PNAS
www.pnas.org

Supplementary Information for

A non-mutational mechanism of inheritance in the Archaeon *Sulfolobus solfataricus*.

Sophie Payne, Samuel McCarthy, Tyler Johnson, Erica North, and Paul Blum.

Paul Blum
Email: pblum1@unl.edu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1808221115

Supplemental methods

Genome and transcriptome sequencing.

Sequencing of parental, partially-evolved and fully evolved SARC strains. Paired-end sequencing was performed as in McCarthy et al (1). Genome coverage was ≥500 for all samples, read length was 100bp, and insert size was about 300-500bp. The SARC strains were aligned to their respective parental strains using Bowtie2 v2.3.4.1 (default settings with –sensitive, –end-to-end and –interleaved options) (2) and Samtools v1.9 (default settings) (3). To avoid false negatives for genomic rearrangements and transpositions, genome sequences were also assembled into contigs *de novo* using SPAdes v33.7.1 (default settings with –careful option) (4) and into scaffolds using Ragout reference-assisted assembly (default settings) (5). For genome sequencing, lineage traceability was verified by ensuring the SARC strains contained the same mutations as their ancestral partially evolved strain. This consistency of mutations also confirms mutation authenticity. Genome and transcriptome sequence information is available in the NCBI databank. Prior to sequencing, genomic DNA was analyzed by electrophoresis and formed a single high molecular weight band lacking smearing. The absence of additional lower molecular weight bands indicated a lack of plasmids that might act to facilitate genome rearrangement or transposition.

Sequencing of passaging control strain (SUL120). The genome of the SUL120 passaging control strain was sequenced using the PacBio sequencing platform. The point mutation accuracy of this method is comparable to that of Illumina and identifies mobile element transfer efficiently (6). The DNA sample was purified by phenol/chloroform extraction and sheared using a Megaruptor instrument (Diagenode, Liege, Belgium) to a 10kb average size. The fragments were converted into a sequencing library using the SMRTbell Template Prep Kit 1.0 (Pacific Biosciences). The library was sequenced using P6C4 chemistry on a single-molecule real-time (SMRT) cell with a 240-min collection protocol on a PacBio RSII sequencer. The PacBio hierarchical genome assembly pipeline (RS_HGAP.2) was used to assemble the data *de novo*, with polishing using the PacBioSMRT-Portal embedded Quiver software. Sequencing data filtering parameters were set to have a minimum polymerase read quality of 0.8 and a minimum seed length of 6kp for the assembly. The BLASR aligner (Pacific Biosciences) was used to align the RAW reads to the *de novo* assembly at default settings. For other analyses filtered subreads were extracted from the PacBio data using the RS_Subreads.1 protocol (Pacific Biosciences) at default settings. PacBio data produced 120-fold genome coverage.

Analysis of genomic data.

Point mutation identification. Evolved strain (SARC-B, SARC-C, SARC-H, SARC-I, SARC-N, SARC-O) genomes were compared to parental genomes to identify mutations using variant calling software bcftools v1.9 (3) and vcflib v1.9 (7) (default settings). Variant calls were filtered for DP>50 to allow coverage at least 10% of the average, and QUAL >30 for calls with less than 0.1% chance of error. Putative mutations were visually inspected using Integrative Genomics Viewer (IGV). Mutations were classified as “likely” when between 50 and 80% of aligned bases differed from the reference genome. Mutations were classified as “real” when > 80% of aligned bases differed from the reference genome. All “likely” mutations were experimentally tested using PCR and resequencing (8) to be reclassified as “real”. ”Real” mutations were bioinformatically analyzed further and recorded in Table S2. For SUL120 however, the high mutation count precluded resequencing of all 102 ≤80% point mutations. Instead, to estimate the accuracy of point mutations, a random sample of 11 mutations where ~75% of aligned

bases at that position differed from the genome were resequenced. Based on the 75% alignment, the mutated position was predicted to be erroneous 25% of the time (3/11). Instead, all were found to be true mutations.

The point mutations identified using the methods above were bioinformatically analyzed to determine codon positions and effect on protein sequence or regulation. For the SARC strains, mutation location was also cross-referenced to known functional domains to identify potential sources of changes in protein function. Potential relationships between mutated genes were investigated through manual annotation and literature searches, in addition to searching the KEGG database for presence in similar pathways.

Genome rearrangement identification. Mauve 2.4.0 progressive Mauve aligner was used to identify genome rearrangements in reference-mapped genomes using default settings (9). Deleted regions were validated to be pre-existing transposase ORFs, and insertion sequences were analyzed using BLAST to identify a potential source ORF. The identities of all active transposases were identified based on BLAST sequence similarity. Of the 255 putative transposases annotated in the SULA genome, the families of 129 IS elements could be determined using the annotation available for the *S. solfataricus* SULA genome on NCBI (Genbank accession number CP011057.1 (1, 10)). Genome rearrangements were validated using PCR and agarose gel size comparison. For SUL120, the accuracy of a representative set of 10 transposons was extrapolated to the larger dataset. Because 4 of 4 tested insertions were real, all insertion events were included in SUL120 transposition analysis. Of 6 deletion events detected in sequencing data, none were found to have existed in the parental SULA strain via PCR. As a result, many SUL120 deletion events might be false positives arising from improper assembly of IS regions in SULA. Although 18 deletion events were originally detected in SUL120, a conservative estimate of 0 events was used for transposition analysis.

As the SARC-I strain was found to have no point mutations or apparent transpositions, DELLY v0.7.8 (default settings) (11) was used to identify genome rearrangements such as transposition, inversions and copy number variation. A simulated inversion was called with a PASS filter and a PE score of 1000. On real datasets (SULG and SARC-I), potential SVs and CNV were filtered to those with a PASS filter and PE score > 100. Those that only appeared in SARC-I were analyzed further by visualization in IGV. The four potential inversions that passed filtering and were supported by split reads in IGV were tested using PCR. Primers flanked the breakpoints and would produce amplicons if there was no rearrangement. Primers were also chosen that would amplify if an inversion occurred. With this, all putative rearrangements were excluded. Pindel v0.2.5b9 (default settings) was used as an alternative method to test for structural rearrangements. Results were filtered to have a read depth of 100, and that each potential event required that the reads that supported it in SARC-I were ≥2x greater than in SULG. No additional putative structural rearrangements were detected. Point mutation rate was calculated as mutations per base per cell division (Cell divisions = 120, bases = 2.7Mb). The transposition rate was calculated as transpositions each cell division cycle. Point mutation and transposition counts were determined as described above. The cell division count was determined by tracking cell growth over the course of passaging.

Analysis of transcriptomic data

Determination of the SARC transcriptome. Transcriptomic data from the evolved SARC strains was compared to their parental strains using EdgeR (12) and analyzed to find conserved gene expression

patterns likely to contribute to acid resistance. To avoid the introduction of artifacts, only genes whose differential expression results had p-values ≤ 0.05 in all three SARC lines were analyzed further. Altered expression of a gene that was conserved in all SARC strains was likely to be phenotypically important (part of the SARC transcriptome) and not a result of random chance. The number of genes with conserved expression predicted to be false positives was determined by multiplying the highest p-values of each dataset (0.05), and then multiplying by the total genes in the transcriptome. For example, the false discovery rate for genes with conserved expression in three fully evolved SARC strains was $\sim 0.05^3 \times 2924 = 0.37$ genes. The false discovery rate for genes conserved in all six fully evolved and partially evolved intermediates was $\sim 0.05^6 \times 2924 = 0.00005$ genes. To be considered a SARC transcriptome gene: (1) The fold change direction was consistent between the three fully evolved SARC strains, such that all are upregulated or all are downregulated. (2) The average fold change between the 3 strains was greater than 2 fold. (3) The SARC-I acid-stress control showed an expression shift in the same direction, showing that the transcriptomic change was heritable and not due to stress. (4) The range of standard error for the average does not bring the fold change to less than 2. (5) The gene was not a transposase, as transposases are unlikely to contribute to the SARC phenotype.

Identification of factors that could facilitate the SARC transcriptome. To determine whether transcriptomic changes correlated with GC content, small regions of high GC content were identified by calculating the GC% for every 5bp and looking for regions of where a stretch >20bp in length was >20% different than the norm up to 100bp away from a transcriptionally altered gene. Proximity to tRNAs and origins of replication was determined using the NCBI genome browser. To determine whether the SARC transcriptome could be co-regulated, DAVID v6.8 (default settings) (13) was used to determine gene ontology and functional enrichment relative to the *S. solfataricus* genome. A Benjamini-Hochberg corrected p-value of 0.1 was used as a cutoff for significant enrichment (14, 15). To determine the existence of potential transcription factor binding sites, the 100bp 5' flanking regions of the 12 most upregulated genes and 12 most downregulated genes in the SARC transcriptome were collected in addition to 17 unaltered genes. The sequences were submitted to MEME v4.11.2 (default settings) (16) and analyzed on normal mode, searching for zero or one motif occurrence per sequence, and searching for 5 motifs total. The occurrence of each motif, regardless of p-value, within the SULA genome was determined using FIMO v4.11.2 (default settings), with a p-value limit of 0.0001.

Recombination with phenotypic and expression analysis

Genes for recombination. SARC genes selected for homologous recombination were SULA_2869, a pyruvate ferredoxin oxidoreductase, and SULA_2027, a formate hydrogenlyase/proton translocating membrane protein. Both were upregulated in the SARC strains and likely to be involved in acid resistance (17, 18). As a non-SARC control gene, SULA_0895 (aldehyde dehydrogenase) was chosen because it had unaltered expression in the SARC strains and was shown previously to be non-essential (19).

