. The recovery of several identical sites was not surprising, as replicated sibling plasmids were necessarily present in the plasmid library prepared originally from 28,000 *E. coli* colonies.

**Table 4 Pigmentation-deficient mutants**

Mutant	Integration site; orientation <sup>#</sup>	Gene	Protein
W-20	2,389,801; +	HVO_2524, <i>crtB</i>	Phytoene synthase
W-5, W-17	2,389,806; -	HVO_2524, <i>crtB</i>	Phytoene synthase
W-6, W-7, W-14	2,390,154; +	HVO_2524, <i>crtB</i>	Phytoene synthase
W-4, W-8, W-9, W-10, W-12, W-13, W-15, W-16, W-19	2,390,422; -	HVO_2524, <i>crtB</i>	Phytoene synthase

<sup>#</sup>The first change compared to *H. volcanii* chromosomal sequence (NC\_013967.1) is indicated as the transposon integration site. The orientation is denoted as *trpA* direction compared to *H. volcanii* chromosome (5' to 3'): +, same direction; -, opposite direction.

## Auxotrophic mutants

We screened a total of 4,100 mutants for auxotrophy by initially testing their ability to grow on Hv-Min minimal medium plates. Sixteen mutants did not grow, and their growth capacity was further studied on Hv-Min plates each supplemented with a single amino acid (Table 5). All biologically relevant amino acids were tested, and auxotrophic mutants were found for arginine (5 clones), histidine (2 clones), proline (2 clones), tyrosine (1 clone), isoleucine (1 clone), and phenylalanine (1 clone). Four of the sixteen mutants did not grow with any of these single amino acid supplements.

**Table 5 Amino acid auxotrophic mutants**

Mutant	Auxotrophy	Integration site; orientation <sup>#</sup>	Gene	Protein
H295/6-78	Arginine	43,942; +	HVO_0044, <i>argB</i>	Acetylglutamate kinase
H295/6-100	Arginine	44,698; +	HVO_0045, <i>argC</i>	N-acetyl-gamma-glutamyl-phosphate reductase
H295/12-27	Arginine	44,905; -	HVO_0045, <i>argC</i>	N-acetyl-gamma-glutamyl-phosphate reductase
H295/10-36	Arginine	43,144; +	HVO_0043, <i>argD</i>	Acetylornithine aminotransferase
H295/17-48	Arginine	2,374,426; -	HVO_2508, <i>carA</i>	Carbamoyl-phosphate synthase small subunit
H295/13-54	Histidine	1,198,248; +	HVO_0431	HAD-superfamily hydrolase
H295/12-18	Histidine	1,198,248; +	HVO_0431	HAD-superfamily hydrolase
H295/21-77	Proline	786,914; +	HVO_0869	Glutamate synthase [NADPH] large chain
H295/24-65	Proline	1,051,586; -	HVO_1153	Hypothetical protein
H295/12-1	Tyrosine	1,198,243; +	HVO_1312	Prephenate dehydrogenase
H295/16-1	Isoleucine	578,651; -	HVO_0644, <i>leuA1</i>	2-isopropylmalate synthase / (R)-citramalate synthase
H295/24-63	Phenylalanine	399,114; -	HVO_0449, <i>pheA1</i>	Prephenate dehydratase
H295/9-8	Valine, Leucine, Isoleucine	1,374,516; -	HVO_1506, <i>ilvC</i>	Ketol-acid reductoisomerase
H295/4-26	Valine, Leucine, Isoleucine	1,375,827; +	HVO_1508, <i>ilvB1</i>	Acetolactate synthase large subunit
H295/12-44	Glycine, Serine, Threonine	2,805,618; +	HVO_2969, <i>thrC3</i>	Threonine synthase

<sup>#</sup>The first change compared to *H. volcanii* chromosomal sequence (NC\_013967.1) is indicated as the transposon integration site. The orientation is denoted as *trpA* direction compared to *H. volcanii* chromosome (5' to 3'): +, same direction; -, opposite direction.

The transposon insertion sites in the growth-defective mutants were then determined (Table 5). All the five arginine auxotrophs had the transposon inserted in the genes known to be involved in arginine biosynthesis: *argB*, *argC*, *argD*, and *carA*, of which *argC* was targeted at two different sites. The two histidine auxotrophs were identical with the transposon inserted in the same position in HVO\_0431 (similarity to HAD-superfamily hydrolases). One proline auxotroph had the insertion in HVO\_0869 (glutamate synthase subunit) and the other in HVO\_1153 (hypothetical protein). In the tyrosine auxotroph the transposon interrupted HVO\_1312 (prephenate dehydrogenase, a known enzyme on the tyrosine biosynthesis pathway). The isoleucine auxotroph harbored the transposon in *leuA1*, proposed to encode either 2-isopropylmalate synthase or highly similar (R)-citramalate synthase. In the phenylalanine auxotroph the transposon was inserted in *pheA1* (prephenate dehydratase, known to participate in phenylalanine biosynthesis). In summary, the transposon insertion sites pinpointed both genes known to be involved in amino acid biosynthesis and genes with no prior annotation.

Of the four mutants not growing with single amino acid supplements, three harbored the transposon in a gene involved in amino acid biosynthesis. Two of them, *ilvC* (ketol-acid reductoisomerase) and *ilvB1* (acetolactate subunit), are needed for valine, leucine, and isoleucine biosynthesis. Indeed, the addition of the mixture of these three supplements restored the growth of these mutants. The third identified gene, *thrC3* (threonine synthase) is required in glycine, serine, and threonine metabolism, and the addition all of these three amino acids restored the mutant growth. The sequence data from the fourth growth-deficient mutant suggested that a larger scale genome rearrangement might have taken place, possibly due to an initial insertion of two transposons (see Discussion). As the revelation of the exact alteration would have required substantial extra efforts in sequencing, this mutant was not studied further.

## Discussion

More than 100 sequenced archaeal genomes are currently available, and the number is rapidly increasing (UCSC Archaeal Genome Browser [43-45]). Although the sequence data is critically important for gene discovery, it cannot be used to predict archaeal gene or protein functions if homologous counterparts have not been characterized in other organisms. For example, only ~66% of the predicted genes in seven sequenced haloarchaeal genomes was assigned with a hypothetical protein function in a recent study [46], leaving a significant portion of the genomes non-annotated. As simple sequence comparisons cannot reveal genes behind unique archaeal features, experimental approaches are needed to advance our knowledge about the biology of these intriguing organisms. Various types of mutant libraries constitute an important resource for genetics/genomics studies, and those containing randomly distributed genomic insertions have been particularly useful for the identification of gene functions. Transposon-mediated mutagenesis provides a straightforward means to produce insertion libraries and has been widely applied in bacteria and eukaryotes [47-50]. However, the use of insertion mutagenesis has been very limited with archaea [20,21,23]. This is presumably due to the difficulties in methods development caused by the challenging biology of most archaeal species. Furthermore, inadequately characterized native transposon systems [51] and potential incompatibility problems with heterologous systems may have hindered the progress towards the efficient utilization of transposition-based techniques in archaea.

The aim of this study was to develop a robust insertion mutagenesis methodology for *H. volcanii*. We chose to use minimal-component Mu *in vitro* transposition [33,52] for the distribution of insertions in DNA and homologous recombination machinery of *H. volcanii* for genomic integration [30-32]. Mu *in vitro* reaction yields transposition products highly efficiently and with a relatively low target site selectivity [33,53]. In practice, the distribution of integrations is essentially random along any longer segment of DNA. These characteristics have made the Mu reaction ideal for a variety of molecular biology [54-56], protein engineering [57-63], and genomics [27,64-67] applications. Also in this study, the reaction provided a straightforward means to deliver selection cassettes into random positions along the *H. volcanii* chromosomal DNA. In order to maximize the frequency of gene targeting via homologous recombination into *H. volcanii* genome, we chose to use the strain H295, which is devoid of Mre11 and Rad50 proteins. In the wild type strains these proteins form a complex restraining homologous recombination, and in H295 mutant strain this activity is reported to be increased ~100-fold [37]. As the targeting frequency with H295 in our experiments was  $\sim 2 \times 10^4$  CFU/ $\mu$ g DNA, it is likely that the generation of sizeable mutant libraries can be accomplished also with strains exhibiting the wild type level of homologous recombination activity.

