ABSTRACT

MAPPING ECOLOGICALLY IMPORTANT VIRUS-HOST INTERACTIONS IN GEOGRAPHICALLY DIVERSE SOLAR SALTERNS WITH METAGENOMICS

by Abraham Ghoreishi Moller

Viruses that infect microbes are critical players in the world's ecosystems. By lysing microbes, viruses turn over nutrients, regulate microbial populations, and maintain global biogeochemical cycles. Despite this ecological importance, determining viral-microbial interactions - especially interactions with lytic viruses, which do not integrate into their host's genome - remains a major challenge. In this work, we determine viral interactions with microbes in salt-collecting ponds (salterns), where viruses are the dominant predator of the microbial community. Low microbial diversity, environmental stability, and high viral density also make solar salterns excellent model ecosystems for studying viral-archaeal interactions. By using a suite of bioinformatics tools to analyze saltern metagenomes, we mapped virus-host interactions across geographically diverse salterns and related them to carbon cycling. Our studies suggest viruses are critical players in saltern carbon cycling, and the loss of CRISPRs in archaea hosts may play an important role in regulating virus-mediated nutrient cycling in these environments.

MAPPING ECOLOGICALLY IMPORTANT VIRUS-HOST INTERACTIONS IN
GEOGRAPHICALLY DIVERSE SOLAR SALTERNS WITH METAGENOMICS

A Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master of Science

by

Abraham Ghoreishi Moller

Miami University

Oxford, Ohio

2016

Advisor: Chun Liang

Reader: Michael Crowder

Reader: Gary Lorigan

This Thesis titled

MAPPING ECOLOGICALLY IMPORTANT VIRUS-HOST INTERACTIONS IN
GEOGRAPHICALLY DIVERSE SOLAR SALTERNS WITH METAGENOMICS

by

Abraham Ghoreishi Moller

has been approved for publication by

The College of Arts and Science

and

Department of Biology

Chun Liang

Michael Crowder

Gary Lorigan

Table of Contents

Summary	1
Introduction	1
Results	3
Discussion	7
Materials and Methods	8
References	11
Figures	16
Supplementary Information	24

List of Tables

Table S1: Additional information about the Santa Pola, Isla Cristina, and Cahuil solar saltern metagenomes provided in the NCBI Sequence Read Archive (SRA) and associated genome announcements (21–23). Site names (as listed in the Figures) are listed in bold. NCBI SRA accession numbers (SRX for experiment and SRR for run) are included in the table.

Table S2: Additional information about the Chula Vista saltern metagenomes obtained from the iMicrobe Collaborative (55) and combined into three metagenomes based on salt concentration. Site names (as listed in the Figures) are in bold. The datasets can be found at this web address: <http://data.imicrobe.us/project/view/58>.

Table S3: The library of haloviral genomes examined included 326 contigs more than 500 bp in length assembled from Chula Vista metaviromes (<http://data.imicrobe.us/project/view/58>) and the following haloviral genomes obtained from NCBI GenBank.

Table S4: Sequence assembly statistics

Table S5: Alignments of combined Santa Pola CRISPR direct repeats identified with Crass against a library of taxonomically annotated direct repeat sequences obtained from CRISPRdb

Table S6: Query sequences used for reference-guided CRISPR search and associated archaeal hosts

Table S7: Summary of CRISPR virus-host pairings in Cahuil (C34) metagenome; CRISPR spacers were detected with the reference-guided method

Table S8: Summary of CRISPR virus-host pairings in SS13 metagenome; CRISPR spacers were detected with the reference-guided method

Table S9: Summary of CRISPR virus-host pairings in SS19 metagenome; CRISPR spacers were detected with the reference-guided method

Table S10: Summary of CRISPR virus-host pairings in IC21 metagenome; CRISPR spacers were detected with the reference-guided method

Table S11: Summary of CRISPR virus-host pairings in SS33 metagenome; CRISPR spacers were detected with the reference-guided method

Table S12: Summary of CRISPR virus-host pairings in SS37 metagenome; CRISPR spacers were detected with the reference-guided method

Table S13: The library of *cas* genes identified in Halobacteriales genomes that were aligned against Newbler contigs assembled from the Cahuil/C34, combined Chula Vista, and combined Santa Pola/Isla Cristina metagenomes

Table S14: The library of *cas* genes identified in Haloferacales genomes that were aligned against Newbler contigs assembled from the Cahuil/C34, combined Chula Vista, and combined Santa Pola/Isla Cristina metagenomes

Table S15: The library of *cas* genes identified in Natrilbales genomes that were aligned against Newbler contigs assembled from the Cahuil/C34, combined Chula Vista, and combined Santa Pola/Isla Cristina metagenomes

List of Figures

Figure 1: Taxonomic profiles (A) and principal component analysis (B) of selected saltern metagenomes. Metagenomes were profiled using MetaPhyler and resulting profiles clustered using MATLAB. Different clusters are marked in different colors and sites corresponding to individual metagenomes are labeled within clusters.

Figure 2: Glycerol metabolism gene family abundances determined using ShotMAP. Abundances are normalized by gene lengths and by sequencing depths through read counts. Abundances are labeled both by Pfam gene family (e.g., Dak1) and by site corresponding to metagenome.

Figure 3: Heat maps of glycerol metabolism gene family abundances across genus. Abundances were clustered by genus and site (A), by genus and family (B), and by genus, site, and family (C). The first two heat maps (A and B) were generated with the heatmap R package, while the last heat map (C) was generated with ggplot2.

Figure 4: Number of CRISPR spacers detected in each metagenome either with the *de novo* detection method (Crass) or the reference-guided method (and corresponding maximum edit distances) described in Materials and Methods that used a query of 29 halobacterial CRISPR direct repeat sequences.

Figure 5: Comparison of CRISPR spacers and direct repeats detected with *de novo* (Crass) and reference-guided methods. All reference-guided spacers compared here with Crass spacers were detected with a maximum allowed edit distance of 3 (insertions, deletions, or mismatches with query direct repeat sequence). CRISPR spacers detected in SS13, SS19, SS33, and SS37 metagenomes with *de novo* (A) and reference-guided (B) methods were compared among sites (not methods) to determine identical spacers among sets. Across IC21 (C) and Cahuil/C34 (D) data sets, spacers detected were compared across the two detection methods. Direct repeats detected with Crass were compared amongst Cahuil (C34), combined Chula Vista, and combined Santa Pola and Isla Cristina metagenomes (E).

Figure 6: Map of virus-host interactions generated by aligning Crass spacers detected in the combined Santa Pola and Isla Cristina metagenome against a library of haloviral genomes (see Supplementary Table S3). Viruses are marked in orange and spacers in blue. Visualization was performed with Cytoscape.

Figure 7: Map of virus-host interactions generated by aligning spacers detected with reference-guided methods in SS13 (A), SS19 (B), SS33 (C), SS37 (D), IC21 (E), and Cahuil/C34 (F) metagenomes against a library of haloviral genomes (see Supplementary Table S3). Viruses are marked in orange and spacers in blue. Visualization was performed with Cytoscape.