Recombination procedure. A 1kb segment of the gene of interest and its promoter was amplified using primers (Table S7) and inserted into a marker-containing suicide plasmid, pUC19-*lacS*. Plasmid inserts were sequenced to avoid introducing mutations. SARC-I was transformed, and the insertion and excision of the plasmid was traced using the *lacS* marker. Integration at the target site was validated using PCR. Segregants were collected from the same recombinant intermediate having the integrated plasmid and both copies of the target locus to ensure that the final recombinants arose from the same

genetic intermediate. To enrich for unique segregation events, passage number from this intermediate was limited.

Expression analysis after recombination. Recombinants at SARC loci that acquired acid-sensitive phenotypes were predicted to have also acquired decreased expression of the target SARC gene. Similarly, recombinants that retained acid resistant phenotypes were predicted to have unaltered expression levels of the target SARC gene. Because recombinants at SARC loci varied in phenotype, three acid-sensitive (altered phenotype) and three acid-resistant isolates (unchanged phenotype) for each recombinant set (SULA_2027, SULA_2869) were measured for expression of the targeted SARC gene. A control non-SARC gene recombinant set (SULA_0895) was similarly tested.

To determine whether expression changes incurred by HR were locus specific, expression levels of genes targeted for HR were compared to expression of the same gene in isolates that underwent HR at a different locus. For expression analysis of recombinants, all RNA was extracted from mid-exponential phase cell cultures that had been passaged twice at pH 1.2 to stabilize physiology and expression patterns after the transformation. RT-PCR amplifications were performed in parallel the gene SULA_2002, which was used for expression normalization because it had consistent expression across all transcriptomes. The amount of expression difference between the SARC and parental strains were consistent with RNAseq data, but of a lower intensity as is sometimes observed for qRTPCR (20). SigmaPlot 11.0 was used for statistical analysis of growth and expression data.

Data availability

The genome sequences for *S. solfataricus* 98/2 strain SULA, partially evolved strain SARC-B (SULB) and terminal SARC strain SARC-C (SULC) are available at GenBank under the accession numbers CP011057.2, CP011055.2 and CP011056.2, respectively (1, 10). Newly resequenced SULG and its derived lineages SARC-H and SARC-I are available under accession numbers CP033235, CP033236, CP033237. SULM and its derived lineages SARC-N and SARC-O are available under accession numbers CP033238, CP033239, CP033240. SULA120 is available under accession number CP033241. Raw DNA-seq data are available in the JGI Genome portal under JGI Project IDs 1019966 (SULA), 1019969 (SARC-B), 1019972 (SARC-C), 1019975 (SULG), 1019978 (SARC-H), 1019981 (SARC-I), 1019984 (SULM), 1019987 (SARC-N) and 1019990 (SARC-O). Raw RNA-seq data are available in the JGI Genome portal under JGI Project IDs 1019993 (SULA), 1019996 (SARC-B), 1019999 (SARC-C), 1020002 (SULG), 1020005 (SARC-H), 1020008 (SARC-I), 1036511 (SARC-I stress control) 1020011 (SULM), 1020014 (SARC-N) and 1020017 (SARC-O). RNA-seq data is also available in the Sequence Read Archive under accession numbers SRX872629 and SRX872630 (SULA), SRX872631 and SRX872632 (SARC-B), SRX712375 and SRX712376 (SARC-C), SRX872634 and SRX872633 (SULG), SRX875205 (SARC-H), SRX712379 and SRX712378 (SARC-I), SRX872636 and SRX872635 (SULM), SRX712413 and SRX712412 (SARC-N), and SRX872640 and SRX872639 (SARC-O).

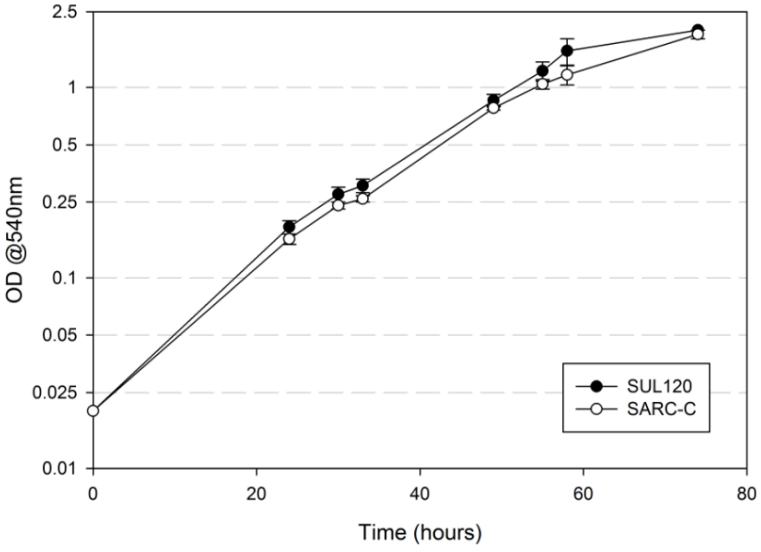
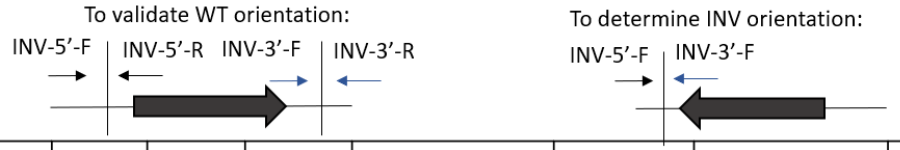


Fig. S1. Growth of SUL120 and SARC-C at pH 3.0 (n=2).

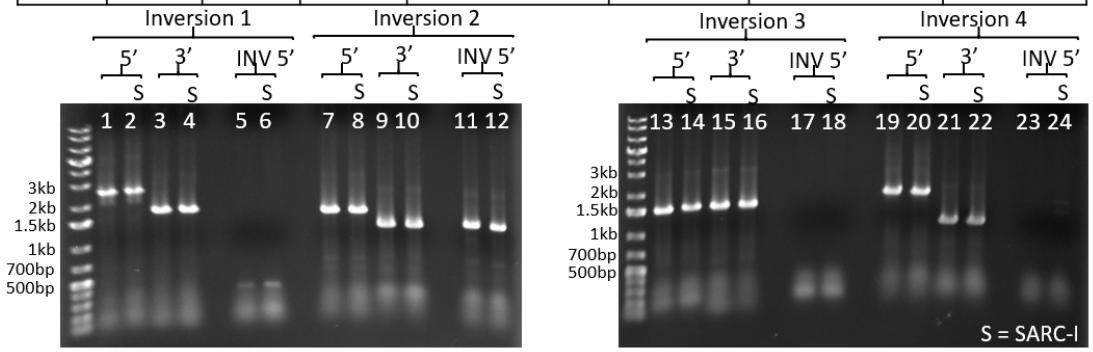
A.1



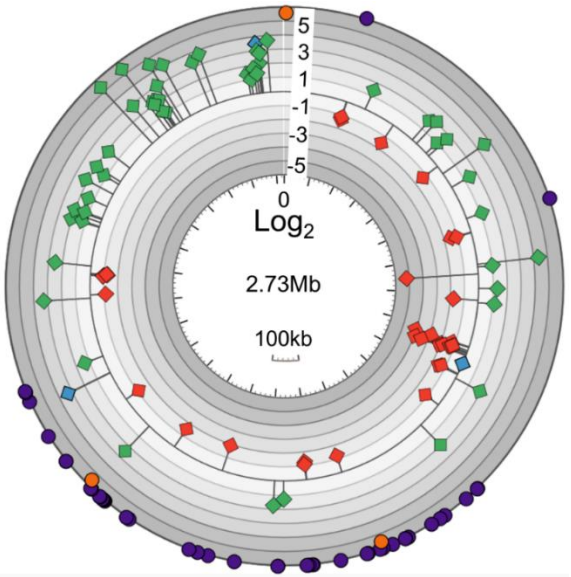
A.2

Predicted Inversion (INV)	Start coordinate	End coordinate	Length of predicted inversion	Primers for WT orientation	Length of predicted amplicon if WT	Primers for INV orientation	Length of predicted amplicon if INV
1	606715	1973957	1367kb	5': INV1-5'-F, INV1-5'-R 3': INV1-3'-F, INV1-3'-R	5' – 2600bp 3' – 1751bp	5': INV1-5'-F, INV1-3'-F	2422bp
2	675554	678396	2.8kb	5': INV2-5'-F, INV2-5'-R 3': INV2-3'-F, INV2-3'-R	5' – 1671bp 3' – 1300bp	5': INV2-5'-F, INV2-3'-F	1743bp
3	1477076	1483552	6.5kb	5': INV3-5'-F, INV3-5'-R 3': INV3-3'-F, INV3-3'-R	5' – 1518bp 3' – 1509bp	5': INV3-5'-F, INV3-3'-F	2008bp
4	2414765	2419640	4.9kb	5': INV4-5'-F, INV4-5'-R 3': INV4-3'-F, INV4-3'-R	5' – 1853bp 3' – 1100bp	5': INV4-5'-F, INV4-3'-F	1757bp