In order to produce a high quality insertion mutant library for *H. volcanii*, we paid careful attention to the details of the strategy design and optimized several key steps in the techniques used. (i) Three different restriction endonucleases and partial digestions were used to fragment *H. volcanii* genomic DNA, effectively reducing the bias caused by restriction site distribution and guaranteeing an extensive coverage of the genome with overlapping fragments. (ii) The size-selection following the *in vitro* transposition reaction prior to ligation ensured that 4-6 kb genomic inserts were enriched in the plasmid library. Accordingly, the effective elimination of short fragments guaranteed that practically all of the cloned inserts were of suitable length for gene targeting. (iii) Our preliminary *en masse* targeting trials with circular plasmids resulted in a large fraction of *H. volcanii* clones that had not experienced genomic insertions but contained replicative plasmids harboring the selection cassette used and an *H. volcanii* origin of replication (data not shown). These plasmids evidently represented molecules, in which the transposon was originally integrated in the vicinity of an *H. volcanii* origin of replication, and ultimately resulted in shuttle plasmids capable of replication in both *E. coli* and *H. volcanii*. This prompted us to use linear DNA fragments for targeting and to eliminate even trace amounts of DNA circles from the targeting preparation. As alkaline plasmid preparation methods produce a fraction of collapsed supercoiled plasmid forms, which enter the cells effectively and cannot be digested with restriction endonucleases, we removed these plasmid forms by purifying the plasmid library using cesium chloride gradient centrifugation. In addition, the DNA to be transformed was size-selected prior to transformation using a preparative agarose gel, a technique which also selects against circular plasmid forms. We believe that particularly the circle-elimination procedures were pivotal for the successful generation of a high-quality insertion mutant library.

The plasmid library was made from 28,000 *E. coli* colonies. Theoretically, assuming random transposon insertion into *H. volcanii* DNA, the probability ( $P$ ) at which every gene is targeted at least once can be calculated using the formula  $P = 1 - (1 - x/G)^n$  [68], where  $x$  is the average gene size,  $G$  is the genome size and  $n$  is the number of mutants. The size of *H. volcanii* genome is 4.01 Mb and the number of genes is 4,130 [10]. Thus, the average gene size is 971 bp. With these numbers, the probability of every gene being tagged in the plasmid library was 0.9989, indicating that all of the *H. volcanii* genes were represented in this library with a very high likelihood. Note that these calculations do not take into account the known

coding density of 86% [10]. It should be emphasized that the plasmid library generated contains a full repertoire of both essential and non-essential *H. volcanii* genes. However, the original diversity of the plasmid library could not be maintained in the final *H. volcanii* genomic insertion mutant library. This is because an insertion into any essential gene would result in cell death upon mutant genome homogenization, leading to a situation where the cell would not be able to survive following the loss of wild type chromosome copies. Likewise, it should be remembered that not all of the insertions generate a loss-of-function mutation.

As dam-methylation had earlier been shown to reduce transformation efficiency in *H. volcanii* [69], we used both Dam<sup>+</sup> and Dam<sup>-</sup> DNA in our experiments. The transformation efficiency with plasmid DNA was shown to be dependent on its methylation status, being ten times higher with unmethylated than methylated DNA. This difference is smaller than the 100- or 1000-fold difference described earlier for *H. volcanii* [69,70]. The reason for this apparent discrepancy is not known but may plausibly relate to the use of different *H. volcanii* strains. Intriguingly, the methylation status had no effect on the transformation efficiency with linear fragments. This is a novel finding, as (to the best of our knowledge) comparative transformation experiments with linear DNA fragments have not previously been conducted with *H. volcanii*. These data might indicate that targeting is a relatively rapid process and/or that the homologous recombination machinery of *H. volcanii* protects the incoming DNA from degradation. This conclusion is in line with earlier studies with *H. volcanii*, showing that integration into the chromosome can rescue methylated foreign DNA from being destroyed [70]. An important practical consequence of our findings is that the linear DNA used for transformation can be purified from a variety of sources, and it does not need to be passaged through a Dam<sup>-</sup> microbial strain.

We performed important quality measures to control the library generation process. Initially, we studied the accuracy of the targeting procedure by introducing three separate linear transposon-tagged *H. volcanii* DNA fragments into cells. With each fragment, the transposon was found in the expected locus in the genome, confirming that targeting proceeded accurately via homologous recombination. Thus, linear tagged DNA fragments can be used effectively in a fast, one-step procedure to generate site-specific insertions into *H. volcanii* genome. However, in theory there exists a possibility that the integration of incoming DNA could also follow a non-homologous recombination pathway. On the basis of our data, we believe that such pathways, if operational, do not nevertheless prevail in *H. volcanii*. This is evidenced by our results from the gene discovery experiments: sixteen out of eighteen phenotypically screened insertion sites (in 4 pigmentation-defective and 12 auxotrophic mutants) identified a known gene on a logical biochemical pathway, a result incompatible with a dominating integration via a non-homologous recombination machinery. As the second quality measure, we determined the number of transposon copies present in the genome of twelve insertion mutant library clones. In each case Southern analysis revealed only one resident copy. As the stoichiometry in the transformation of linear fragments should favor single-copy genomic integrations (only one out of ~2000 cells becomes transformed, data not shown), the result was not surprising. Thus, each mutant clone should harbor only one tagged gene, establishing the genomic insertion library as a powerful tool for genetic screens.

To validate the insertion mutant library for gene identification, we initially searched for mutants defective in pigment formation. The *crtB* gene, identified in the search, encodes phytoene synthase (Figure 4). This enzyme, which has been characterized in several bacterial,

archaeal, and plant species, catalyzes the first specific step in the carotenoid biosynthesis pathway, i.e. the formation of phytoene from two molecules of geranylgeranyl pyrophosphate (GGPP) [41,42]. The loss of pigmentation was thus an expected outcome upon the interruption of *crtB*. We were not able to identify insertion mutations in the genes required for the synthesis of GGPP. This molecule is required not only for the carotenoid biosynthesis, but it is also a precursor of isoprenoid lipids, essential components of the archaeal cell membrane [71]. Thus, it is likely that insertions into any of the genes needed for GGPP biosynthesis would result in cell death, and our data are consistent with this supposition. In archaea, the colorless phytoene is converted into the red-colored lycopene by phytoene dehydrogenase [42,72]. In *H. volcanii*, the *crtI* gene (HVO\_2528) has been annotated to encode this enzyme [10]. If this gene is solely responsible for the phytoene hydrogenase activity in *H. volcanii*, we should have pinpointed it with our screen. A closer look at the *H. volcanii* genome annotation revealed HVO\_0817, annotated to encode a flavin-containing amine-oxidoreductase. This enzyme family contains various amine oxidases, also including phytoene dehydrogenases and related enzymes. Homology searches with other halophilic archaea indicated that the deduced protein sequence of HVO\_0817 shares a high degree of similarity with *Haloferax mediterranei* and *Halobacterium salinarum* phytoene dehydrogenases: 88.2 % and 57.2 % identical amino acids with the products of *crtI* (HFX\_0786) and *crtI3* (OE1808F), respectively (Figure 4). Thus, *H. volcanii* HVO\_0817 most likely encodes a phytoene dehydrogenase. Considering that the theoretical likelihood for an average gene not being tagged in our study was as low as 0.0011, the fact that we were unable to pinpoint either HVO\_2528 or HVO\_0817 strongly suggests that these two genes encode overlapping functions and thus complement each other. As similar redundant phytoene dehydrogenase genes are present also in *H. mediterranei* and *H. salinarum* (Figure 4), such a gene arrangement may be a common feature among haloarchaea.