Figure 8: Measures of difference in dimer and codon usages between host and virus genomes matched with CRISPR spacers. The differences between each possible dimer and codon usage was subtracted, squared, and summed across all possible dimers or codons to generate a measure of dimer or codon difference. The average dimer usage and codon usage was calculated for all combinations of viruses with a particular host (A and B, respectively) and all virus-host combinations detected for a particular metagenome (C and D, respectively). All averages are reported with a 95% confidence interval (within two standard error) in black.

Figure 9: Detection of *cas* operons in metagenomes and affiliation of detected genes to representative taxa. Three metagenomes (Cahuil/C34, combined Chula Vista, and combined Santa Pola and Isla Cristina) were assembled and searched for contigs matching *cas* genes. Detection results are reported for five halobacterial *cas* operons (A). *cas* genes detected in no metagenomes are shown in red; genes detected in one metagenome are shown in blue; genes detected in two metagenomes are shown in green (A). Detected contigs assembled with Newbler and Velvet were further aligned against a library of *cas* genes from three halobacterial orders (Halobacteriales, Haloferacales, and Natrilbales). The taxonomic affiliations of the best hits for each contig were tabulated into profiles for each order and assembly method (B).

Figure S1: Percentage of reads mapping to reference library in taxonomic profiling steps (MetaPhyler)

Figure S2: Percentage of reads mapping to reference library in functional profiling steps (ShotMAP)

Figure S3: Number of CRISPR spacers detected by *de novo* and reference-guided methods, normalized by millions of reads in each metagenome

Figure S4: Network of virus-host interactions generated from *Haloquadratum walsbyi* strain C23 spacer alignments against the haloviral genome library used in this study. Viruses are marked in blue, while spacers are in orange.

Figure S5: Measures of the number of contig BLAST hits to *cas* genes extracted from three different halobacterial orders (Halobacteriales, Haloferacales, and Natrilbales)

Figure S6: Taxonomic affiliation of all BLAST hits of *cas* genes against contigs (not only best hits)

Dedication

I dedicate my thesis to my parents Akram and James Moller.

Acknowledgements

The project was supported by an Academic Challenge grant from Miami University. Thanks to Chun Liang, Michael Crowder, and Gary Lorigan for feedback on the project and manuscript.

Summary

Solar salterns are excellent model ecosystems for studying virus-microbial interactions because of their low microbial diversity, environmental stability, and high viral density. By using the power of CRISPR spacers to link viruses to their prokaryotic hosts, we explored virus-host interactions in geographically diverse salterns and related them to carbon cycling. Using taxonomic profiling, we identified hosts such as archaeon *Haloquadratum*, *Halorubrum*, and *Haloarcula* and bacteria *Salinibacter*, and we found that community composition related to not only salinity but also local environmental dynamics. Characterizing glycerol metabolism genes in these metagenomes suggested most dihydroxyacetone kinase genes affiliate to *Halorubrum* and *Haloquadratum* while most glycerol-3-phosphate dehydrogenase genes affiliate to *Salinibacter*. We identified CRISPR spacers in the metagenomes with two different methods and found more spacers in the generally Halobacteriaceae-dominated IC21 and C34 salterns compared with the specifically *Haloquadratum*-dominated SS19, SS33, and SS37 salterns, suggesting a low level of CRISPR diversity and possibly a high rate of CRISPR loss in the *Haloquadratum*-dominated salterns. After CRISPR detection, spacers were aligned against haloviral genomes to map virus to host. While most alignments linked viruses to *Haloquadratum walsbyi*, there were clusters of interactions with less abundant *Haloarcula* and *Haloferax*. Further examination of the dimer and codon usage differences between paired viruses and their hosts and detection of *cas* genes in the salterns confirmed both the plausibility of virus-host interactions and the possibility of CRISPR activity. Taken together, our studies suggest CRISPR loss in archaeal hosts controls the level of virus proliferation and the nutrient turnover they induce in these environments.

Introduction

Viruses, the most abundant lifeforms on the planet, are critical for maintaining global biogeochemical cycles (1, 2). Many viruses infect and lyse microbial hosts, releasing stored nutrients back into the ecosystem through a mechanism called the viral shunt (1, 3). This relationship suggests viral-microbial interactions are important for environmental nutrient cycling. However, viruses' host specificities are poorly understood. Clustered regularly interspaced palindromic repeat (CRISPR) arrays, which are found in many archaeal and bacterial genomes (4), can help us better understand the virus-host interactions that mediate nutrient cycling. Acting as an adaptive immune system against viral infection, CRISPRs incorporate short spacers cleaved from viral DNA into host genomes, providing a record of past infections and thus associations between viruses and prokaryotic hosts (5).

Virus-host interactions are particularly important in solar saltern ecosystems. These salt-collecting ponds are hypersaline ecosystems that vary from 10 up to 37% (saturation) salt concentration by weight (6). The salterns at Santa Pola near Alicante, Spain contain a low diversity community of halophile microbial hosts, making it a good model ecosystem (6). The square archaeon *Haloquadratum walsbyi* and bacterioidete *Salinibacter ruber* are the two major components of saltern microbial communities at the highest salinities (7). In addition, hypersaline environments are estimated to have the highest densities of viruses ($10^9/\text{mL}$) among studied aquatic environments (8). Reduced viral decomposition in hypersaline conditions and a lack of protozoan predation may explain this extraordinarily high density (8). Though heterotrophic protozoans have been identified in Korean salterns (9), recent studies suggest these species

primarily graze upon algae, not bacteria (8). Though protozoa probably do not engulf bacteria, viruses may indeed prey upon halophilic bacteria and archaea.

While saltern microbial eukaryotes may not be predators of archaea and bacteria, they do serve as the primary producers in these ecosystems. The halotolerant, unicellular green alga *Dunaliella salina* synthesizes vast quantities of glycerol as an osmoprotectant in response to hyperosmotic shock (11). Glycerol that leaks out of these algae may be the primary carbon source for the bacteria and archaea that inhabit these ecosystems (12–14). The rare saltern archaeon *Haloferax volcanii* converts glycerol to glycerol-3-phosphate and then dihydroxyacetone phosphate (DHAP) using glycerol kinase and glycerol-3-phosphate dehydrogenase, respectively (15, 16). Bacteria *Salinibacter ruber*, on the other hand, oxidizes glycerol to dihydroxyacetone (17, 18). Both *Haloferax* and *Haloquadratum walsbyi* can convert dihydroxyacetone to glycolysis intermediate DHAP using dihydroxyacetone kinase (17, 19, 20). Considering this metabolic context, haloviruses that lyse halophile microbes may promote carbon cycling either by turning over glycerol metabolism intermediates or reducing rates of glycerol uptake within the ecosystem.