A.3



B

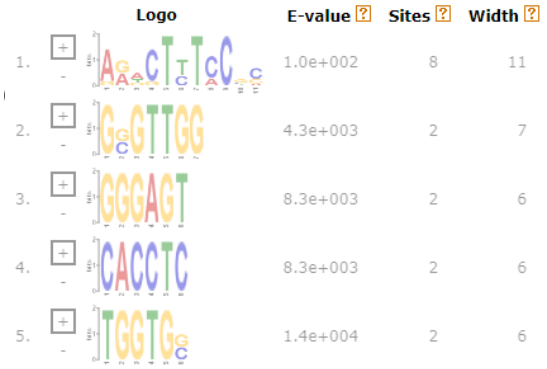


C

Annotation Cluster 1	**Enrichment Score: 1.53**	**G**			**Count**	**P_Value**	**Benjamini**
GOTERM_MF_DIRECT	pyruvate synthase activity	RT			3	9.5E-3	2.3E-1
UP_KEYWORDS	Oxidoreductase	RT			8	1.9E-2	4.5E-1
UP_KEYWORDS	Pyruvate	RT			3	1.4E-1	6.8E-1
Annotation Cluster 2	**Enrichment Score: 0.94**	**G**			**Count**	**P_Value**	**Benjamini**
UP_KEYWORDS	Transmembrane	RT			20	7.9E-2	7.2E-1
UP_KEYWORDS	Transmembrane helix	RT			20	7.9E-2	7.2E-1
UP_KEYWORDS	Membrane	RT			20	8.3E-2	5.9E-1
GOTERM_CC_DIRECT	integral component of membrane	RT			19	3.3E-1	9.6E-1
Annotation Cluster 3	**Enrichment Score: 0.21**	**G**			**Count**	**P_Value**	**Benjamini**
GOTERM_MF_DIRECT	ATPase activity	RT			3	3.4E-1	1.0E0
SMART	AAA	RT			3	4.7E-1	9.8E-1
INTERPRO	AAA+ ATPase domain	RT			3	5.3E-1	1.0E0
GOTERM_MF_DIRECT	ATP binding	RT			5	7.1E-1	1.0E0
INTERPRO	P-loop containing nucleoside triphosphate hydrolase	RT			4	7.4E-1	1.0E0
UP_KEYWORDS	Nucleotide-binding	RT			4	7.9E-1	1.0E0
UP_KEYWORDS	ATP-binding	RT			3	8.8E-1	1.0E0

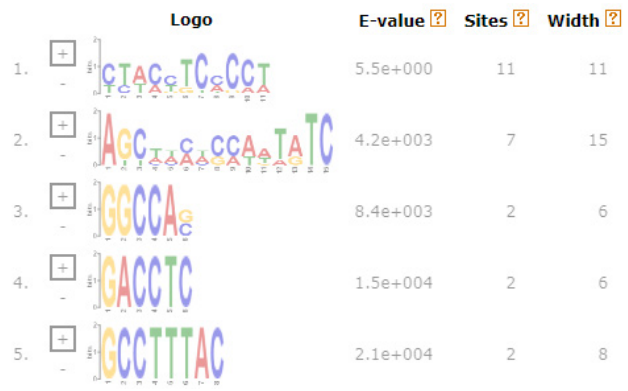
D

Upregulated SARC transcriptome genes



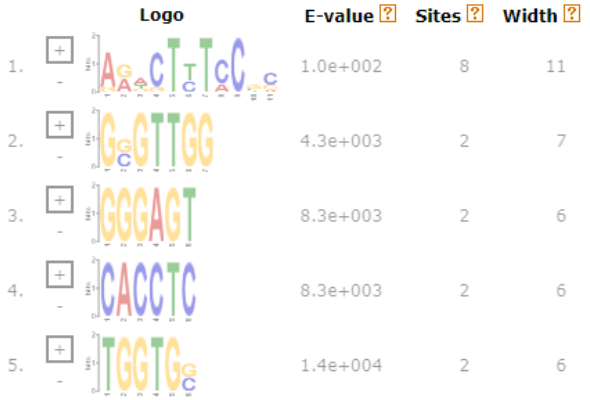
	Logo	E-value	Sites	Width
1.		1.0e+002	8	11
2.		4.3e+003	2	7
3.		8.3e+003	2	6
4.		8.3e+003	2	6
5.		1.4e+004	2	6

Downregulated SARC transcriptome genes



	Logo	E-value	Sites	Width
1.		5.5e+000	11	11
2.		4.2e+003	7	15
3.		8.4e+003	2	6
4.		1.5e+004	2	6
5.		2.1e+004	2	8

Unaltered SARC transcriptome genes



	Logo	E-value	Sites	Width
1.		1.0e+002	8	11
2.		4.3e+003	2	7
3.		8.3e+003	2	6
4.		8.3e+003	2	6
5.		1.4e+004	2	6

E

	Occurrences in SULA genome		
Motif #	Upregulated	Downregulated	Unaltered
1	509	857	740
2	0	884	846
3	979	0	2077
4	0	1010	305
5	2356	328	0

Fig. S2. Excluding known mechanisms of transcriptomic inheritance. A. Exclusion of the occurrence of predicted genome rearrangements in SARC-I. A.1. Schematic for validating Delly-predicted inversions using PCR primers. Arrows indicate gene region (black) and primers (green and blue). WT, wild type orientation, INV, inverted orientation. A.2. Locations of inversions (INV) predicted using DELLY and primers used to determine occurrence of inversions. A.3. Results for PCR determination of the presence of inversions in SULG (WT) and SARC-I (adapted). “S” indicates SARC-I lanes while blank lanes are SULG. Identical amplicons for SULG and SARC-I indicate no inversion event. B. Location of tRNAs (purple circles) and origins of replication (orange circles) in relation to the SARC transcriptome (see figure 1B). C. DAVID gene ontology clustering and functional enrichment analysis results for the SARC transcriptome. D. MEME motif discovery results from upregulated, downregulated SARC transcriptome genes or random unaltered genes. E. Occurrence frequency of each discovered motif (regardless of e-value) in the SULA genome using FIMO

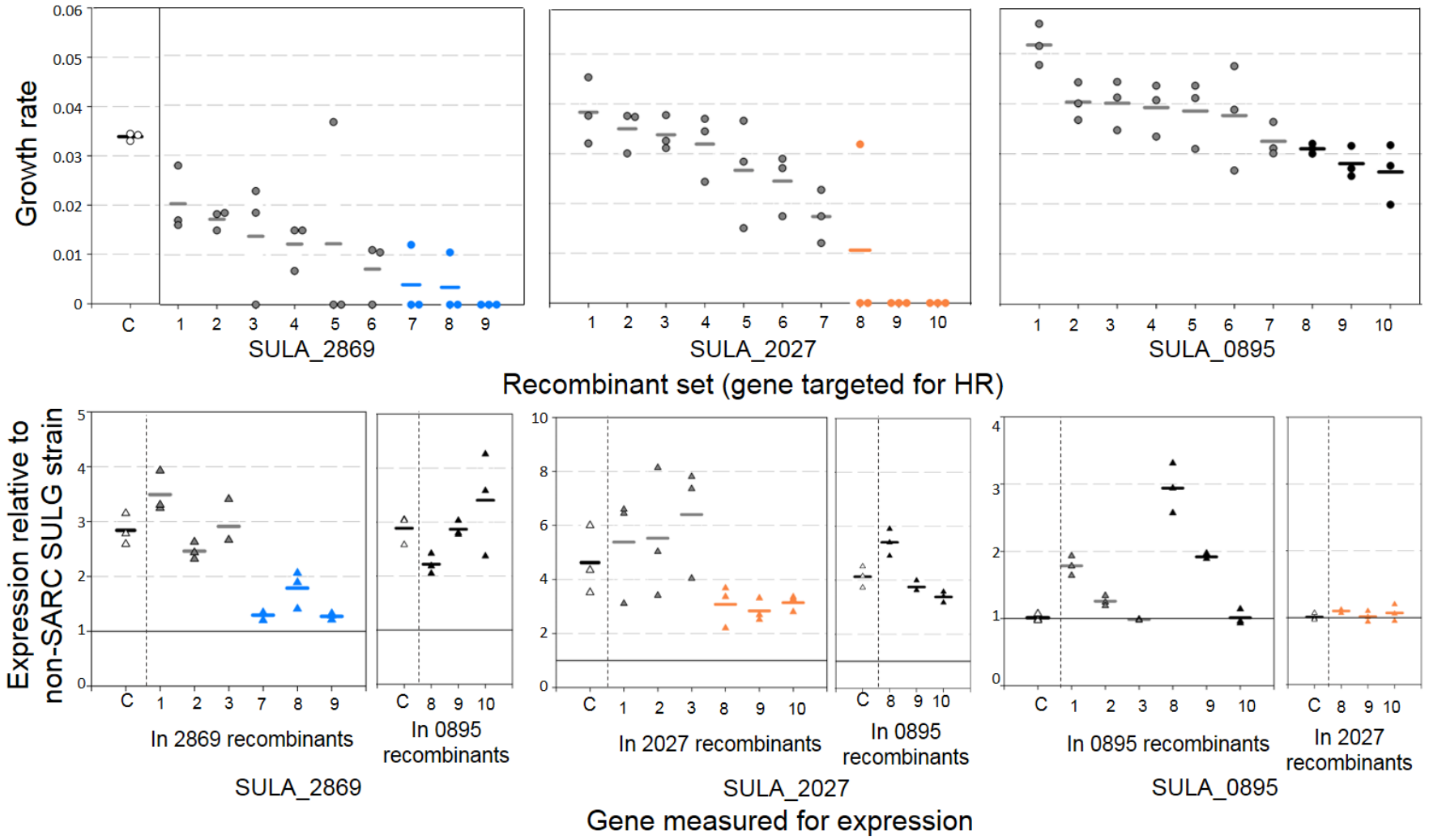


Fig. S3. Growth rate at low pH and expression data for SARC recombinants and controls. A) Recombinant growth rates at low pH for averaged relative growth data presented in Fig 4A. Open circles: SARC-I parental control (C) that did not undergo recombination. Colored circles: isolates with highly altered growth that were tested for expression. Horizontal bars: average of triplicates. B) Individual data and recombinant strain identities for averaged expression data presented in Fig 4B. Expression levels are measured relative to non-SARC strain SULG. Open circles: SARC-I parental control (C) that did not undergo recombination. Colored circles: recombinants with highly altered growth that were tested for expression. Grey circles: recombinants with mostly unaltered growth that were tested to determine whether SARC expression levels were retained.