---

**Figure 4 Carotenoid biosynthesis pathway in haloarchaea.** The pathway from geranylgeranyl pyrophosphate to lycopene is shown. The insert portrays the comparison of the deduced phytoene dehydrogenase sequences of *Haloferax volcanii* (NC\_013967.1) to those of *Haloferax mediterranei* (NC\_017941.2) and *Halobacterium salinarum* (NC\_010364.1). The numbers indicate the percentage of identical amino acids between the compared proteins. Pairwise analyses were made by EMBOSS Water [73].

---

The screen for amino acid auxotrophic mutants mostly identified genes known to be involved in amino acid biosynthesis: *argB*, *argC*, *argD* and *carA* for arginine, HVO\_1312 (prephenate dehydrogenase) for tyrosine, *ilvC* and *ilvB1* for valine, leucine, and isoleucine, and *thr3* for glycine, serine, and threonine. Interestingly, the transposon insertion in the isoleucine auxotroph was in the *leuA1* gene, which has been proposed to encode either a 2-isopropylmalate synthase required for leucine biosynthesis, or an (R)-citramalate synthase required for isoleucine biosynthesis [10]. Our results thus support the latter hypothesis. The proline auxotroph (H295/21-77) had the transposon inserted in HVO\_0869, which encodes glutamate synthase [NADPH] large chain. Glutamate synthase is a key enzyme in ammonia assimilation, and the interruption of HVO\_0869 is expected to result in unbalanced glutamate and nitrogen metabolism. More work would be needed to find out why the growth defect of this mutant is restored by the addition of proline, but not by any other amino acid. In addition to these known amino acid biosynthesis genes, we obtained two auxotrophic mutants having insertions in putative or unidentified genes: HVO\_0431 (HAD-superfamily hydrolase) for histidine and HVO\_1153 (hypothetical protein) for proline. Our study thus proposes the involvement of these genes in amino acid biosynthesis pathways, illustrating the huge potential of the library for the discovery of novel genes.

The copy number of *H. volcanii* major chromosome is dependent on growth phase and conditions, being as high as 30, while the minichromosome copy numbers are somewhat lower (~20) [11]. A study with *H. mediterranei* minichromosome pHM300 suggests that the replication of the major and minor chromosomes is controlled independently [74], and we suspect that this may be the case also in *H. volcanii*. In transposition mutagenesis, the transposon is integrated into DNA of one chromosomal copy, while others remain wild type. Therefore, during the first few cell generations following mutagenesis, the transposon-containing genome copies are necessarily in a minority. As the mixture of mutated and non-mutated chromosomes may persist in a cell for some time, the cells thus being heterozygous, it may be intuitively somewhat surprising that loss-of function mutants could be isolated in our study relatively easily.

Our results suggest that chromosome homogenization takes place under the selection pressure applied, and the PCR data on the targeting into the major chromosome supports this conclusion. The findings are consistent with the results of an earlier *H. volcanii* study, showing that genome copies will be equalized under selection pressure relatively rapidly [19]. In the paper the authors suggested an efficient gene conversion machinery as a primary driving force behind the phenomenon, although mutually non-exclusive uneven segregation could not entirely be ruled out. Our pHV3 targeting data showed variation in chromosome homogenization, possibly indicating that this process may be slow or inefficient with minichromosomes.

There are a few important issues that need to be considered when generating and using mutant libraries for microbes as described in this study: (i) Two copies of the transposon may be integrated *in vitro* into the same target DNA fragment, although rarely, as the reaction stoichiometry favors single insertions. If such a fragment passes the size selection applied, it may experience recombination between the transposon copies. This can occur either in *E. coli* or in the target organism, and it could ultimately lead to a deletion or inversion, both relatively uninformative with regard to gene discovery. (ii) The mutagenesis strategy produces siblings in the plasmid library. Thus, following transformation into the target organism, the recovery of identical genomic insertion clones is expected, which was particularly well illustrated with our *crtB* mutants. Therefore, a large number of mutants with a desired phenotype should be collected in order to guarantee the finding of all possible insertion sites. (iii) The target genome may contain regions that can not be cloned into a plasmid vector, and such regions will not be recovered as transposon-tagged clones in the final insertion library. (iv) The possibility of polar effects should be remembered when working with insertion mutant libraries of microbes. (v) Although rarely, spontaneous mutations may generate false positive results, in which case the transposon tag will identify a gene not involved in the observed phenotype. (vi) The genetic screens should always be regarded a first step towards gene discovery. In order to verify the genotype-phenotype connection, complementation or site-directed mutagenesis studies should be executed with each potential candidate gene.

## Conclusions

To conclude, we present here the generation of the first genome-wide insertion mutant library for haloarchaea using the model archaeon *H. volcanii*. The library includes clones that harbor a single transposon insertion in their genome, and it can be used for gene identification in *H. volcanii*. The strategy used provides a straightforward general means to generate comprehensive mutant libraries, is characteristically species non-specific, and should readily

be transferable to other archaeal species. Thus, the library represents a significant step towards advanced genetics studies in archaea.

## Methods

### Strains and culture conditions

Microbial strains and their use are described in Table 2. *E. coli* strains were cultured at 37°C in Luria-Bertani (LB) medium or on LB-agar plates [75] supplemented with ampicillin (Amp, 100 µg/ml) and/or chloramphenicol (Cam, 10 µg/ml) when required. For *H. volcanii*, cultured at 45°C, MGM (18%) was used as a rich medium, Hv-Ca as a medium to select for the Trp marker, and Hv-Min as a minimal medium [76]. For culture dishes, these media were solidified by the addition of Bacto agar (Difco, 15 g/l). In Hv-Min dishes most amino acids were used at a concentration of 50 µg/ml. Exceptions were histidine (at 65 µg/ml) and threonine (at 10 µg/ml). Uracil was used at 50 µg/ml.

### Enzymes, reagents, oligonucleotides, DNA, and molecular biology techniques

Restriction endonucleases, calf alkaline phosphatase, and T4 DNA ligase were from New England Biolabs and used as recommended by the supplier. MuA transposase and Phusion DNA polymerase were from Finnzymes (currently Thermo Scientific). Plasmid DNA isolation kits and RNase A were from Macherey Nagel. Antibiotics, amino acids, and uracil were from Sigma. Oligonucleotides used are described in Table 6 and plasmids in Table 1. Standard DNA manipulation methods were used [75]. *E. coli* competent cells were prepared and electrotransformed as described previously [27]. *H. volcanii* transformation was done using the published protocol [31]. *H. volcanii* chromosomal DNA was isolated at room temperature as follows: cells were collected by centrifugation from three milliliters of late exponential phase culture. A total of 200 µl of ST buffer (1 M NaCl, 20 mM Tris-Cl, pH 7.5) was added and the cells were resuspended, after which they were lysed by the addition of 200 µl of lysis solution (100 mM EDTA, pH 8.0, 0.2% SDS) [76]. DNA was extracted twice with 1 volume of phenol with 1 h incubation at 60°C during the first extraction phase. Subsequently, DNA was extracted with 1.5 volume of chloroform and ethanol precipitated. Finally, DNA was dissolved in TE buffer (10 mM Tris-Cl, pH 7.5) containing RNase A (0.3 mg/ml). For the amplification of *H. volcanii* genomic DNA by Phusion DNA polymerase, PCR conditions recommended for GC-rich DNA were used (GC-buffer and 5% DMSO).