In recent years, whole metagenome sequencing has made it possible to study the total microbial DNA from solar salterns across the planet. A subset of the Santa Pola salterns have recently been sampled and sequenced (21), in addition to salterns in Isla Cristina (Spain) (22), Cahuil (Chile) (23), and Chula Vista (California) (24). Metagenomics has also made it possible to study the vast haloviral populations that inhabit these ecosystems. A number of viral genomes from the CR30 (crystallizer, 30% salt) saltern in Santa Pola have recently been isolated and sequenced using fosmid cloning (8). Many of these fosmid clones failed to hybridize with microbial DNA, suggesting most of isolated CR30 viruses are lytic (25). Haloviral genomes have also been assembled from viral metagenomes sequenced from Lake Tyrell in Australia (26) and the Chula Vista salterns in California (24).

While metagenomic sequencing has provided extraordinary glimpses into the viral and microbial diversity of hypersaline ecosystems, it has also enabled detection of microbial CRISPR arrays. Recent studies in a variety of ecosystems have exploited CRISPRs to link environmental viruses to their microbial hosts (26–32). Among these, one study examined virus-archaea interactions in an ecosystem similar to salterns – the hypersaline Lake Tyrrell, Australia (26). Although this study showed that CRISPRs in the lake were remarkably dynamic over time, it also indicated a number of viruses could be mapped to saltern microbes with CRISPR spacers (26). In the saltern context, the abundant community member *Haloquadratum walsbyi* encodes two clusters of *cas* genes (20, 33) and contains CRISPR spacers that map to many Santa Pola haloviral genomes (34). These studies suggest CRISPRs have great potential to indicate virus-archaea interactions within saltern ecosystems.

In this study, we analyze saltern metagenomes to link haloviruses to their hosts and propose a potential ecological role for these interactions. Furthermore, we incorporate knowledge gained from studying not one but many hypersaline ecosystems across the planet. We not only use metagenomic CRISPR sequences to map virus-host interactions across different saltern microbial populations, but also place these interactions in the broader context of carbon cycling within these ecosystems. In this context, we find that CRISPRs and viruses together help regulate glycerol consumption and dynamics in these hypersaline ecosystems.

Results

Taxonomic profiling identifies predominant members of saltern microbial communities

We profiled the taxonomic abundances of metagenomes sequenced from Cahuil (C34), Chula Vista (CV6-8, CV12-14 and CV27-30), Isla Cristina (IC21), and Santa Pola Saltern (SS13, SS19, SS33, SS37) sites (where the numbers indicate salinity percentages) to better understand the population structures of saltern microbial communities and to compare host abundances amongst sites. Prior studies suggested seasonal shifts in ion concentrations control taxonomic abundances in salterns (35). The studies found *Haloquadratum* abundance increases with potassium, magnesium, and sulfate concentrations, while *Halorubrum*, *Haloarcula*, *Halonotius*, and *Halobaculum* abundances decrease with these same factors (35). Research groups have previously taxonomically profiled the Santa Pola and Isla Cristina metagenomes through mapping of reads to 16S rRNAs obtained from the Ribosomal Database Project (6, 21, 22). These 16S rRNA profiling studies revealed that the lowest salinity saltern (SS13) contained a diverse community of bacterial Actinobacteria and Proteobacteria and archaeal Euryarchaeota, while members of the Euryarchaeota formed an increasing proportion of the saltern microbial communities as salinity increased (7, 36, 37).

We used a whole metagenome profiling tool (MetaPhyler) that maps reads to clade-specific markers to determine taxonomic abundances in each of the metagenomes (38). We found the low salinity Chula Vista (CV6-8) and SS13 salterns had the most diverse microbial communities, containing significant proportions of bacterial Actinobacteria, Proteobacteria and Rhodobacteraceae (Figure 1A). In addition, we found *Haloquadratum* dominated the SS19, SS33, and SS37 salterns, while *Halorubrum* and other members of the archaeal Halobacteriaceae dominated the Cahuil and IC21 salterns (Figure 1A). Bacterial *Salinibacter* was an abundant component of the SS19 and more noticeably the SS33 and SS37 salterns. While *Haloquadratum* dominated higher salinity Santa Pola salterns, various Halobacteriaceae dominated medium and high salinity Chula Vista salterns. We realize that the Halobacteriaceae component of the saltern communities may also include species in the new order Haloferacales (39), but this taxon was defined after the MetaPhyler marker library was constructed (38).

We further compared the taxonomic profiles through principal component analysis, which clustered SS19, SS33, and SS37 profiles; low salinity Chula Vista (CV6-8) and SS13 profiles; and medium salinity Chula Vista (CV12-14), high salinity Chula Vista (CV27-30), IC21, and C34 profiles (Figure 1B). We also determined the percentage of reads mapping to phylogenetic markers to compare profiling quality among metagenomes (Supplementary Figure S1). The percentage mapping was consistent among Santa Pola, Isla Cristina, and Cahuil (C34) metagenomes, but much lower among the Chula Vista genomes due to lower coverage and shorter read length. These taxonomic profiles and clustering results further guided our analyses of host CRISPR arrays, as we used taxonomy to guide our searches for CRISPR spacers.

Functional profiling reveals abundances of glycerol metabolism genes and their taxonomic affiliations

Glycerol is the main carbon source for saltern microbes. We next used two functional profiling tools to study the glycerol metabolism genes critical for carbon cycling in these ecosystems. We used the tool ShotMAP to determine abundance of glycerol metabolism gene

families (40) and metAnnotate to study the taxonomic distribution of reads mapping to these gene families (41). As with MetaPhyler, Cahuil, Isla Cristina, and Santa Pola metagenomic reads mapped to the ShotMAP reference library in higher proportions than the Chula Vista reads (Supplementary Figure S2). Using these tools, we examined the abundances and taxonomic affiliations of the glycerol kinase (FGGY_N and FGGY_C), NAD⁺-dependent glycerol-3-phosphate dehydrogenase (NAD_Gly3P_dh_N and NAD_Gly3P_dh_C), glycerol dehydrogenase (iron-containing alcohol dehydrogenase – Fe-ADH and Fe-ADH_2), and dihydroxyacetone kinase (Dak1, Dak1_2, and Dak2) Pfam gene families. Prior studies of the Santa Pola and Isla Cristina salterns found that glycerol metabolism genes, such as dihydroxyacetone kinase, glycerol kinase, and NAD⁺-dependent glycerol-3-phosphate dehydrogenase, increased in abundance with salt concentrations (36, 37).