Table S1. Strains used in study.

Name	Strain type	Generations from parental strain	Growth pH	Parental strain identity	Parental strain difference from SULA	GenBank Accession #
SULA	Parental	-	3.0	98/2 (8, 18)	-	CP011057
SARC-B	Partially evolved SARC	50	1.5			CP011055
SARC-C	Fully evolved SARC	120	1.0			CP011056
SULG	Parental	-	3.0	PBL2025 (19)	del(SULA_0784-0827)	XXXXX
SARC-H	Partially evolved SARC	50	1.5			XXXXX
SARC-I	Fully evolved SARC	120	1.0			XXXXX
SULM	Parental	-	3.0	PBL2091 (19)	del(SULA_1707-1708) del(SULA_0784-0827)	XXXXX
SARC-N	Partially evolved SARC	50	1.5			XXXXX
SARC-O	Fully evolved SARC	120	1.0			XXXXX
SUL120	Passaging control without selection	120	3.0	98/2 (8, 18)	-	XXXXX
2027-1…2027-10	Recombination at SULA_2027	~50	3, phenotyping at 1.13	SARC-I	del(SULA_0784-0827), SARC traits	NA
2869-1…2869-9	Recombination at SULA_2869					NA
0895-1..0895-10	Recombination at SULA_0895					NA

NA = not whole-genome resequenced for upload to GenBank

Table S2. Location and type of mutations occurring in SARC strains and SUL120.

Strain	Coordinate	Nucleotide Change	Transition (Ts) or Transversion (Tv)	Amino acid change	ORF #	Annotation	KEGG Pathway
Partially evolved and terminal SARC-C	507,235	C → T	Ts	Synonymous	SULA_0550	Hypothetical protein	NA
Partially evolved and terminal SARC-C	590,768	C → T	Ts	Gly → Asp	SULA_0631	Formate dehydrogenase / Oxidoreductase	ssoa00630 - Glyoxylate and dicarboxylate metabolism
Partially evolved and terminal SARC-C	962,913	G → A	Ts	Thr → Ile	SULA_0993	Amino acid permease	NA
SARC-C	1,253,935	C → A	Tv	Arg → Leu	SULA_1334	Hypothetical protein (predicted MFS superfamily domain)	NA
SARC-C	2,569,782	C → T	Ts	NA	SULA_2764	pseudogene	NA
Partially evolved and terminal SARC-O	1,476,035	C → T	Ts	NA	SULA_1594	Hypothetical protein (predicted transcriptional regulator domain)	NA
Partially evolved and terminal SARC-O	1,476,069	C → T	Ts	NA	NA	Intergenic region	NA
Partially evolved and terminal SARC-O	1,480,239	C → T	Ts	NA	NA	Intergenic region	NA
SARC-O	111,938	C → A	Tv	Arg → Met	SULA_0122	Flagellar protein FlaI	NA
SARC-O	231,532	T → C	Ts	Synonymous	SULA_0260	Pyruvate carboxylase/ Biotin carboxylase(1st step in fatty acid synthesis)	ssoa00630 - Glyoxylate and dicarboxylate metabolism ssoa00680 - Methane metabolism

							ssoa01200 - Carbon metabolism
							ssoa00280 - Valine, leucine and isoleucine degradation
							ssoa00640 - Propanoate metabolism
							ssoa00720 - Carbon fixation pathways in prokaryotes
							ssoa01130 - Biosynthesis of antibiotics
SARC-O	269,889	T → A	Tv	Synonymous	SULA_0304	Hypothetical protein	NA
SARC-O	543,980	A → G	Ts	Synonymous	SULA_0588	Transposase	NA
SARC-O	544,124	A → T	Tv	Synonymous	SULA_0588	Transposase	NA
SARC-O	544,127	A → G	Ts	Synonymous	SULA_0588	Transposase	NA
SARC-O	544,130	A → G	Ts	Synonymous	SULA_0588	Transposase	NA
SARC-O	576,251	A → G	Ts	Lys → Glu	SULA_0618	Transposase	NA
SARC-O	986,237	G → A	Ts	Arg → Lys	SULA_1019	Phosphomethyl-pyrimidine kinase	NA
SARC-O	1,447,126	G → T	Tv	Synonymous	SULA_1555	Transposase	NA
SARC-O	1,447,152	G → A	Ts	Synonymous	SULA_1555	Transposase	NA

SARC-O	1,447,230	G → A	Ts	Synonymous	SULA_1555	Transposase	NA
SARC-O	1,447,235	G → A	Ts	Arg → Lys	SULA_1555	Transposase	NA
SARC-O	1,447,242	G → T	Tv	Synonymous	SULA_1555	Transposase	NA
SARC-O	1,532,625	G → A	Ts	Met → Ile	SULA_1663	Glutamate/proline-tRNA ligase / Prolyl-tRNA synthetase	ssoa00970 - Aminoacyl-tRNA biosynthesis
SARC-O	1,672,086	C → T	Ts	Val → Ile	SULA_1841	Spermidine synthase	ssoa00270 - Cysteine and methionine metabolism ssoa00330 - Arginine and proline metabolism ssoa00410 - beta-Alanine metabolism ssoa00480 - Glutathione metabolism
SARC-O	1,865,679	G → A	Ts	Ala → Thr	SULA_2045	Dephospho-CoA kinase	NA
SARC-O	1,903,384	C → T	Ts	Leu → Phe	SULA_2087	Hypothetical protein	NA
SARC-O	1,991,035	C → T	Ts	NA	SULA_2177	Pseudogene	NA
SARC-O	1,992,437	G → A	Ts	Synonymous	SULA_2180	Transposase	NA
SARC-O	1,992,438	C → T	Ts	His → Tyr	SULA_2180	Transposase	NA
SARC-O	1,992,440	C → T	Ts	His → Tyr	SULA_2180	Transposase	NA
SARC-O	2,251,472	C → T	Ts	NA	SULA_2418	Pseudogene	NA

SARC-O	2,334,041	G → A	Ts	NA	NA	Intergenic region	NA
SARC-O	2,479,423	A → T	Tv	NA	NA	Intergenic region	NA
SARC-O	2,621,131	A → G	Ts	Val → Ala	SULA_2823	Peptidase S53	NA
SUL120	14,502	G→T	Tv	Leu→Ile	SULA_0016	Hypothetical	-
SUL120	169,536	G→A	Ts	Gly→Asp	SULA_0188	type II secretion system protein E	-
SUL120	234,876	T→C	Ts	-	NA	Intergenic	-
SUL120	278,602	T→G	Tv	Trp→Gly	SULA_0313	acyl-CoA dehydrogenase	-
SUL120	288,650	C→T	Ts	Val→Ile	SULA_0321	alpha/beta hydrolase	-
SUL120	307,152	C→T	Ts	-	NA	Intergenic	-
SUL120	307,153	A→C	Tv	-	NA	Intergenic	-
SUL120	307,430	A→G	Ts	Tyr→Cys	SULA_0338	transposase	-
SUL120	312,543	G→T	Tv	Arg→Ser	SULA_0343	Hypothetical	-
SUL120	340,210	T→C	Ts	Synonymous	SULA_370	Transposase	-
SUL120	340,259	T→C	Ts	Synonymous	SULA_370	Transposase	-
SUL120	394,425	C→T	Ts	Gln→STOP	SULA_0431	Phenylacetate-CoA ligase	-
SUL120	412,164	G→A	Ts	Ile→Met	SULA_0453	ATPase P	-
SUL120	435,756	A→G	Ts	-	SULA_0478 (pseudogene)	Pseudo N acetylglocosamine 6 phosphate deacetylase	-
SUL120	507,235	C→T	Ts	Synonymous	SULA_0550	Hypothetical	-
SUL120	590,768	C→T	Ts	Gly→Asp	SULA_0631	Oxidoreductase	-
SUL120	606,818	A→C	Tv	Synonymous	SULA_0648	Sugar ABC transporter permease	-
SUL120	637,799	C→T	Ts	-	SULA_0682 (pseudogene)	Pseudo transposase	-
SUL120	657,394	A→T	Tv	-	SULA_0705 (pseudogene)	Pseudo transposase	-