**Table 6 Oligonucleotides used in the work**

Oligonucleotide	Sequence (5' to 3')	Comment
HSP675	GCGCGCGGATCCAGTTATGTGCGTTCGGATG	Cloning of <i>Pfdx</i> + <i>trpA</i>
HSP683	GCGCGCGGATCCCCTGGATAAAACCCCTCGTT	Cloning of <i>Pfdx</i> + <i>trpA</i>
HSP734	CGTTGTAAAACGACGGCCAGTG	Sequencing from pBlueScript SK+, forward
HSP735	GGGAACAAAAGCTGGAGCTCC	Sequencing from pBlueScript SK+, reverse
HSP742	AACGCCTCGTAGAGCGTGTA	PCR spanning the insertion site of iSKT10. Forward, 5' at 228,414 of pHV3 <sup>#</sup>
HSP743	CGTCGTGGAACGTACCTCAT	PCR spanning the insertion site of iSKT10. Reverse, 5' at 230,985 of pHV3 <sup>#</sup>
HSP746	TGAGTCGAGACGGAGCGAGA	PCR spanning the insertion site of iSKT11. Forward, 5' at 1,530,341 of <i>H. volcanii</i> chromosome <sup>#</sup>
HSP747	CCCACGAGAAAGGCGAGAAC	PCR spanning the insertion site of iSKT11. Reverse, 5' at 1,534,771 of <i>H. volcanii</i> chromosome <sup>#</sup>
HSP744	GTTTCTTGGCGAGGGGTTTC	PCR spanning the insertion site of iSKT12. Forward, 5' at 385,560 of <i>H. volcanii</i> chromosome <sup>#</sup>
HSP745	CGAGACCTTCTCCAGCTCGT	PCR spanning the insertion site of iSKT12. Reverse, 5' at 390,228 of <i>H. volcanii</i> chromosome <sup>#</sup>
HSP750	CCCCGTGGAGGTAATAATTGACG	Sequencing of TrpA-cat-Mu insertion site, PCR of iSKT10 insertion site
HSP751	CGTCGCAACGCCACCGC	Sequencing of TrpA-cat-Mu insertion site

<sup>#</sup>Numbering according to *H. volcanii* D2 pHV3 (NC\_013964.1) and chromosome (NC\_013967.1), respectively.

### ***E.coli/H.volcanii* shuttle transposon**

TrpA-cat-Mu transposon is a derivative of Cat-Mu [33] with an added *H. volcanii*-specific selectable gene cassette (*trpA* gene driven by *Pfdx* promoter). It was constructed by initially amplifying a PCR fragment with the primer pair HSP683/HSP675 (Table 6) and plasmid pTA351 (Table 1) as a template. The fragment was trimmed with BamHI and ligated into a BamHI site of Cat-Mu such that the selectable marker genes (*trp* and *cat*) are transcribed in opposite directions (Figure 1A). TrpA-cat-Mu was maintained within its carrier plasmid pMPH20 (Table 1), from which it was isolated by BglII digestion and purified using anion exchange chromatography as described [33].

### **Plasmid library of transposon-tagged *H. volcanii* DNA**

*H. volcanii* strain H53 genomic DNA was digested partially with HpaII, AciI, and TaqI in three separate reactions, purified by phenol extraction, and ethanol precipitated. The DNA fragments were used as targets in three separate Mu *in vitro* transposition reactions [33] using TrpA-cat-Mu as a donor DNA. The reaction products were then pooled and purified as above. Anion exchange chromatography [33] was used to size-select 4-6 kb fragments, which were then ligated into pBlueScript SK+ cleaved with ClaI. The ligation mixture was electrotransformed into *E. coli* DH10B, and transposon-containing plasmid clones were selected on culture plates containing Amp and Cam. Approx. 28,000 colonies were pooled, and the cells were grown for 2.5 h in liquid culture containing Amp and Cam. Plasmid DNA was isolated from the pool, yielding a primary plasmid library with *H.volcanii* genome segments tagged with transposon DNA as inserts. An aliquot from the library was electrotransformed into *E. coli* DH5 $\alpha$  using the above selection, and approx. 600,000 colonies were pooled. The cells were grown and plasmid DNA was isolated as above. To eliminate

non-supercoiled plasmid forms, the plasmid preparation was further purified using CsCl-gradient ultracentrifugation [75].

### ***H. volcanii* insertion mutant library**

Transposon-tagged *H. volcanii* DNA fragments were released from the plasmid library using XhoI + HindIII double digestion. Linear 4-6 kb fragments were isolated using preparative 1% agarose gel (SeaPlaque GTG, Lonza). Several one microgram aliquots of linear DNA were used to transform [31] *H. volcanii* H295 cells. Following transformation, cells were collected by centrifugation and resuspended in 1 ml of transformant dilution solution [31], after which 335  $\mu$ l of 80% glycerol – 6% Salt Water [76] was added. The cells were frozen in aliquots under liquid nitrogen and stored at -75°C. For the evaluation of transformation efficiency, one aliquot was thawed, the cells were plated on selective media and cultured for seven days. To avoid the potential diversity decrease caused by unfavorable restriction site distribution, another library was made by releasing the transposon-tagged *H. volcanii* DNA fragments using KpnI + EcoRV double digestion.

### **Southern analysis**

EcoRI + KpnI double-digested chromosomal DNA (2.5  $\mu$ g) was separated on a 0.8% SeaKem LE agarose gel. The DNA was transferred onto Hybond-N+ -membrane (Amersham). Southern hybridization was performed as described [75] with biotin-labeled (Biotin DecaLabel DNA Labeling Kit, Thermo Scientific) *cat*-gene-specific DNA (PvuII-ScaI fragment of TrpA-*cat*-Mu). Biotin Chromogenic Detection Kit (Thermo Scientific) was used for visualization.

### **Mutant isolation**

An aliquot of *H. volcanii* H295 insertion mutant library was plated on selective media and cultured for seven days. Colonies were visually inspected for pigmentation deficiency, and white normal-sized colonies were isolated. To isolate auxotrophic mutants, 4,100 mutant clones were streaked onto Hv-Ca and Hv-Min plates. Mutants failing to grow on Hv-Min were isolated and studied for their ability to grow on supplemented Hv-Min plates containing one or several amino acids.

### **Determination of transposon location**

Initially, chromosomal DNA from isolated mutant clones was cleaved with KpnI, EcoRI, or SphI (not cleaving the transposon DNA) in separate reactions. From each reaction, the ensuing fragments were cloned into the corresponding site of pUC19 (Table 1). Plasmid clones containing transposon-tagged *H. volcanii* genomic DNA fragments were selected on LB-plates containing Amp and Cam. Plasmid DNA was isolated and the sequence from each transposon border was determined using transposon-specific primers HSP750 and HSP751 (Table 6). DNA sequence determination was performed at the sequencing facility of the Turku Centre for Biotechnology by using 3130xl Genetic Analyzer (Applied Biosystems). Genomic transposon insertion site loci were identified by using NCBI blastn suite [77] and the publicly available genome sequence of *H. volcanii* D2 strain [10]. Functions of enzymes along the carotenoid and amino acid biosynthesis pathways were assessed by using KEGG Pathway Maps [78].

## Description of additional data file

The following additional data are available with the online version of this paper. Additional data file 1 is a picture of pigmentation-deficient *H. volcanii* insertion mutants.

## Abbreviations

Amp: Ampicillin; bp: Base pairs; Cam: Chloramphenicol; CFU: Colony-forming unit; DNA: Deoxyribonucleic acid; EMS: Ethylmethanesulfonate; kb: Kilo-base pairs; LB: Luria-Bertani; PCR: Polymerase chain reaction.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

SK designed, generated, and verified the insertion mutant library; isolated and analyzed the mutants; and wrote the manuscript. MIP designed and generated the plasmid library, and drafted the manuscript. KD participated in the library constructions and analyses, and drafted the manuscript. SS participated in the design and generation of the libraries. MP conceived the study, participated in its design, and drafted the manuscript. HS conceived the study, participated in its design, and wrote the manuscript. All authors have read and approved the manuscript.