We first determined gene family abundances with ShotMAP. In the Santa Pola (13-37% salinity: SS13, SS19, SS33 and SS37), Isla Cristina (21%: IC21), and Cahuil (34%: C34) metagenomes, we found the abundance of the dihydroxyacetone kinase (Dak1 and Dak2) and alcohol dehydrogenase (Fe-ADH_2) gene families increased from 13% to 19% salinity, but remained consistent from 19% to 37% (Figure 2). The alcohol dehydrogenase (Fe-ADH), glycerol kinase, and NAD⁺-dependent glycerol-3-phosphate dehydrogenase families had consistent levels among 19-37% salinity salterns, but higher levels in the SS13 saltern, and especially so for the NAD⁺-dependent glycerol-3-phosphate dehydrogenase families. Among the Chula Vista metagenomes, gene family abundance generally increased with salinity, with the notable exceptions of the alcohol dehydrogenase and NAD⁺-dependent glycerol-3-phosphate dehydrogenase families. We also noted unusually high family abundances among the Chula Vista metagenomes. A combination of smaller total numbers of reads and more short reads mapping to these gene families may account for these relatively high values compared with the Santa Pola, Isla Cristina, and Cahuil metagenomes.

We next determined the taxonomic affiliations of reads mapping to these gene families using metAnnotate. We generated heat maps of these affiliations clustered both with regard to metagenome (Figure 3A) and gene family (Figure 3B). As with taxonomic profiles, the taxonomic affiliations of glycerol metabolism genes in the *Haloquadratum*-rich SS19, SS33, and SS37 metagenomes cluster together, as do those of the more diverse Chula Vista and the *Halorubrum*-rich IC21 and Cahuil (C34) metagenomes (Figure 3A). Across gene family, taxonomic affiliations of the dihydroxyacetone kinase families cluster together, as well as those of the alcohol dehydrogenase (Fe-ADH) and glycerol kinase gene families (Figure 3B). The two NAD⁺-dependent glycerol-3-phosphate dehydrogenase families clustered together, but had a large proportion of unclassified sequences. We also analyzed these taxonomic affiliations for each gene family across each individual metagenome (Figure 3C). The dihydroxyacetone kinase families most strongly affiliated to *Haloquadratum* in the SS19, SS33, and SS37 metagenomes and to *Halorubrum* in the IC21 and Cahuil metagenomes (Figure 3C). The two NAD⁺-dependent glycerol-3-phosphate dehydrogenase families, on the other hand, affiliated mainly to *Salinibacter* in the Cahuil (C34), high salinity Chula Vista, SS19, SS33, and SS37 metagenomes and to *Spiribacter* in the SS13 and IC21 metagenomes. Unlike the previous families, alcohol dehydrogenase and glycerol kinase had far less clear affiliations across the metagenomes, with significant affiliation to *Haloquadratum* and *Halorubrum* but also many unclassified reads. In addition, glycerol kinase families affiliated appreciably to *Spiribacter* and the alcohol dehydrogenase family affiliated to *Halobellus* across several metagenomes.

Metagenomic CRISPRs detected with two independent methods map haloviruses to saltern hosts

We detected CRISPR arrays in metagenomes using both *de novo* and reference-guided methods. We used Crass for *de novo* detection (42) and a custom pipeline we developed for reference-guided detection (see Materials and Methods). We developed this custom pipeline both to constrain CRISPR searches based on the taxa we found to be present in each metagenome and to improve search sensitivity. Crass was not sensitive enough to identify spacers that aligned to viruses in several individual metagenomes (e.g., SS13). We used a combination of Halobacteriales and Haloferacales CRISPR direct repeats from CRISPRdb as our query for reference-guided detection analyses across all metagenomes. Overall, many more spacers were detected in IC21 and Cahuil (C34) metagenomes compared with the high salinity Chula Vista, SS13, SS19, SS33, and SS37 metagenomes (Figure 4). Significantly more spacers were detected with Crass compared to reference-guided methods in the combined Santa Pola, SS13, and SS33 metagenomes. For the reference-guided detection analyses, the number of spacers identified steadily increased as the allowable error (insertions, deletions, or substitutions different from the query DR sequence) in alignment was increased from 0 to 3 (Figure 4). No spacers were detected with either method in the low or medium salinity Chula Vista metagenomes. When we normalized by total number of metagenomic reads, we found that the Cahuil (C34) metagenome by far had the largest number of spacers detected per million reads (Supplementary Figure S3).

We then compared the detected direct repeats and spacers among metagenomes to discover common CRISPR array types or spacers (Figure 5). Remarkably, a number of common spacers were identified amongst the SS19, SS33, and SS37 metagenomes using both detection methods (Figures 5A and 5B). Amongst the metagenomes with a higher number of detected spacers (IC21 and Cahuil/C34), some common spacers were detected with both methods (Figures 5C and 5D). Amongst the direct repeats detected with Crass in the combined Chula Vista, combined Santa Pola, and Cahuil metagenomes, there was one direct repeat (DR) sequence common to both Cahuil and combined Santa Pola (Figure 5E). We also searched the combined Santa Pola direct repeats against a full database of direct repeats from CRISPRdb to determine possible taxonomic affiliations. We found that 14 of the 67 detected combined Santa Pola DRs matched DRs in the CRISPRdb database (Table S5). Of these 14 matching DRs, 10 matched haloarchaeal DRs, while 4 matched bacterial DRs.

We then aligned the spacers detected with both methods against a library of haloviral genomes obtained from GenBank to determine virus-host interactions in these salterns. Unlike CRISPR spacers detected with Crass in several Santa Pola metagenomes, spacers detected in a combination of all Santa Pola and Isla Cristina metagenomes did align to many haloviral genomes (Figure 6). In this Crass analysis of the combined Santa Pola dataset, we detected clusters of virus interactions most likely with *Haloquadratum* (G467), *Haloarcula* (G42 and G82), and Pseudonocardiales (G57) hosts, based on our previous comparison of Crass DRs (listed by group number, such as G467) against CRISPRdb DRs (Figure 6). For many of the interactions (e.g., G467, G42), we could determine a possible host from by aligning the detected Crass DRs against CRISPRdb DRs from assembled genomes. For several notable clusters, however, we could not determine possible hosts from such alignments (e.g., G85, G1290).

While Crass only provided a picture of virus-host interactions in the combined Santa Pola metagenome, our reference-guided pipeline detected spacers that mapped to haloviral genomes in

each of the Santa Pola, Isla Cristina, and Cahuil metagenomes (Figure 7). Only one interaction was detected in the SS13 dataset (Figure 7A), but there were strikingly similar clusters of virus-host interactions detected in the SS19, SS33, and SS37 metagenomes (Figures 7B, 7C, and 7D). While many clusters were similar to those expected from the *Haloquadratum walsbyi* strain C23 CRISPR array (Supplementary Figure S4), no C23 spacers matched any detected metagenomic spacers, while at least three *Haloquadratum walsbyi* strain DSM 16790 spacers did match spacers detected in SS19, IC21, SS33, and SS37 metagenomes. Aside from *Haloquadratum*, in the SS33 metagenome, we also detected a cluster of *Haloarcula*-virus interactions with eHP-4, eHP-8, and eHP-13 (Figure 7C). Quite surprisingly, we noticed remarkably different virus-Halobacteriaceae interactions in the IC21 and Cahuil (C34) metagenomes (Figures 7E and 7F), perhaps due to differences in composition at the genus level.