SUL120	719,918	T→C	Ts	Synonymous	SULA_0770	Transposase	-
SUL120	720,056	T→C	Ts	Synonymous	SULA_0770	Transposase	-
SUL120	756,842	G→A	Ts	-	NA	Intergenic	-
SUL120	762,524	A→T	Tv	-	SULA_0800	Pseudo transposase	-
SUL120	764,518	G→A	Ts	Leu→Phe	SULA_0805	Transposase	-
SUL120	765,535	A→C	Tv	-	NA	Intergenic	-
SUL120	769,242	T→C	Ts	-	NA	Intergenic	-
SUL120	769,782	T→G	Tv	-	NA	Intergenic	-
SUL120	769,817	G→A	Ts	-	NA	Intergenic	-
SUL120	774,072	C→T	Ts	Synonymous	SULA_0812	Transposase	-
SUL120	779,637	G→A	Ts	-	SULA_0815	Pseudo Hypo	-
SUL120	780,137	T→C	Ts	Synonymous	SULA_0818	Beta-glucosidase	-
SUL120	786,643	A→T	Tv	-	NA	Intergenic	-
SUL120	789,437	A→G	Ts	Synonymous	SULA_0824	ABC transporter permease	-
SUL120	789,460	T→A	Tv	Asn→Tyr	SULA_0824	ABC transporter permease	-
SUL120	789,473	G→A	Ts	Synonymous	SULA_0824	ABC transporter permease	-
SUL120	789,522	T→A	Tv	Synonymous	SULA_0825	ABC transporter permease	-
SUL120	789,537	G→A	Ts	Synonymous	SULA_0825	ABC transporter permease	-
SUL120	789,540	T→G	Tv	Synonymous	SULA_0825	ABC transporter permease	-
SUL120	789,570	A→G	Ts	Synonymous	SULA_0825	ABC transporter permease	-
SUL120	789,576	A→G	Ts	Synonymous	SULA_0825	ABC transporter permease	-
SUL120	789,963	C→T	Ts	Synonymous	SULA_0825	ABC transporter permease	-
SUL120	789,973	T→C	Ts	Synonymous	SULA_0825	ABC transporter permease	-

SUL120	873,595	A→G	Ts	Synonymous	SULA_0906	Fe binding protein	-
SUL120	882,546	G→C	Tv	Thr→Arg	SULA_0914	Hypothetical	-
SUL120	896,436	T→C	Ts	Synonymous	SULA_0926	Transposase	-
SUL120	962,913	G→A	Ts	Thr→Ile	SULA_0933	Amino acid permease	-
SUL120	977,171	G→T	Tv	Met→Ile	SULA_1007	Dolichol-phosphate mannosultransferase	-
SUL120	1,006,331	A→C	Tv	Asp→Ala	SULA_1038	FAD-dependent oxidoreductase	-
SUL120	1,019,944	G →A	Ts	Arg→Gln	SULA_1055	Transposase	-
SUL120	1,060,522	A→T	Tv	Phe→Leu	SULA_1108	Phosphoglycolate phosphate	-
SUL120	1,101,605	A→G	Ts	-	NA	Intergenic	-
SUL120	1,101,716	A→C	Tv	Asn→Thr	SULA_1151	Transposase	-
SUL120	1,101,728	T→A	Tv	Ile→Asn	SULA_1151	Transposase	-
SUL120	1,117,782	C→T	Ts	Synonymous	SULA_1171	AsnC family transcriptional regulator	-
SUL120	1,134,701	G→T	Tv	Arg→Ser	SULA_1192	ATPase AAA	-
SUL120	1,158,040	C→T	Ts	Gly→Asp	SULA_1216	S-adenosylmethionine synthetase	-
SUL120	1,158,784	G→T	Tv	Pro→His	SULA_1216	S-adenosylmethionine synthetase	-
SUL120	1,176,737	A→G	Ts	Val→Ala	SULA_1243	DNA directed RNA pol subunit A	-
SUL120	1,253,935	C→A	Tv	Arg→Leu	SULA_1334	Hypothetical	-
SUL120	1,488,465	G→A	Ts	-	NA	Intergenic	-
SUL120	1,533,012	G→A	Ts	-	NA	Intergenic	-
SUL120	1,693,808	G→A	Ts	Glu→Lys	SULA_1870	GTPase	-
SUL120	1,847,144	C→T	Ts	Synonymous	SULA_2027	Oxidoreductase	-
SUL120	1,847,693	G→A	Ts	Synonymous	SULA_2027	Oxidoreductase	-
SUL120	1,985,046	A→G	Ts	Synonymous	SULA_2172	Hypothetical	-
SUL120	2,017,967	T→C	Ts	Synonymous	SULA_2209	Aldehyde oxidase	-
SUL120	2,018,193	G→A	Ts	Gly→Ser	SULA_2209	Aldehyde oxidase	-
SUL120	2,035,452	A→C	Tv	-	SULA_2227	Pseudo transposase	-

SUL120	2,035,553	T→C	Ts	-	NA	Intergenic	-
SUL120	2,035,557	C→T	Ts	-	NA	Intergenic	-
SUL120	2,035,745	T→C	Ts	Synonymous	SULA_2228	transposase	-
SUL120	2,035,748	G→T	Tv	Synonymous	SULA_2228	transposase	-
SUL120	2,035,770	T→C	Ts	Synonymous	SULA_2228	transposase	-
SUL120	2,035,832	T→C	Ts	Synonymous	SULA_2228	transposase	-
SUL120	2,044,050	G→T	Tv	Asp→Tyr	SULA_2238	VapB-type antitoxin	-
SUL120	2,060,463	C→A	Ts	Glu→STOP	SULA_2256	Glutamine amidotransferase	-
SUL120	2,078,239	A→G	Ts	Synonymous	SULA_2272	Transposase	-
SUL120	2,078,246	G→A	Ts	Val→Ile	SULA_2272	Transposase	-
SUL120	2,078,248	G→A	Ts	Val→Ile	SULA_2272	Transposase	-
SUL120	2,080,644	A→G	Ts	Tyr→Cys	SULA_2274	Hypothetical	-
SUL120	2,099,267	G→A	Ts	Asn→Asn	SULA_2290	Hypothetical	-
SUL120	2,099,278	C→T	Ts	Ala→Thr	SULA_2290	Hypothetical	-
SUL120	2,105,230	A→G	Ts	-	NA	Intergenic	-
SUL120	2,145,623	C→T	Ts	Synonymous	SULA_2321	Transposase	-
SUL120	2,213,187	A→G	Ts	-	SULA_2382	Pseudo Hypo	-
SUL120	2,216,941	T→C	Ts	Pro→Phe	SULA_2388	Hypothetical	-
SUL120	2,216,942	C→T	Ts	Pro→Phe	SULA_2388	Hypothetical	-
SUL120	2,216,943	T→C	Ts	Pro→Phe	SULA_2388	Hypothetical	-
SUL120	2,216,960	C→T	Ts	Ala→Val	SULA_2388	Hypothetical	-
SUL120	2,286,621	G→A	Ts	-	NA	Intergenic	-
SUL120	2,287,079	G→A	Ts	Val→Ile	SULA_2451	Transposase	-
SUL120	2,287,120	C→T	Ts	Leu→Leu	SULA_2451	Transposase	-
SUL120	2,287,131	G→T	Tv	Arg→Met	SULA_2451	Transposase	-
SUL120	2,287,140	G→A	Ts	Ser→Asn	SULA_2451	Transposase	-
SUL120	2,287,240	G→A	Ts	Synonymous	SULA_2451	Transposase	-
SUL120	2,288,511	T→C	Ts	Synonymous	SULA_2452	Amidohydrolase	-
SUL120	2,334,580	T→C	Ts	-	SULA_2496	Pseudo transposase	-
SUL120	2,359,468	T→C	Ts	-	SULA_2526	Pseudo transposase	-
SUL120	2,414,630	G→T	Tv	Arg→Ser	SULA_2584	Hypothetical	-

SUL120	2,417,603	G→A	Ts	Pro→Ser	SULA_2586	Potassium transporter TrkA	-
SUL120	2,428,508	C→A	Tv	-	SULA_2601	intergenic	-
SUL120	2,437,816	C→A	Tv	Arg→Ser	SULA_2610	Hypothetical	-
SUL120	2,437,844	C→T	Ts	Gly→Asp	SULA_2610	Hypothetical	-
SUL120	2,438,247	T→G	Tv	Synonymous	SULA_2611	Transposase	-
SUL120	2,438,314	A→C	Tv	-	NA	intergenic	-
SUL120	2,442,149	A→G	Ts	-	NA	Intergenic	-
SUL120	2,442,176	A→G	Ts	-	NA	Intergenic	-
SUL120	2,442,320	T→G	Tv	-	NA	Intergenic	-
SUL120	2,442,337	A→G	Ts	-	NA	Intergenic	-
SUL120	2,442,463	T→C	Ts	Val→Ala	SULA_2616	Transposase	-
SUL120	2,442,681	A→G	Ts	Asn→Ser	SULA_2616	Transposase	-
SUL120	2,447,853	G→C	Tv	-	SULA_2621	Pseudo amino acid permease	-
SUL120	2,454,054	C→A	Tv	Gly→Val	SULA_2627	Amidohydrolase	-
SUL120	2,460,502	A→C	Tv	-	NA	Intergenic	-
SUL120	2,476,229	G→A	Ts	-	SULA_2654	Pseudo aldehyde dehydrogenase	-
SUL120	2,484,873	C→G	Tv	Asn→Lys	SULA_2665	Transposase	-
SUL120	2,484,875	T→C	Ts	Leu→Ser	SULA_2665	Transposase	-
SUL120	2,484,876	G→A	Ts	Leu→Ser	SULA_2665	Transposase	-
SUL120	2,484,909	T→C	Ts	Synonymous	SULA_2665	Transposase	-
SUL120	2,484,912	T→C	Ts	Synonymous	SULA_2665	Transposase	-
SUL120	2,484,918	G→T	Tv	Synonymous	SULA_2665	Transposase	-
SUL120	2,493,794	G→A	Ts	Synonymous	SULA_2675	Transposase	-
SUL120	2,493,797	A→G	Ts	Synonymous	SULA_2675	Transposase	-
SUL120	2,569,783	C→T	Ts	-	SULA_2764	Pseudo transposase	-
SUL120	2,581,994	T→C	Ts	Lys→Asn	SULA_2777	Transposase	-
SUL120	2,582,006	A→T	Tv	Synonymous	SULA_2777	Transposase	-
SUL120	2,582,010	G→A	Ts	Thr→Val	SULA_2777	Transposase	-
SUL120	2,582,011	T→C	Ts	Thr→Val	SULA_2777	Transposase	-