## Acknowledgements

This work was supported by the Academy of Finland (grants 124039 and 251168; [79]). Support was provided to SS by a DFG postdoctoral fellowship (Sto 804/1-1; [80]), to KD by a National Science Foundation (grant MCB02-39215; [81]), and MP by a National Aeronautics and Space Administration (grant NNX10AR84G; [82]). Thorsten Allers is acknowledged for providing the *H. volcanii* strain H295. Pirjo Rahkola, Tatjana Saarinen, and Tuuli Hoikkala are thanked for technical assistance.

## References

1. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: the primary kingdoms.** *Proc Natl Acad Sci U S A* 1977, **74**:5088–5090.
2. Woese CR, Fox GE: **The concept of cellular evolution.** *J Mol Evol* 1977, **10**:1–6.
3. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM: **The archaeobacterial origin of eukaryotes.** *Proc Natl Acad Sci* 2008, **105**:20356–20361.
4. Williams TA, Foster PG, Cox CJ, Embley TM: **An archaeal origin of eukaryotes supports only two primary domains of life.** *Nature* 2013, **504**:231–236.

5. Brochier-Armanet C, Forterre P, Gribaldo S: **Phylogeny and evolution of the Archaea: one hundred genomes later.** *Curr Opin Microbiol* 2011, **14**:274–281.
6. Koga Y: **Thermal adaptation of the archaeal and bacterial lipid membranes.** *Archaea* 2012, **2012**:789652.
7. Matsumi R, Atomi H, Driessen AJ, van der Oost J: **Isoprenoid biosynthesis in Archaea—biochemical and evolutionary implications.** *Res Microbiol* 2011, **162**:39–52.
8. Sumper M, Berg E, Mengele R, Strobel I: **Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*.** *J Bacteriol* 1990, **172**:7111–7118.
9. Ellen AF, Zolghadr B, Driessen AM, Albers SV: **Shaping the archaeal cell envelope.** *Archaea* 2010, **2010**:608243.
10. Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, Robinson J, Khouri H, Ren Q, Lowe TM, Maupin-Furlow J, Pohlschroder M, Daniels C, Pfeiffer F, Allers T, Eisen JA: **The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon.** *PLoS One* 2010, **5**:e9605.
11. Zerulla K, Chimileski S, Näther D, Gophna U, Papke RT, Soppa J: **DNA as a phosphate storage polymer and the alternative advantages of polyploidy for growth or survival.** *PLoS One* 2014, **4**:e94819.
12. Allers T, Mevarech M: **Archaeal genetics - the third way.** *Nat Rev Genet* 2005, **6**:58–73.
13. Leigh JA, Albers SV, Atomi H, Allers T: **Model organisms for genetics in the domain Archaea: methanogens, halophiles. *Thermococcales* and *Sulfolobales*.** *FEMS Microbiol Rev* 2011, **35**:577–608.
14. Giménez MI, Dilks K, Pohlschröder M: ***Haloferax volcanii* twin-arginine translocation substrates include secreted soluble, C-terminally anchored and lipoproteins.** *Mol Microbiol* 2007, **66**:1597–1606.
15. Storf S, Pfeiffer F, Dilks K, Chen ZQ, Imam S, Pohlschroder M: **Mutational and bioinformatic analysis of haloarchaeal lipobox-containing proteins.** *Archaea* 2010, **2010**:doi: 10.1155/2010/410975.
16. Zhou G, Kowalczyk D, Humbard MA, Rohatgi S, Maupin-Furlow JA: **Proteasomal components required for cell growth and stress responses in the haloarchaeon *Haloferax volcanii*.** *J Bacteriol* 2008, **190**:8096–8105.
17. Cohen A, Lam WL, Charlebois RL, Doolittle WF, Schalkwyk LC: **Localizing genes on the map of the genome of *Haloferax volcanii*, one of the Archaea.** *Proc Natl Acad Sci U S A* 1992, **89**:1602–1606.
18. Watrin L, Prieur D: **UV and ethyl methanesulfonate effects in hyperthermophilic archaea and isolation of auxotrophic mutants of *Pyrococcus* strains.** *Curr Microbiol* 1996, **33**:377–382.