Nucleotide usage comparison suggests biological plausibility of virus-host associations

To determine whether our virus-host mappings are plausible, we compared the dimer and codon frequencies of CRISPR-determined virus-host pairings. We expected these frequencies to be similar between virus and host genomes if viruses were to replicate and express their genes within these hosts. We generated non-redundant lists of virus-host associations for each possible host and each metagenome for this analysis. To determine the overall differences in nucleotide usage, we calculated the total of the squared differences between each possible virus and host dimer (or codon) frequency for each pair in these lists. We compared these frequencies both among the predicted hosts and analyzed metagenomes (Figure 8). Among predicted hosts, *Haloquadratum walsbyi*, *Haloferax volcanii*, and *Haloarcula marismortui* showed the smallest overall differences in dimer and codon usage, while *Haloarcula hispanica* and *Halorhabdus utahensis* showed the highest overall differences (Figures 8A and 8B). Only one virus associated to *Halogeometricum borinquense*, and it had a greater difference in codon frequency compared with dimer frequency relative to the other predicted hosts. Among the metagenomes, the lowest differences were found in the SS33 metagenome; the next lowest differences were found in Cahuil (C34), SS19, and SS37 metagenomes, while the highest difference was found in the single virus-host pair detected in the SS13 metagenome (Figures 8C and 8D).

Assembled metagenomic contigs contain halobacterial *cas* genes

To further confirm our CRISPR detection and mapping results, we searched through metagenomic contigs for the *cas* genes required for CRISPR activity. Sequencing the genome of one strain of *Haloquadratum walsbyi* DSM 16790 showed complete loss of *cas* genes and near total loss of CRISPR arrays (33). We both searched for contigs that matched several different haloarchaeal *cas* operons (Figure 9A) and profiled the taxonomic affiliations of haloarchaeal *cas* genes that best matched contigs (Figure 9B). We were able to assemble almost all *cas* genes in three different metagenomes (Cahuil, combined Chula Vista, combined Santa Pola) from several different Type I-B and Type I-D *cas* operons (Figure 9A). Overall, more genes from Type I-B operons were assembled among multiple metagenomes.

We also searched a library of *cas* genes from three different haloarchaeal orders (Halobacteriales, Haloferacales, and Natrilbales) and generated profiles of taxonomic affiliations for the genes that best matched contigs (Figure 9B). *Haloarcula* composed a substantial proportion of Halobacteriales affiliations, especially amongst the Cahuil (C34) contigs. On the other hand,

Haloquadratum and *Halorubrum* composed major proportions of the Haloferacales annotations, especially for the combined Santa Pola dataset. The Chula Vista dataset had far fewer matching contigs and thus a radically different profile. We also calculated the total number of hits in the search against the *cas* gene libraries (Supplementary Figure S5) and profiled the taxonomic affiliations of all *cas* gene alignments to the contigs (Supplementary Figure S6). There were far lower *Haloarcula*, *Haloferax*, *Halorubrum*, and *Haloquadratum* proportions in the total hit profiles (Supplementary Figure S6) compared with the best hit profiles (Figure 9B).

Discussion

Our studies have not only confirmed prior understandings of saltern taxonomic and functional diversity but also provided new insights into virus-host interactions in many salterns worldwide. We have built upon prior metagenomics studies to show that *Haloquadratum* dominates the archaeal communities of many salterns (43), particularly those that are environmentally stable with high salinity. Compared with the Santa Pola salterns, Cahuil and Isla Cristina salterns may be more dynamic, with more rapidly changing daily and yearly temperatures and higher rainfall (37). This may lead to lower concentrations of potassium, magnesium, and sulfate in those salterns, thus leading to much lower abundances of *Haloquadratum*, and higher abundances of other haloarchaea (35, 37).

Our functional studies found rather similar levels of glycerol metabolism genes across high salinity salterns. Taxonomic affiliations, however, differed significantly amongst salterns and different gene families. Dihydroxyacetone kinase most strongly affiliated to *Haloquadratum*, except in the IC21 and Cahuil (C34) salterns, where it affiliated strongly to *Halorubrum* instead. NAD⁺-dependent glycerol-3-phosphate dehydrogenase, on the other hand, very strongly affiliated to *Salinibacter*. Based on prior understanding of saltern glycerol cycling (13, 18, 44), these results together suggest archaea such as *Haloquadratum* and *Halorubrum* consume the glycerol metabolism intermediate dihydroxyacetone in these ecosystems, while bacteria such as *Salinibacter* may first phosphorylate and then oxidize glycerol to metabolize it (18).

Our CRISPR detection results showed a remarkably small number of CRISPR spacers in *Haloquadratum*-dominated compared with more diverse Halobacteriaceae-dominated salterns. Furthermore, comparison of detected metagenomic spacers to those found in the genomes of two different *H. walsbyi* strains (C23 and DSM 16790) strongly suggests these archaea may be undergoing spacer loss in the salterns. Strain C23 contains a full set of *cas* genes and several CRISPR arrays, while strain DSM 16790 lost all its *cas* genes and all but five spacers (33). While we could identify no C23 spacers amongst our metagenomes, we could find at least three DSM 16790 spacers in several metagenomes. A successful mobile element integration event in spite of CRISPR defense is thought to have promoted loss of spacers in *Haloquadratum* genomes (33). Taken together, this evidence strongly suggests that the *Haloquadratum* species in these salterns are undergoing net spacer loss.

Our CRISPR mapping results not only showed consistent viral interactions with *Haloquadratum* across salterns but also numerous interactions with less abundant taxa such as *Haloferax* and *Haloarcula*. As expected from our comparison of taxonomic and functional profiles, there were consistent sets of interactions amongst *Haloquadratum*-dominated SS19, SS33, and SS37 salterns. These interactions suggest similar viral infection patterns across these environments, consistent with “global pond” theories of microbial and viral exchange (33, 45). Unclear biological requirements for CRISPR spacer incorporation make it difficult to make any

conclusion about the abundances of the viruses. Though some models suggest viruses must have reached a critical number within their hosts to induce CRISPR spacer incorporation into host genomes (46), characterization of the *Haloferax volcanii* CRISPR systems suggests plasmid copy number has little effect on incorporation (47).

There were also significant numbers of interactions detectable with Crass that could not map to any known hosts. Some of these hosts may be poorly understood members of the Nanohaloarchaea, as recent studies have suggested for eHP-12 and eHP-16 (48). Surprisingly, we did not detect any spacers or virus-host interactions with *Halorubrum* with either CRISPR detection method. Either limitations of our detection tools or a lack of *Halorubrum* CRISPRs could account for this observation.