SUL120	2,582,320	A→C	Tv	Phe→Val	SULA_2777	Transposase	-
SUL120	2,582,323	C→T	Ts	Asp→Asn	SULA_2777	Transposase	-
SUL120	2,582,333	C→T		Lys→Lys	SULA_2777	Transposase	-
SUL120	2,582,355	C→G	Tv	Gly→Thr	SULA_2777	Transposase	-
SUL120	2,582,356	C→T	Ts	Gly→Thr	SULA_2777	Transposase	-
SUL120	2,582,357	C→T	Ts	Glu→Lys	SULA_2777	Transposase	-
SUL120	2,582,359	C→T	Ts	Glu→Lys	SULA_2777	Transposase	-
SUL120	2,582,441	T→C	Ts	Lys→Arg	SULA_2777	Transposase	-
SUL120	2,582,442	T→C	Ts	Lys→Arg	SULA_2777	Transposase	-
SUL120	2,582,447	A→T	Tv	Synonymous	SULA_2777	Transposase	-
SUL120	2,582,470	A→G	Ts	Synonymous	SULA_2777	Transposase	-
SUL120	2,597,755	T→C	Ts	-	NA	Intergenic	-
SUL120	2,654,491	T→C	Ts	-	SULA_2848	Pseudo transposase	-
	INDELS						
SUL120	116,284-292	-	-	-	SULA_0127	Flagellin	-
SUL120	207,040-43	-	-	-	SULA_0234	Pseudo sugar isomerase	-
SUL120	480,128	-	-	-	SULA_0522	Pseudo hypothetical	-
SUL120	480,156	-	-	-	SULA_0522	Pseudo hypothetical	-
SUL120	834,832	-	-	-	SULA_0869	SirA family protein	-
SUL120	884,571	-	-	-	SULA_0916	Hypothetical	-
SUL120	1,106,104-5	-	-	-	NA	Intergenic	-
SUL120	1,491,571-77	-	-	-	SULA_1620	Hypothetical	-
SUL120	1,469,652	-	-	-	SULA_1590	Conjugal transfer protein TraG	-
SUL120	2,010,786	-	-	-	SULA_2198	DDE endonuclease	-
SUL120	2,148,521	-	-	-	SULA_2323	Pseudo CRISPR Cas1	-
SUL120	2,225,549-552	-	-	-	NA	Intergenic	-
SUL120	2,384,883	-	-	-	NA	Intergenic	-
SUL120	2,405,143	-	-	-	NA	Intergenic	-
SUL120	2,425,586-7	-	-	-	SULA_2595	Hemerythrin	-

SUL120	2,442,708-710	-	-	-	SULA_2617	Hypothetical	-
SUL120	2,472,839	-	-	-	NA	Intergenic	-
SUL120	2,597,744-5	-	-	-	NA	Intergenic	-
SUL120	2,523,446-448	-	-	-	NA	Intergenic	-

Table S3. Identity of SARC transcriptome genes and fold-change in expression compared to parental strains.

Gene ORF#	Annotation	Partially evolved SARC-C	SARC-C	Partially evolved SARC-I	SARC-I	Partially evolved SARC-O	SARC-O	Average FC for 3 fully evolved strains (Fig 2)	Standard error for average	Non-acid stress control SARC-O	Putative involvement in acid resistance
Conserved in all partially and fully adapted lineages - Altered 5 fold or more											
SULA_0674	Radical SAM domain protein	5.2	43.79	27.79	11.26	7.88	4.87	**19.97**	12.05	8.48	
SULA_2620	Hypothetical protein - membrane protein	1.3	6.74	85.19	9.37	10.13	108.21	**41.44**	33.39	3.59	Membrane permeability
SULA_2869	Pyruvate ferredoxin/flavodoxin oxidoreductase, delta subunit	2.7	17.03	290.46	8.71	2.66	8.35	**11.36**	2.84	16.69	Central energy metabolism
Conserved in all partially and fully adapted lineages - Altered more than 2 fold, less than 5 fold											
SULA_0339	Hypothetical protein	1.61	2.54	6.02	4.88	13.07	2.96	**3.46**	0.72	4.16	
SULA_0352	Sso7D	1.23	4.77	5.18	5.73	2.46	2.21	**4.24**	1.05	10.1	Chromatin
SULA_0394	Hypothetical protein	1.73	1.46	5.4	5.11	10.28	3.44	**3.34**	1.06	4.9	
SULA_0486	Hypothetical protein	1.33	2.07	4.11	3.65	11	2.57	**2.76**	0.47	2.05	
SULA_0765	Nucleoside-diphosphate sugar epimerase	1.53	2.01	3.47	2.6	6.81	2.11	**2.24**	0.18	2.19	
SULA_1632	Hypothetical protein	-1.7	-15.29	-3.66	-2.96	-0.41	-2.32	**-6.86**	4.22	-2.68	
SULA_2209	Putative Xanthine dehydrogenase/CO dehydrogenase/Aldehyde oxidase (XdhA domain)	2.07	18.95	23.23	8.58	2.32	1.72	**9.75**	5.01	13.34	Oxidation resistance (14)
SULA_2423	Hypothetical protein	2.14	2.35	47.74	4.3	2.87	7.08	**4.58**	1.37	4.64	
SULA_2448	MFS transporter	1.6	1.95	169.34	2.02	6.35	3.16	**2.37**	0.39	7.49	Membrane permeability
SULA_2630	Hypothetical protein	2.29	2.81	4.57	5.52	5.13	3.95	**4.09**	0.79	2.74	
SULA_2631	Putative Sulfocyanin	2.4	3.68	3.22	3.84	2.22	6.77	**4.76**	1	2.23	
SULA_2640	Peptidase	1.83	2.56	2.84	3.14	2.7	1.68	**2.46**	0.42	4.6	
SULA_2642	Hypothetical protein	1.67	5.47	23.66	6.35	0.22	2.16	**4.66**	1.27	5.34	

SULA_2646	GMP-synthase	2.12	6.13	9.53	4.23	1.88	2.74	**4.37**	0.98	4.42	Oxidative resistance (15,16)
SULA_2767	Thiamine biosynthesis protein ThiC	2.06	29.46	60.95	7.56	0.18	2.05	**13.02**	8.37	33.27	Oxidation resistance (17)
SULA_2845	Hypothetical protein	1.49	2.23	4.22	3.24	21.25	2.88	**2.79**	0.3	2.71	
SULA_2872	AsnC family transcriptional regulator	1.37	1.89	8.71	1.97	6.11	10.32	**4.73**	2.8	6.81	
SULA_2873	Glycerol kinase	1.87	12.69	4.67	3.2	2.02	4.88	**6.92**	2.92	18.03	
Conserved in all fully adapted - Altered more than 5 fold											
SULA_0452	Hypothetical protein	-1.49	19.72	7.05	2.56	0.41	13.26	**11.85**	5	3.11	
SULA_0698	Sulfite reductase	-2.69	-9.3	6.68	-1.1	0.07	-87.12	**-32.51**	27.41	-2.59	
SULA_0844	Sugar ABC transporter	-4.44	-35.89	-	-1.29	-0.08	-8.7	**-15.29**	10.52	-3.12	Membrane permeability
SULA_0871	Hypothetical protein	-2.15	-23.15	-	-2.89	-0.67	-16.28	**-14.11**	5.95	-2.21	
SULA_0872	Terminal quinol oxidase subunit	-2.07	-10.2	-	-1.95	-0.52	-18.3	**-10.15**	4.72	-2.26	Proton extrusion
SULA_2569	Hypothetical protein	-3.39	56.96	5.1	24.35	-	40.74	**40.68**	9.41	7.82	
SULA_2749	superfamily 1 DNA and RNA helicase subunit	-	2.98	19.16	9.31	5.78	18.12	**10.13**	4.39	9.67	
SULA_2870	Pyruvate ferredoxin/flavodoxin oxidoreductase, delta subunit	2.18	11.3	237.56	6.9	-	5.51	**7.9**	1.74	18.45	Energy metabolism
Conserved in all fully adapted - Altered more than 2 fold, less than 5 fold											
SULA_0152	FAD-dependent oxidoreductase	-1.46	-1.46	-	-1.79	-2.21	-6.2	**-3.15**	1.53	-3.1	
SULA_0153	Hypothetical protein	-1.83	-14.17	-	-1.26	-0.21	-2.84	**-6.09**	4.07	-2.75	
SULA_0202	AsnC family transcriptional regulator	-	1.77	3.71	3.52	4.03	3.01	**2.77**	0.52	3.42	
SULA_0286	Hypothetical protein	-1.86	-4.78	-	-1.72	-	-3.68	**-3.4**	0.89	-2.03	
SULA_0385	Twitching motility protein PilT	1.75	1.67	-	2.08	1.72	2.53	**2.09**	0.25	3.53	
SULA_0446	Rubrerythrin	1.31	13	-	1.63	0.37	2.6	**5.74**	3.64	2.56	
SULA_0566	Electron transfer flavoprotein subunit alpha	-1.27	3.68	3.09	1.28	0.4	1.76	**2.24**	0.73	2.68	Proton extrusion
SULA_0600	Component of anaerobic dehydrogenase	-1.31	-3.47	4.36	-1.35	0.55	-18.61	**-7.81**	5.43	-3.41	