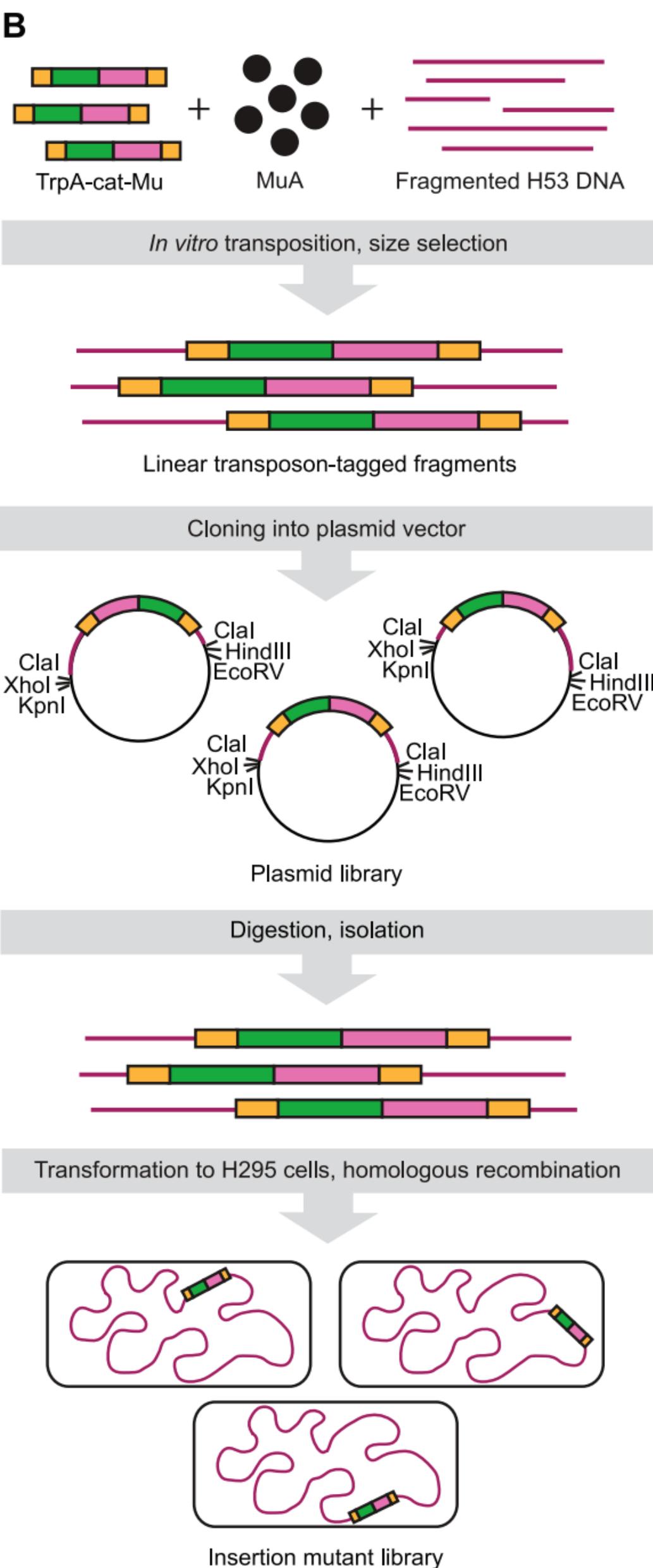
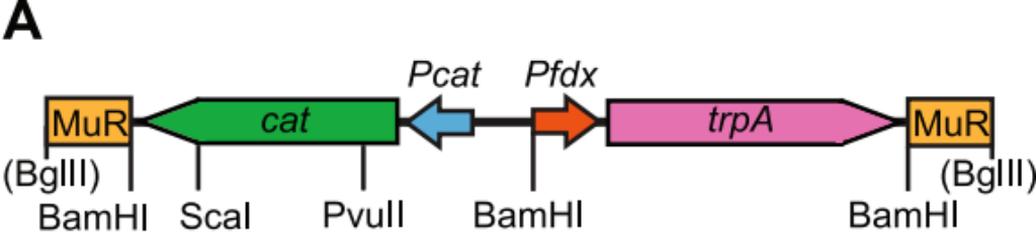
19. Lange C, Zerulla K, Breuert S, Soppa J: **Gene conversion results in the equalization of genome copies in the polyploid haloarchaeon *Haloferax volcanii***. *Mol Microbiol* 2011, **80**:666–677.
20. Zhang JK, Pritchett MA, Lampe DJ, Robertson HM, Metcalf WW: **In vivo transposon mutagenesis of the methanogenic archaeon *Methanosarcina acetivorans* C2A using a modified version of the insect *mariner*-family transposable element *HimarI***. *Proc Natl Acad Sci U S A* 2000, **97**:9665–9670.
21. Sarmiento F, Mrazek J, Whitman WB: **Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis***. *Proc Natl Acad Sci U S A* 2013, **110**:4726–4731.
22. Dyall-Smith ML, Doolittle WF: **Construction of composite transposons for halophilic Archaea**. *Can J Microbiol* 1994, **40**:922–929.
23. Woods WG, Ngui K, Dyall-Smith ML: **An Improved Transposon for the Halophilic Archaeon *Haloarcula hispanica***. *J Bacteriol* 1999, **181**:7140–7142.
24. Craig NL, Craigie R, Gellert M, Lambowitz AM: *Mobile DNA II*. Washington D.C.: ASM Press; 2002.
25. Skipper KA, Andersen PR, Sharma N, Mikkelsen JG: **DNA transposon-based gene vehicles - scenes from an evolutionary drive**. *J Biomed Sci* 2013, **20**:92-0127-20-92.
26. Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS: **Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes**. *Nat Biotechnol* 2000, **18**:97–100.
27. Lamberg A, Nieminen S, Qiao M, Savilahti H: **Efficient insertion mutagenesis strategy for bacterial genomes involving electroporation of *in vitro*-assembled DNA transposition complexes of bacteriophage Mu**. *Appl Environ Microbiol* 2002, **68**:705–712.
28. Hamer L, Adachi K, Montenegro-Chamorro MV, Tanzer MM, Mahanty SK, Lo C, Tarpey RW, Skalchunes AR, Heiniger RW, Frank SA, Darveaux BA, Lampe DJ, Slater TM, Ramamurthy L, DeZwaan TM, Nelson GH, Shuster JR, Woessner J, Hamer JE: **Gene discovery and gene function assignment in filamentous fungi**. *Proc Natl Acad Sci U S A* 2001, **98**:5110–5115.
29. Kumar A, Seringhaus M, Biery MC, Sarnovsky RJ, Umansky L, Piccirillo S, Heidtman M, Cheung K, Dobry CJ, Gerstein MB, Craig NL, Snyder M: **Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon**. *Genome Res* 2004, **14**:1975–1986.
30. Cline SW, Schalkwyk LC, Doolittle WF: **Transformation of the archaeobacterium *Halobacterium volcanii* with genomic DNA**. *J Bacteriol* 1989, **171**:4987–4991.
31. Cline SW, Lam WL, Charlebois RL, Schalkwyk LC, Doolittle WF: **Transformation methods for halophilic archaeobacteria**. *Can J Microbiol* 1989, **35**:148–152.

32. Bitan-Banin G, Ortenberg R, Mevarech M: **Development of a gene knockout system for the halophilic archaeon *Haloferax volcanii* by use of the *pyrE* gene.** *J Bacteriol* 2003, **185**:772–778.
33. Haapa S, Taira S, Heikkinen E, Savilahti H: **An efficient and accurate integration of mini-Mu transposons *in vitro*: a general methodology for functional genetic analysis and molecular biology applications.** *Nucleic Acids Res* 1999, **27**:2777–2784.
34. Yanisch-Perron C, Vieira J, Messing J: **Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.** *Gene* 1985, **33**:103–119.
35. Allers T, Ngo HP, Mevarech M, Lloyd RG: **Development of additional selectable markers for the halophilic archaeon *Haloferax volcanii* based on the *leuB* and *trpA* genes.** *Appl Environ Microbiol* 2004, **70**:943–953.
36. Norais C, Hawkins M, Hartman AL, Eisen JA, Myllykallio H, Allers T: **Genetic and physical mapping of DNA replication origins in *Haloferax volcanii*.** *PLoS Genet* 2007, **3**:e77.
37. Delmas S, Shunburne L, Ngo HP, Allers T: **Mre11-Rad50 promotes rapid repair of DNA damage in the polyploid archaeon *Haloferax volcanii* by restraining homologous recombination.** *PLoS Genet* 2009, **5**:e1000552.
38. Grant SG, Jessee J, Bloom FR, Hanahan D: **Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants.** *Proc Natl Acad Sci U S A* 1990, **87**:4645–4649.
39. Meissner PS, Sisk WP, Berman ML: **Bacteriophage lambda cloning system for the construction of directional cDNA libraries.** *Proc Natl Acad Sci U S A* 1987, **84**:4171–4175.
40. Ronnekleiv M, Liaaen-Jensen S: **Bacterial carotenoids 53\*, C-50-carotenoids 23; Carotenoids of *Haloferax volcanii* versus other halophilic bacteria.** *Biochem Syst Ecol* 1995, **23**:627–634.
41. Dogbo O, Laferriere A, D’Harlingue A, Camara B: **Carotenoid biosynthesis: Isolation and characterization of a bifunctional enzyme catalyzing the synthesis of phytoene.** *Proc Natl Acad Sci U S A* 1988, **85**:7054–7058.
42. Sieiro C, Poza M, de Miguel T, Villa TG: **Genetic basis of microbial carotenogenesis.** *Int Microbiol* 2003, **6**:11–16.
43. **The UCSC Archaeal Genome Browser.** <http://archaea.ucsc.edu/>.
44. Schneider KL, Pollard KS, Baertsch R, Pohl A, Lowe TM: **The UCSC Archaeal Genome Browser.** *Nucleic Acids Res* 2006, **34**:D407–D410.
45. Chan PP, Holmes AD, Smith AM, Tran D, Lowe TM: **The UCSC Archaeal Genome Browser: 2012 update.** *Nucleic Acids Res* 2012, **40**:D646–D652.