Taken together, our results suggest that both viruses and CRISPR systems may regulate saltern ecology. While common sets of viruses infect haloarchaea across salterns, these hosts may have strikingly different levels of CRISPR defenses. *Haloquadratum*, it seems, may exist in either virus-resistant and virus-sensitive strains depending on the presence or absence of its CRISPR/Cas system. CRISPR/Cas systems may be part of the previously examined *Haloquadratum* accessory gene pool (49). While it is surprising that virus-sensitive strains would be present in such a virus-rich environment, even these forms may have some innate halovirus resistance through extensive genomic variation of phage recognition proteins (50). These patterns may not apply to other haloarchaea and saltern inhabitants, however. While *Haloquadratum* encodes CRISPRs chromosomally, both *Haloarcula* and *Haloferax* encode CRISPR elements on both their chromosomes and plasmids, suggesting the possibility of horizontally transferred CRISPR systems (4, 51–54). Viral predation of *Salinibacter* remains far more elusive, as this bacterium lacks a CRISPR system (4).

Nonetheless, the loss of CRISPR systems and resulting loss of phage resistance in *Haloquadratum* may carefully control its predation. Together both haloviral predation and CRISPR defense could tightly regulate the population of *Haloquadratum*, which is both the dominant component of many salterns and consumer of the glycerol metabolism intermediate dihydroxyacetone. Levels of CRISPR defense may thus account for the turnover of this organism. Future studies should further analyze the number and genomic arrangement of spacers in different *Haloquadratum* strains and examine how abundances of these populations change over time. Such work would build upon recent studies of the temporal dynamics of virus-host interactions in Lake Tyrrell, an Australian hypersaline ecosystem (26). Taken together, our studies and such future projects will help us better understand how viruses regulate halophile populations across the planet.

Materials and Methods

Obtaining metagenomic data sets and haloviral genomes

The Santa Pola, Isla Cristina, and Cahuil metagenomes were downloaded from the NCBI Sequence Read Archive (SRA). Santa Pola and Isla Cristina metagenomes (SS13, SS19, IC21, SS33, and SS37) were analyzed individually except for assembly and *de novo* CRISPR detection steps. The Cahuil (C34) metagenome was analyzed individually throughout. Chula Vista metagenomes were obtained from the iMicrobe Collaborative (55). Chula Vista metagenomes were grouped by corresponding salt concentration (low, medium, or high) and concatenated to form single low (CV6-8), medium (CV12-14), and high (CV27-30) salt concentration

metagenomes. Haloviral genomes as well the Lake Tyrell viral contigs were downloaded from NCBI GenBank (Supplementary Table S3). Additionally, Chula Vista solar saltern viral metagenomes were obtained through iMicrobe (55) and then combined into one metagenome. This combined viral metagenome was then assembled into Chula Vista viral contigs that were also included in the haloviral library (Supplementary Table S3).

Profiling taxonomic and functional compositions of selected metagenomes

The selected metagenomes were profiled for taxon abundances with MetaPhyler (38). MetaPhyler aligns reads against a library of clade-specific marker genes. Taxonomic profiles were clustered and compared in MATLAB using principal component analysis (PCA). The percent of reads mapping to markers for each metagenome were calculated and compared to detect differences in profile quality. The metagenomes were further profiled for functional abundances with ShotMAP (40) and metAnnotate (41). ShotMAP determination of gene family abundance was conducted with a Hidden Markov Model (HMM)-based search algorithm against an indexed library of Pfam family HMMs. Abundance was calculated from the numbers of metagenomic reads mapping to particular Pfam families. metAnnotate was used to determine the abundance of particular Pfam gene families for particular taxa. We used metAnnotate to taxonomically profile seven gene families known to be related to glycerol metabolism – dihydroxyacetone kinase (Dak1 – PF02733; Dak2 – PF02734), NAD⁺-dependent glycerol dehydrogenase (Fe-ADH – PF00465), NAD⁺-dependent glycerol-3-phosphate dehydrogenase (both N- and C-terminal domains: NAD_Gly3p_dh_N – PF01210; NAD_Gly3p_dh_C – PF07479), and glycerol kinase (FGGY_N – PF00370; FGGY_C – PF02782). We then calculated the proportion of reads mapping to a taxon among all those mapping to a gene family to determine the taxonomic profiles of these gene families.

***De novo* and reference-guided CRISPR detection**

CRISPR arrays were detected using two different methods (reference-guided and *de novo*). CRISPR arrays are highly difficult to assemble due to their highly repetitive nature. CRISPR arrays were thus identified purely from unassembled metagenomic reads. *De novo* detection was performed with Crass (42) with default parameters, which identifies repeat-spacer-repeat patterns in individual reads, uses these as “seeds” to recruit other CRISPR-containing reads, and thus detects new CRISPR arrays. We further clustered Crass spacers using CD-HIT (56) with a similarity threshold of 0.9 and aligned them against known direct repeats from CRISPRdb (4) to determine the associated host organism.

Because Crass does not directly indicate the putative host encoding the detected CRISPRs, we developed a pipeline to constrain spacer detection by expected host species. This pipeline searches each read for direct repeat (DR) sequences matching query DRs known to be present in genomic CRISPR arrays. Direct repeats within a certain Levenshtein edit distance (i.e., number of insertions, deletions, or substitutions necessary to convert one sequence to another) are identified in reads using the Wu-Manber multi-pattern search algorithm (57). In the first step of the pipeline, individual reads are searched for two or more copies of the query direct repeat sequence. In the second step, the sequence fragments between the direct repeats detected in these sequence reads are extracted as the spacers. These spacers are then collected as a non-redundant set in a hash, and then further clustered using CD-HIT (56). We performed reference-guided CRISPR detection

using a group of 29 query direct repeat sequences previously detected in Haloferacales and Halobacteriales genomes (Supplementary Table S6). Acceptable errors - the allowed differences between query and detected DR sequences - were set from 0 through 3 (insertions, deletions, or substitutions). Detected spacers were then clustered using CD-HIT with a clustering similarity threshold of 0.9.

Mapping virus-host interactions with CRISPR spacers

CRISPR spacers extracted from the metagenomes with Crass and reference-guided methods and clustered with CD-HIT were then aligned against the library of haloviral genomes using NCBI BLAST (58) with an e-value cutoff of 1e-03. All spacers detected with Crass for the combined Santa Pola dataset (SS13, SS19, IC21, SS33, and SS37) were aligned against this library to create its corresponding interaction network. For the individual metagenomes (e.g., SS13), no Crass spacers mapped to the haloviral genomes, necessitating the more sensitive reference-guided search. Thus, for individual metagenomes, spacers detected with the reference-guided method using a maximum Levenshtein edit distance of 3 were aligned for network creation. BLAST results from each method were then processed to form lists of virus-host pairs. Networks of virus-host interactions were drawn with Cytoscape (59).