SULA_0602	Sulfide reductase	-2.18	-2.14	9.18	-1.57	0.4	-2.93	**-2.21**	0.4	-2.68	
SULA_0681	Pyridine nucleotide-disulfide oxidoreductase	1.17	1.47	-	1.75	2.34	2.98	**2.07**	0.46	3.04	
SULA_0759	Hypothetical protein	-6.29	-9.26	-	-3.25	2.02	-2.03	**-4.85**	2.23	-2.87	
SULA_0845	Sugar ABC transporter	-4.08	-7.83	-	-2.2	-0.47	-10.46	**-6.83**	2.44	-4	Membrane permeability
SULA_0846	Sugar ABC transporter ATP-binding protein	-3.19	-3.28	-	-1.99	-0.82	-2.18	**-2.49**	0.4	-2.6	Membrane permeability
SULA_0854	Twitching motility protein PilT	-2.19	-1.76	-5.03	-5.26	-7.01	-1.62	**-2.88**	1.19	-2.68	
SULA_0860	Hypothetical protein	-1.58	-4.82	-	-1.44	-0.35	-3.03	**-3.1**	0.98	-2.88	
SULA_0865	Sulfur transporter	-1.47	-3.32	-	-2.15	-1.49	-8.95	**-4.81**	2.1	-1.73	Membrane permeability
SULA_0870	Hypothetical protein	-1.86	-5.68	-	-2.46	-1.48	-5.42	**-4.52**	1.03	-1.72	
SULA_0914	Hypothetical protein	-1.68	-6.81	-	-1.34	-0.72	-3.09	**-3.75**	1.61	-1.77	
SULA_0916	Hypothetical protein	-2	-16.24	-	-1.62	-0.26	-2.5	**-6.79**	4.73	-1.66	
SULA_0994	Putative Dolichyl-phosphate-mannose-protein mannosyltransferase	-2.12	-6.41	-	-1.55	-0.27	-1.82	**-3.26**	1.58	-3.63	
SULA_1059	Putative cytochrome or quinol oxidase	-	1.79	14.36	4.14	2.07	5.99	**3.97**	1.22	1.85	Proton extrusion
SULA_1305	ABC transporter related ATPase	-1.19	-3.1	-	-1.31	-2.29	-4.52	**-2.97**	0.93	-1.8	Membrane permeability
SULA_1399	Glucosamine--fructose-6-phosphate aminotransferase	-1.96	-2.47	-	-1.75	-0.8	-2.1	**-2.1**	0.21	-2.22	
SULA_1481	TrmB family transcriptional regulator	1.18	3.01	9.3	2.61	1.52	4.55	**3.39**	0.59	4.01	
SULA_1770	Hypothetical protein	-1.21	-22.71	-	-1.86	-0.22	-2.03	**-8.87**	6.92	-1.82	
SULA_2067	Glycine cleavage system protein H	-1.29	4.94	-	1.1	-3.51	1.74	**2.59**	1.19	2.07	
SULA_2217	ATPase	-3.08	-2.29	-	-1.66	-0.76	-2.59	**-2.18**	0.27	-3.83	
SULA_2262	Peptide ABC transporter permease	-1.22	-2.28	-	-1.38	-0.39	-2.19	**-1.95**	0.29	-3.17	Membrane permeability
SULA_2263	Peptide ABC transporter permease	-1.22	-1.56	-	-1.53	-0.68	-3.22	**-2.1**	0.56	-3.42	Membrane permeability

SULA_2279	Putative Xanthine and CO dehydrogenase maturation factor (XdhC domain)	-	5.29	10.86	4.27	6.85	8.09	**5.88**	1.14	3.92	Oxidation resistance (14)
SULA_2343	Hypothetical protein	-	7.59	4.13	1.48	0.42	3.4	**4.16**	1.81	2.6	
SULA_2345	Putative DNA polymerase subunit	-	1.56	6.79	3.01	5.89	2.73	**2.43**	0.44	5.33	
SULA_2356	Hypothetical protein	-1.16	3.51	-	2.38	-2.64	2.78	**2.89**	0.33	3.09	
SULA_2361	Metallophosphoesterase	2.05	4.81	6.82	3.38	-	4	**4.06**	0.42	4.54	
SULA_2396	Homogentisate 1,2-dioxygenase	-1.48	2.94	4.31	3.34	1.4	1.96	**2.75**	0.41	3.12	
SULA_2454	D-arabinose dehydrogenase	-	7.16	3.47	1.51	0.65	3.14	**3.94**	1.68	3.39	
SULA_2591	Putative SirA / TusA like sulfurtransferase	1.17	13.84	12.87	3.22	-	4.91	**7.32**	3.29	2.47	
SULA_2644	Protease	1.14	2.4	-	1.98	2.87	2.24	**2.21**	0.12	3.08	
SULA_2657	Hypothetical protein	-	3.69	4.64	3.54	3.52	17.1	**8.11**	4.5	2.6	
SULA_2693	Hypothetical protein	-	5.82	9.45	2.64	5.93	42.53	**16.99**	12.8	4.02	
SULA_2850	Amidohydrolase	-	1.6	4.35	2.19	7.44	2.39	**2.06**	0.24	2.93	
SULA_2859	Peroxiredoxin	1.79	7.2	-	1.66	2.8	7.02	**5.29**	1.82	3.64	
SULA_2860	Aspartyl/glutamyl-tRNA amidotransferase	-1.25	4.66	3.28	1.93	2.11	3.77	**3.46**	0.81	2.51	
SULA_2865	Lactate permease	-1.94	4.38	-	1.2	1.54	2.28	**2.62**	0.93	2.24	
SULA_2871	Pyruvate synthase	1.97	5.73	114.48	3.08	-	3.97	**4.26**	0.78	10.15	Central energy metabolism
SULA_2889	Putative Xanthine dehydrogenase/CO dehydrogenase/Aldehyde oxidase (XdhA)	-	2.1	6.15	2.35	2.23	32.9	**12.45**	10.23	9.8	Oxidation resistance (14)
SULA_2027	NADH/Ubiquinone/plastoquinone (complex I)	9.3	7.73	7.34	12.27	7.36	-5.15	4.95	5.217292	8.66	Proton extrusion

Table S4. Identity and location of transposition events occurring in SARC and SUL120 strains

Transposition type	Family	IS element	Size of transposon (bp)	Original IS element ORF #	SULA event site (bp)	SUL120 event site (bp)	Insertion site ORF#	Disrupted gene function
SARC-O transpositions								
Insertion	IS4	ISC1058	1058	SULA_0801	112977	NA	SULA_0123	Flagellar accessory protein FlaH
SUL120 Transpositions								
Insertion	IS5	ISC1234	1232	SULA_0292, 0348, 2353, 2457, 2639	170000	170013	SULA_0188	Type II secretion system protein E
MITE Insertion	-	-	98	-	735177	745283	SULA_0784	Oxidoreductase
Insertion	IS4	ISC1439	1446	SULA_0338, 0710	625278	632579	SULA_0669	Long-chain fatty acid CoA ligase
Insertion	IS4	ISC1359	1366	SULA_0256	720386	729152	SULA_0770	Transposase
Insertion	IS5	ISC1234	1238	SULA_0292, 0348, 2353, 2457, 2639	1120366	1122452	SULA_1175	Acetyl-lysine deacetylase, putative FabG
Replicative transposition	IS4	ISC1439	1446	SULA_0812	1390186	1392356	SULA_1490	Transposase Pseudogene
Replicative transposition	IS4	ISC1225	1102	SULA_0728	1390187	1393931	SULA_1490	Transposase Pseudogene
Replicative transposition	IS4	ISC1225	1230	SULA_2307	1907500	1870472	SULA_2091	CoA-transferase
Replicative transposition	IS4	ISC1439 partial	733	SULA_0131	2286721	2263885	SULA_2451	Transposase

Insertion	IS4	ISC1439	1444	SULA_0138, 0270, 0344, 0712, 0880, 2847	1748403	1711286	SULA_1933	Carboxylate-amine ligase
Insertion	IS4	ISC1058	1066	SULA_0981, 2705, 1154	2020457	1983468	SULA_2212	Hypothetical protein
Insertion	IS4	ISC1439	1448	SULA_1394, 2650	2488589	2468634	Intergenic (SULA_2669-2670)	NA
Insertion	IS4	ISC1225	1229	SULA_1394, 2650	2581673	2563500	SULA_2776	Hypothetical protein
MITE Deletion	-	-	335	-	262726	263900	-	-
MITE Deletion			68		2114977	2090505	-	-
MITE Deletion			70		2156460	2131928	-	-
Non-replicative transposition	IS4	ISC1439	1162	SULA_2451	2286715	126361	Intergenic (SULA_0138-0139)	NA
Non-replicative transposition	IS4	ISC1439	1443	SULA_0812	773243	2359687	SULA_2526	Transposase Pseudogene
Non-replicative transposition	IS4	ISC1439	1163	SULA_2338	2172274	2536625	SULA_2723	AMP-dependent synthetase
Non-replicative transposition & Inversion	IS4	ISC1439	1428	SULA_0797	755442	2167837	Intergenic (SULA_2333-2334)	NA
Non-replicative transposition & Inversion	IS4	ISC1439	1160	Part of SULA_2653	2475060	2172243	Intergenic (SULA_2337-2338)	NA
Non-replicative transposition & Inversion	IS4	ISC1439	267	Part of SULA_2653	2474793	2536625	SULA_2723	AMP-dependent synthetase