46. Lynch EA, Langille MG, Darling A, Wilbanks EG, Haltiner C, Shao KS, Starr MO, Teiling C, Harkins TT, Edwards RA, Eisen JA, Facciotti MT: **Sequencing of seven haloarchaeal genomes reveals patterns of genomic flux.** *PLoS One* 2012, **7**:e41389.
47. Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA, Hamer JE: **Recent advances in large-scale transposon mutagenesis.** *Curr Opin Chem Biol* 2001, **5**:67–73.
48. Choi KH, Kim KJ: **Applications of transposon-based gene delivery system in bacteria.** *J Microbiol Biotechnol* 2009, **19**:217–228.
49. Ivics Z, Li MA, Mates L, Boeke JD, Nagy A, Bradley A, Izsvak Z: **Transposon-mediated genome manipulation in vertebrates.** *Nat Methods* 2009, **6**:415–422.
50. Ivics Z, Izsvak Z: **The expanding universe of transposon technologies for gene and cell engineering.** *Mob DNA* 2010, **1**:25. -8753-1-25.
51. Filee J, Siguier P, Chandler M: **Insertion sequence diversity in archaea.** *Microbiol Mol Biol Rev* 2007, **71**:121–157.
52. Savilahti H, Rice PA, Mizuuchi K: **The phage Mu transpososome core: DNA requirements for assembly and function.** *EMBO J* 1995, **14**:4893–4903.
53. Haapa-Paananen S, Rita H, Savilahti H: **DNA transposition of bacteriophage Mu. A quantitative analysis of target site selection *in vitro*.** *J Biol Chem* 2002, **277**:2843–2851.
54. Pajunen M, Turakainen H, Poussu E, Peränen J, Vihinen M, Savilahti H: **High-precision mapping of protein-protein interfaces: an integrated genetic strategy combining *en masse* mutagenesis and DNA-level parallel analysis on a yeast two-hybrid platform.** *Nucleic Acids Res* 2007, **35**:e103.
55. Orsini L, Pajunen M, Hanski I, Savilahti H: **SNP discovery by mismatch-targeting of Mu transposition.** *Nucleic Acids Res* 2007, **35**:e44.
56. Turakainen H, Saarimaki-Vire J, Sinjushina N, Partanen J, Savilahti H: **Transposition-based method for the rapid generation of gene-targeting vectors to produce Cre/Flp-modifiable conditional knock-out mice.** *PLoS One* 2009, **4**:e4341.
57. Poussu E, Vihinen M, Paulin L, Savilahti H: **Probing the  $\alpha$ -complementing domain of *E. coli*  $\beta$ -galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu *in vitro* DNA transposition.** *Proteins* 2004, **54**:681–692.
58. Poussu E, Jäntti J, Savilahti H: **A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: mapping of the Sec1p binding domain in yeast Mso1p by a Mu *in vitro* transposition-based approach.** *Nucleic Acids Res* 2005, **33**:e104.
59. Rasila TS, Vihinen M, Paulin L, Haapa-Paananen S, Savilahti H: **Flexibility in MuA transposase family protein structures: functional mapping with scanning mutagenesis and sequence alignment of protein homologues.** *PLoS One* 2012, **7**:e37922.

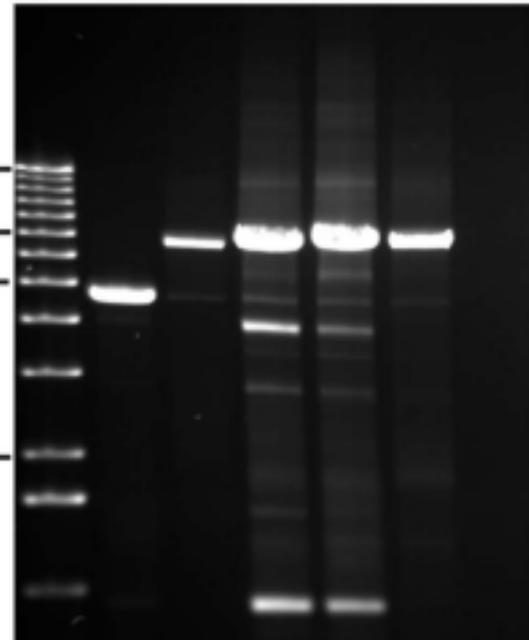
60. Jones DD: **Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1  $\beta$ -lactamase to an amino acid deletion.** *Nucleic Acids Res* 2005, **33**:e80.
61. Simm AM, Baldwin AJ, Busse K, Jones DD: **Investigating protein structural plasticity by surveying the consequence of an amino acid deletion from TEM-1  $\beta$ -lactamase.** *FEBS Lett* 2007, **581**:3904–3908.
62. Edwards WR, Busse K, Allemann RK, Jones DD: **Linking the functions of unrelated proteins using a novel directed evolution domain insertion method.** *Nucleic Acids Res* 2008, **36**:e78.
63. Baldwin AJ, Busse K, Simm AM, Jones DD: **Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx).** *Nucleic Acids Res* 2008, **36**:e77.
64. Pajunen MI, Pulliainen AT, Finne J, Savilahti H: **Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes.** *Microbiology* 2005, **151**:1209–1218.
65. Kiljunen S, Vilen H, Pajunen M, Savilahti H, Skurnik M: **Nonessential genes of phage  $\phi$ YeO3-12 include genes involved in adaptation to growth on *Yersinia enterocolitica* serotype O:3.** *J Bacteriol* 2005, **187**:1405–1414.
66. Krupovic M, Vilen H, Bamford JK, Kivelä HM, Aalto J, Savilahti H, Bamford DH: **Genome characterization of lipid-containing marine bacteriophage PM2 by transposon insertion mutagenesis.** *J Virol* 2006, **80**:9270–9278.
67. Paatero AO, Turakainen H, Happonen LJ, Olsson C, Palomaki T, Pajunen MI, Meng X, Otonkoski T, Tuuri T, Berry C, Malani N, Frilander MJ, Bushman FD, Savilahti H: **Bacteriophage Mu integration in yeast and mammalian genomes.** *Nucleic Acids Res* 2008, **36**:e148.
68. Krysan PJ, Young JC, Sussman MR: **T-DNA as an insertional mutagen in *Arabidopsis*.** *Plant Cell* 1999, **11**:2283–2290.
69. Holmes ML, Nuttall SD, Dyall-Smith ML: **Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase.** *J Bacteriol* 1991, **173**:3807–3813.
70. Blaseio U, Pfeifer F: **Transformation of *Halobacterium halobium*: development of vectors and investigation of gas vesicle synthesis.** *Proc Natl Acad Sci U S A* 1990, **87**:6772–6776.
71. Boucher U, Kamekura M, Doolittle WF: **Origins and evolution of isoprenoid lipid biosynthesis in archaea.** *Mol Microbiol* 2004, **52**:515–527.
72. Britton G, Liaaen-Jensen S, Phander H: *Carotenoids. Vol. 4: Natural Functions.* Switzerland: Birghäuser Verlag; 2008.

73. **EMBOSS Water.** [http://www.ebi.ac.uk/Tools/psa/emboss\\_water/](http://www.ebi.ac.uk/Tools/psa/emboss_water/).
74. Liu X, Miao D, Zhang F, Wu Z, Liu J, Xiang H: **Characterization of the minimal replicon of pHM300 and independent copy number control of major and minor chromosomes of *Haloferax mediterranei*.** *FEMS Microbiol Lett* 2013, **339**:66–74.
75. Sambrook J, Russell DW: *Molecular cloning: a laboratory manual*. 3rd edition. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press; 2001.
76. Dyall-Smith M: **The Halohanbook.**  
[<http://www.haloarchaea.com/resources/halohandbook/>].
77. **NCBI blast.** <http://blast.ncbi.nlm.nih.gov>.
78. **KEGG PATHWAY Database.** <http://www.genome.jp/kegg/pathway.html>.
79. **Academy of Finland.** <http://www.aka.fi/en-GB/A/>.
80. **The Alliance of Science Organizations in Germany.** <http://www.dfg.de/en/>.
81. **The National Science Foundation.** <http://www.nsf.gov>.
82. **National Aeronautics and Space Administration.** <http://www.nasa.gov>.