Comparative analysis of virus and host sequence characteristics

Viruses and hosts matched through CRISPR mapping were further compared in nucleotide k-mer frequencies to determine whether the mappings were plausible. Dinucleotide and trinucleotide (codon) usages were determined using the function *nmercount* from the MATLAB Bioinformatics Toolbox. Dinucleotide and trinucleotide frequencies were compared for a nonredundant list of viruses associated to each of the six detected hosts listed below. All viruses were associated to possible hosts using our in-house reference-based CRISPR detection methods. *Haloquadratum walsbyi* C23, *Haloferax volcanii* DS2, *Halogeometricum borinquense* DSM 11551, *Halorhabdus utahensis* DSM 12940, *Haloarcula marismortui* ATCC 43049 (chromosome 1), and *Haloarcula hispanica* N601 (chromosome 1) host genomes were analyzed along with the matched haloviral genomes. To measure the sequence characteristic differences among virus-host pairs, each virus dinucleotide and trinucleotide frequency was subtracted from each corresponding host genome frequency and squared. All these squared distances were summed for possible dinucleotides and for possible trinucleotides to generate an overall measure of distance between virus and host.

Assembling metagenomes and detecting halobacterial *cas* genes in metagenomic contigs

Santa Pola, Chula Vista, and Cahuil metagenomes were assembled into contigs to detect *cas* operons. All Santa Pola (SS13, SS19, SS33, and SS37) and Isla Cristina (IC21) metagenomes were combined into a combined Santa Pola dataset for assembly, while all Chula Vista metagenomes were similarly combined to a combined Chula Vista dataset. Combined Santa Pola, combined Chula Vista, and Cahuil metagenomes were assembled both with Newbler and Velvet, creating a total of six sets of contigs (60). Both Newbler and Velvet were used to compare the effects of Overlap-Layout-Consensus (Newbler) and de Bruijn graph (Velvet) assembly methods. Because all microbial metagenomes studied were sequenced with 454 technology, and Newbler is

designed to assemble 454 reads, we expected longer contigs to be assembled with Newbler and found this in our assembly results (Supplementary Table S4). Newbler assembly was performed with default parameters, while Velvet assembly was performed with a k-mer size of 31. Newbler and Velvet contigs assembled from each combined metagenome were first aligned against *cas* genes annotated in *Haloferax volcanii* DS2, *Halorubrum lacusprofundi* ATCC 49239, and *Haloquadratum walsbyi* C23 genomes using NCBI blastn (58) with an e-value cutoff of 1e-06 and NCBI tblastx with an e-value cutoff of 1e-30. Color-coded *cas* detection results for genes in the *Haloquadratum*, *Halorubrum*, and *Haloferax* *cas* operons were drawn with the synthetic biology tool Pigeon (61). A library of *cas* genes from the three major halobacterial orders (the Haloferacales, Halobacteriales, and Natrilbales) were downloaded from NCBI Genbank (see Supplementary Tables S13-S15) and aligned against each set of assembled contigs (assembled with each method and from each combined metagenome) to determine the taxonomic affiliation of matching *cas* genes. BLAST alignments were performed with the same parameters described above. Taxonomy information was extracted from matching *cas* gene Genbank records with the NCBI Efetch utility to determine the taxonomic affiliations of the assembled *cas* contigs.

References

1. Suttle CA (2007) Marine viruses — major players in the global ecosystem. *Nat Rev Microbiol* 5(10):801–812.
2. Rohwer F, Prangishvili D, Lindell D (2009) Roles of viruses in the environment. *Environ Microbiol* 11(11):2771–2774.
3. Weitz JS, Wilhelm SW (2012) Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep* 4. doi:10.3410/B4-17.
4. Grissa I, Vergnaud G, Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8(1):172.
5. Sorek R, Kunin V, Hugenholtz P (2008) CRISPR — a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6(3):181–186.
6. Ventosa A, Fernández AB, León MJ, Sánchez-Porro C, Rodríguez-Valera F (2014) The Santa Pola saltern as a model for studying the microbiota of hypersaline environments. *Extremophiles* 18(5):811–824.
7. Ventosa A, de la Haba RR, Sánchez-Porro C, Papke RT (2015) Microbial diversity of hypersaline environments: a metagenomic approach. *Curr Opin Microbiol* 25:80–87.
8. Santos F, et al. (2012) Culture-Independent Approaches for Studying Viruses from Hypersaline Environments. *Appl Environ Microbiol* 78(6):1635–1643.
9. Park JS, Simpson AGB (2015) Diversity of Heterotrophic Protists from Extremely Hypersaline Habitats. *Protist* 166(4):422–437.

10. Heidelberg KB, et al. (2013) Characterization of eukaryotic microbial diversity in hypersaline Lake Tyrrell, Australia. *Front Microbiol* 4. doi:10.3389/fmicb.2013.00115.
11. Chitlaru E, Pick U (1991) Regulation of Glycerol Synthesis in Response to Osmotic Changes in *Dunaliella*. *Plant Physiol* 96(1):50–60.
12. Bardavid RE, Khristo P, Oren A (2006) Interrelationships between *Dunaliella* and halophilic prokaryotes in saltern crystallizer ponds. *Extremophiles* 12(1):5–14.
13. Oren A (1993) Availability, uptake and turnover of glycerol in hypersaline environments. *FEMS Microbiol Ecol* 12(1):15–23.
14. Wegmann K, Ben-Amotz A, Avron M (1980) Effect of Temperature on Glycerol Retention in the Halotolerant Algae *Dunaliella* and *Asteromonas*. *Plant Physiol* 66(6):1196–1197.
15. Rawls KS, Martin JH, Maupin-Furlow JA (2011) Activity and Transcriptional Regulation of Bacterial Protein-Like Glycerol-3-Phosphate Dehydrogenase of the Haloarchaea in *Haloferax volcanii*. *J Bacteriol* 193(17):4469–4476.
16. Sherwood KE, Cano DJ, Maupin-Furlow JA (2009) Glycerol-Mediated Repression of Glucose Metabolism and Glycerol Kinase as the Sole Route of Glycerol Catabolism in the Haloarchaeon *Haloferax volcanii*. *J Bacteriol* 191(13):4307–4315.
17. Bardavid RE, Oren A (2007) Dihydroxyacetone metabolism in *Salinibacter ruber* and in *Haloquadratum walsbyi*. *Extremophiles* 12(1):125–131.
18. Sher J, Elevi R, Mana L, Oren A (2004) Glycerol metabolism in the extremely halophilic bacterium *Salinibacter ruber*. *FEMS Microbiol Lett* 232(2):211–215.
19. Ouellette M, Makkay AM, Papke RT (2013) Dihydroxyacetone metabolism in *Haloferax volcanii*. *Front Microbiol* 4. doi:10.3389/fmicb.2013.00376.
20. Bolhuis H, et al. (2006) The genome of the square archaeon *Haloquadratum walsbyi* : life at the limits of water activity. *BMC Genomics* 7(1):169.
21. Fernandez AB, et al. (2013) Metagenome Sequencing of Prokaryotic Microbiota from Two Hypersaline Ponds of a Marine Saltern in Santa Pola, Spain. *Genome Announc* 1(6):e00933–13.
22. Fernández AB, León MJ, Vera B, Sánchez-Porro C, Ventosa A (2014) Metagenomic Sequence of Prokaryotic Microbiota from an Intermediate-Salinity Pond of a Saltern in Isla Cristina, Spain. *Genome Announc* 2(1):e00045–14.
23. Plominsky AM, et al. (2014) Metagenome Sequencing of the Microbial Community of a Solar Saltern Crystallizer Pond at Cahuil Lagoon, Chile. *Genome Announc* 2(6):e01172–14–e01172–14.