Table S5. Fold change in expression for DNA-repair and replication genes for SARC relative to parental strains

ORF #	Gene product	Partially evolved SARC-C	SARC-C	Partially evolved SARC-I	SARC-I	Partially evolved SARC-O	SARC-O	Non-acid stress control SARC-O
SULA_1812	XPF	-	-1.88	-	3.01	11.32	-	2.48
SULA_1331	XPD	-	-3.39	-	2.78	21.28	-	-
SULA_1969	XPB1	-	-2.34	-	2.69	14.81	-	-
SULA_1498	XPB2	-	-	-	1.74	2.11	0.2	-
SULA_1499	Bax1	-	-1.8	3.14	2.81	6.18	0.19	-
SULA_0265	PhrB	-	-	-	2.09	1.84	-	-
SULA_0242	Dpo4	-	-1.7	-	-	3.76	-6.63	-
SULA_1266	RadA	-	-	-	1.27	2.03	-	-
SULA_0535	UDG1	-1.28	-	-	1.31	2.15	-	-
SULA_0087	UDG2	1.12	1.57	-	1.29	-	-5.63	-
SULA_0283	UDG3	1.37	-3.01	-	1.82	3.44	-13.07	-
SULA_1128	UDG4	1.17	-	-	2.77	9.11	-	-
SULA_1194	XPG/Fen1	1.16	1.79	-	1.55	-1.23	-5.85	-
SULA_0063	NurA	-	0.54	-	2.43	2.23	-2.69	-
SULA_0064	Rad50	-	1.57	3.02	4.08	1.47	1.85	-
SULA_0065	Mre11	-	-	-	1.94	1.51	-1.7	-
SULA_0066	HerA	-	-	-	1.91	1.63	1.56	-
SULA_0257	Hjm	1.16	1.82	-	1.82	-1.54	-7.1	2.35
SULA_1668	Hjc	-1.14	2.02	-	1.52	1.42	-	-
SULA_1973	TopR1	-	0.26	-	1.17	2.25	0.04	-
SULA_1442	TopR2	-	-1.41	-	1.25	1.36	0.16	-
SULA_1920	TopIII (TopA)	1.37	-2.12	-	2.4	3.99	0.23	-

Table S6. Primers used in this work.

Primer	Sequence 5'-3'	Application
0895 F	ATAGTCGACTCCGAATACCTCATCAGAGCTT	amplification of SULA_0895 for vector construction
0895 R	ATAGCATGCACCAGCCTTCGTATACTGTCT	
2027 F	ATAGTCGACTTCTCTTTTCTCCAAGGA	amplification of SULA_2027 for vector construction
2027 R	ATAGCATGCCAATATTTCAATTCATCG	
2869 F	ATAGTCGACTCCTATTATTTTTTATT	amplification of SULA_2869 for vector construction
2869 R	ATAGCATGCTCCGTGAACATCGATTC	
0895q F	GGATCTGCTGAAGGAAGCGT	qRTPCR of SULA_0895 for expression analysis
0895q R	AGCCCAAGAGTTTGCGCTAT	
2027q F	TCACAACGTTTTTCCTATTGGTGA	qRTPCR of SULA_2027 for expression analysis
2027q R	AAGCATTGATGATGCGAAGGA	
2869q F	GCATGGCAAGGAGTTCAACC	qRTPCR of SULA_2869 for expression analysis
2869q R	GTCCCGAAAATGATGGTTCAGT	
2002q F	TTCCAGCATAATGGGAATAGGAA	qRTPCR of SULA_2002 for expression analysis
2002q R	GCCGCACCTATTGGGTGATA	
2027-valF	CCCTATTTTGCCTATTTTCCTTCC	Amplification of SULA_2027 in SULA_2027 HR isolates for sequence validation
2027-valR	GCCTTCCATCTAGGAATCTCA	
2869-valF	TTGCCAAAGATTCCAGAAACGA	Amplification of SULA_2869 in SULA_2869 HR isolates for sequence validation
2869-valR	HCAACAACGTCTACATCGGC	
2027-seqF	GGTTAGATGAACTAAAAGGGTG	Sequence validation of SULA_2027 HR isolates
2027-seqR	AGGATCACCAGTAAAAGTATACCGA	
2869-seqR	TGCCCATAGTCCCGAAAATGA	Sequence validation of SULA_2869 HR isolates
2869-seqR2	TGCTGCATGAGCTACTGCTT	
INV1-5'-F	TGGGTAGATGTTAATCCCTTAGC	Size validation for 5' region of Inversion 1
INV1-5'-R	ACGGGATAAAGAAGTTATTCCATGT	
INV1-3'-F	AGATCTTGTGGAAGCAGATGGT	Size validation for 3' region of Inversion 1
INV1-3'-R	AGTGGAGAAATAGAAGGAAAGGTTG	
INV2-5'-F	GTATGGCGAGCAAGATTGGC	Size validation for 5' region of Inversion 2
INV2-5'-R	ACGGAACTAATTACACTGCGA	
INV2-3'-F	ATTTATGACCCGGGGCACTTT	Size validation for 3' region of Inversion 2
INV2-3'-R	ACAATTCCCCCGTTTAGTGT	
INV3-5'-F	TCAGACAAAACCGAGAGTAATCCT	Size validation for 5' region of Inversion 3
INV3-5'-R	AGGAGGAGAAGGTTGAGCCA	
INV3-3'-F	GGAGGCGGTTAATGGAGGAG	Size validation for 3' region of Inversion 3
INV3-3'-R	ACATGGCTAAAGCCGGAGTA	
INV4-5'-F	TCGCAAAAGAAGCCCGAATC	Size validation for 5' region of Inversion 4
INV4-5'-R	TGCTCAAATCCCTCATGCACT	
INV4-3'-F	ACGAATCAATAAGGTGTGGGGA	Size validation for 3' region of Inversion 4
INV4-3'-R	TGTTACGCACTGTTTTCAAGAAACT	

Supplementary Information References

1. McCarthy S*, et al.* (2016) Expanding the Limits of Thermoacidophily in the Archaeon Sulfolobus solfataricus by Adaptive Evolution. *Appl Environ Microbiol* 82(3):857-867.
2. Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357-359.
3. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987-2993.
4. Bankevich A*, et al.* (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455-477.
5. Kolmogorov M, Raney B, Paten B, & Pham S (2014) Ragout-a reference-assisted assembly tool for bacterial genomes. *Bioinformatics* 30(12):i302-309.
6. Chin CS*, et al.* (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10(6):563-569.
7. Garrison E (Vcflib, a simple C++ library for parsing and manipulating VCF files.
8. Mu WB*, et al.* (2016) Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J Mol Diagn* 18(6):923-932.
9. Darling AE, Mau B, & Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5(6):e11147.
10. McCarthy S*, et al.* (2015) Complete Genome Sequence of Sulfolobus solfataricus Strain 98/2 and Evolved Derivatives. *Genome Announc* 3(3).
11. Rausch T*, et al.* (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28(18):i333-i339.
12. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
13. Huang DW, Sherman BT, & Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44-57.
14. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* 57(1):289-300.
15. Huang da W, Sherman BT, & Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37(1):1-13.
16. Bailey TL & Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28-36.
17. Noguchi K*, et al.* (2010) Hydrogenase-3 Contributes to Anaerobic Acid Resistance of Escherichia coli. *Plos One* 5(4).
18. Lim JK, Mayer F, Kang SG, & Muller V (2014) Energy conservation by oxidation of formate to carbon dioxide and hydrogen via a sodium ion current in a hyperthermophilic archaeon. *Proc Natl Acad Sci U S A* 111(31):11497-11502.
19. Maezato Y, Dana K, & Blum P (2011) Engineering Thermoacidophilic Archaea using Linear DNA Recombination. *Strain Engineering: Methods and Protocols* 765:435-445.
20. Zhang C*, et al.* (2014) Analysis of differential gene expression and novel transcript units of ovine muscle transcriptomes. *PLoS One* 9(2):e89817.
21. Rolfsmeier M & Blum P (1995) Purification and characterization of a maltase from the extremely thermophilic crenarchaeote Sulfolobus solfataricus. *J Bacteriol* 177(2):482-485.
22. Schelert J*, et al.* (2004) Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption. *J Bacteriol* 186(2):427-437.

23. Chung HY*, et al.* (1997) Xanthine dehydrogenase/xanthine oxidase and oxidative stress. *Age (Omaha)* 20(3):127-140.
24. Hersh BM*, et al.* (1996) A glutamate-dependent acid resistance gene in Escherichia coli. *J Bacteriol* 178(13):3978-3981.
25. Richard H & Foster JW (2004) Escherichia coli glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential. *J Bacteriol* 186(18):6032-6041.
26. Rapala-Kozik M, Wolak N, Kujda M, & Banas AK (2012) The upregulation of thiamine (vitamin B1) biosynthesis in Arabidopsis thaliana seedlings under salt and osmotic stress conditions is mediated by abscisic acid at the early stages of this stress response. *BMC Plant Biol* 12:2.