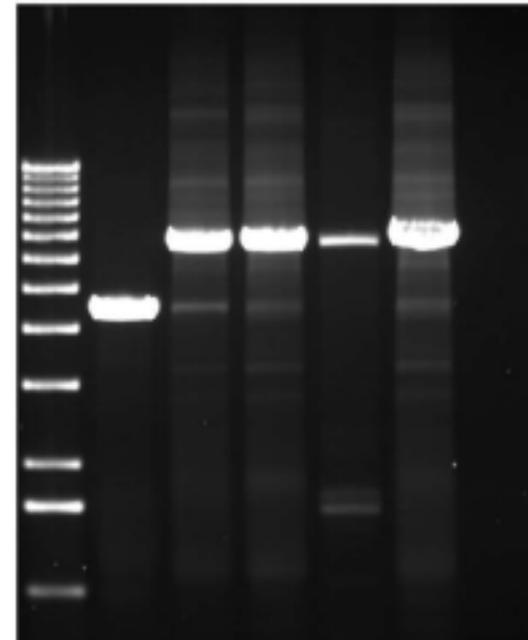


**A**

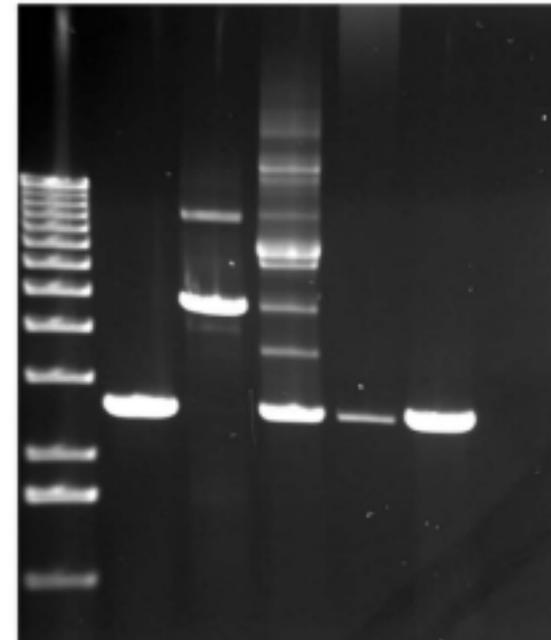
1 kb H295 H295/iSKT11  
1 2 3 4 Ctrl

**B**

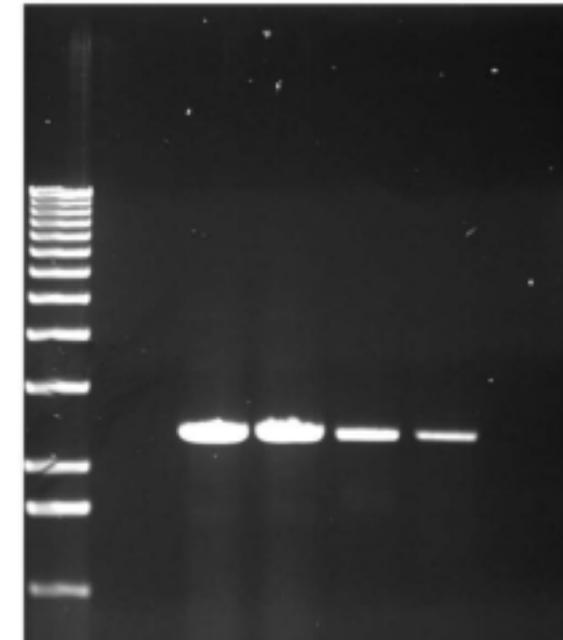
1 kb H295 H295/iSKT12  
1 2 3 4 Ctrl

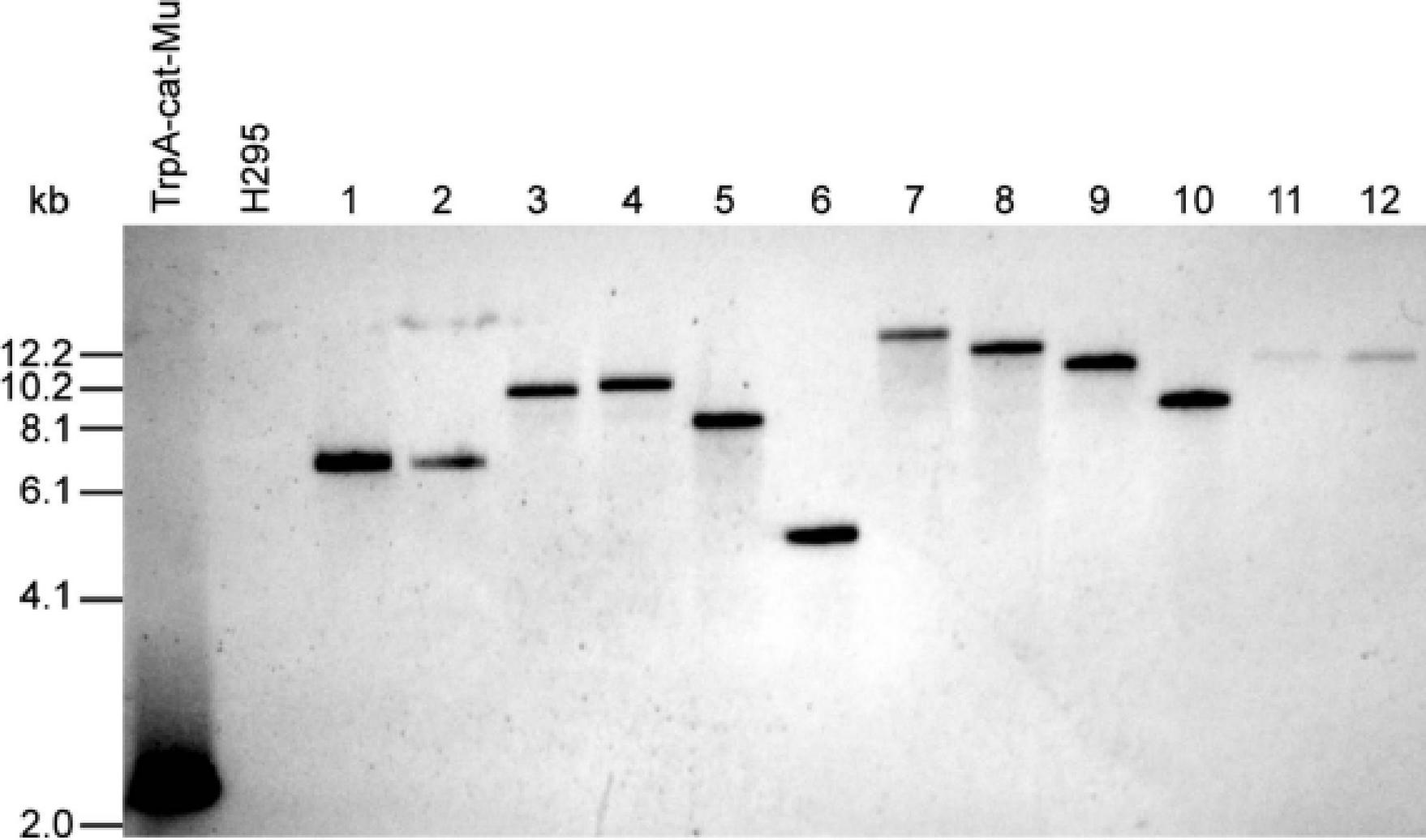
**C**

1 kb H295 H295/iSKT10  
1 2 3 4 Ctrl

**D**

1 kb H295 H295/iSKT10  
1 2 3 4 Ctrl







## **Additional files provided with this submission:**

**Additional file 1: Figure S1.** Pigmentation-deficient mutants. Carotenoid biosynthesis mutants and the control strain H295/pTA231 were cultured on Hv-Ca dish at 45°C for four days and subsequently incubated at RT for four days. One representative mutant for each identified crtB insertion site is shown. Colonies were imaged using Olympus SZX12 microscope and Jenoptic ProgRes SpeedXT core 5 camera (6156k)

<http://www.biomedcentral.com/content/supplementary/s12915-014-0103-3-s1.tiff>