24. Rodriguez-Brito B, et al. (2010) Viral and microbial community dynamics in four aquatic environments. *ISME J* 4(6):739–751.
25. Santos F, Yarza P, Parro V, Briones C, Antón J (2010) The metavirome of a hypersaline environment: The metavirome of a hypersaline environment. *Environ Microbiol* 12(11):2965–2976.
26. Emerson JB, et al. (2013) Virus-Host and CRISPR Dynamics in Archaea-Dominated Hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* 2013:1–12.
27. Berg Miller ME, et al. (2012) Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome: Rumen viral metagenome. *Environ Microbiol* 14(1):207–227.
28. Anderson RE, Brazelton WJ, Baross JA (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage: CRISPR spacers reveal hosts of marine vent viral assemblage. *FEMS Microbiol Ecol* 77(1):120–133.
29. Sanguino L, Franqueville L, Vogel TM, Larose C (2015) Linking environmental prokaryotic viruses and their host through CRISPRs. *FEMS Microbiol Ecol* 91(5):fiv046.
30. Tschitschko B, et al. (2015) Antarctic archaea–virus interactions: metaproteome-led analysis of invasion, evasion and adaptation. *ISME J* 9(9):2094–2107.
31. Zhang Q, Rho M, Tang H, Doak TG, Ye Y (2013) CRISPR-Cas systems target a diverse collection of invasive mobile genetic elements in human microbiomes. *Genome Biol* 14(4):R40.
32. Tyson GW, Banfield JF (2008) Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* 10(1):200–207.
33. Dyall-Smith ML, et al. (2011) Haloquadratum walsbyi : Limited Diversity in a Global Pond. *PLoS ONE* 6(6):e20968.
34. Garcia-Heredia I, et al. (2012) Reconstructing Viral Genomes from the Environment Using Fosmid Clones: The Case of Haloviruses. *PLoS ONE* 7(3):e33802.
35. Podell S, et al. (2014) Seasonal fluctuations in ionic concentrations drive microbial succession in a hypersaline lake community. *ISME J* 8(5):979–990.
36. Fernández AB, et al. (2014) Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol Ecol* 88(3):623–635.
37. Fernández AB, et al. (2014) Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front Microbiol* 5. doi:10.3389/fmicb.2014.00196.

38. Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M (2011) Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* 12(Suppl 2):S4.
39. Gupta RS, Naushad S, Baker S (2015) Phylogenomic analyses and molecular signatures for the class Halobacteria and its two major clades: a proposal for division of the class Halobacteria into an emended order Halobacteriales and two new orders, Haloferacales ord. nov. and Natribacteriales ord. nov., containing the novel families Haloferacaceae fam. nov. and Natribacteraceae fam. nov. *Int J Syst Evol Microbiol* 65(Pt 3):1050–1069.
40. Nayfach S, et al. (2015) Automated and Accurate Estimation of Gene Family Abundance from Shotgun Metagenomes. *PLoS Comput Biol* 11(11):e1004573.
41. Petrenko P, Lobb B, Kurtz DA, Neufeld JD, Doxey AC (2015) MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. *BMC Biol* 13(1):92.
42. Oh D, Porter K, Russ B, Burns D, Dylla-Smith M (2009) Diversity of Haloquadratum and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* 14(2):161–169.
43. Oren A (1994) The ecology of the extremely halophilic archaea. *FEMS Microbiol Rev* 13(4):415–439.
44. Atanasova NS, Roine E, Oren A, Bamford DH, Oksanen HM (2012) Global network of specific virus-host interactions in hypersaline environments: Haloviruses are promiscuous. *Environ Microbiol* 14(2):426–440.
45. Garrett RA, Vestergaard G, Shah SA (2011) Archaeal CRISPR-based immune systems: exchangeable functional modules. *Trends Microbiol* 19(11):549–556.
46. Maier L-K, et al. (2013) Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system I-B. *RNA Biol* 10(5):865–874.
47. Martínez-García M, Santos F, Moreno-Paz M, Parro V, Antón J (2014) Unveiling viral–host interactions within the “microbial dark matter.” *Nat Commun* 5:4542.
48. Legault BA, et al. (2006) Environmental genomics of “Haloquadratum walsbyi” in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species. *BMC Genomics* 7(1):1–13.
49. Cuadros-Orellana S, et al. (2007) Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J* 1(3):235–245.
50. Han J, et al. (2012) Complete Genome Sequence of the Metabolically Versatile Halophilic Archaeon Haloferax mediterranei, a Poly(3-Hydroxybutyrate-co-3-Hydroxyvalerate) Producer. *J Bacteriol* 194(16):4463–4464.
51. Hartman AL, et al. (2010) The Complete Genome Sequence of Haloferax volcanii DS2, a Model Archaeon. *PLoS ONE* 5(3):e9605.

52. Liu H, et al. (2011) Complete Genome Sequence of *Haloarcula hispanica*, a Model Haloarchaeon for Studying Genetics, Metabolism, and Virus-Host Interaction. *J Bacteriol* 193(21):6086–6087.
53. Baliga NS, et al. (2004) Genome sequence of *Haloarcula marismortui*: A halophilic archaeon from the Dead Sea. *Genome Res* 14(11):2221–2234.
54. Hurwitz B (2014) iMicrobe: Advancing Clinical and Environmental Microbial Research using the iPlant Cyberinfrastructure (Plant and Animal Genome). Available at: <https://pag.confex.com/pag/xxii/webprogram/Paper10931.html> [Accessed December 14, 2015].
55. Skennerton CT, Imelfort M, Tyson GW (2013) Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res* 41(10):e105–e105.
56. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658–1659.
57. Wu S, Manber U, Myers E (1995) A Subquadratic Algorithm for Approximate Regular Expression Matching. *J Algorithms* 19(3):346–360.
58. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
59. Shannon P, et al. (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res* 13(11):2498–2504.
60. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18(5):821–829.
61. Bhatia S, Densmore D (2013) Pigeon: A Design Visualizer for Synthetic Biology. *ACS Synth Biol* 2(6):348–350.

Figures

A

B

Figure 1: Taxonomic profiles (A) and principal component analysis (B) of selected saltern metagenomes. Metagenomes were profiled using MetaPhyler and resulting profiles clustered using MATLAB. Different clusters are marked in different colors and sites corresponding to individual metagenomes are labeled within clusters.

Figure 2: Glycerol metabolism gene family abundances determined using ShotMAP. Abundances are normalized by gene lengths and by sequencing depths through read counts. Abundances are labeled both by Pfam gene family (e.g., Dak1) and by site corresponding to metagenome.

Figure 3: Heat maps of glycerol metabolism gene family abundances across genus. Abundances were clustered by genus and site (A), by genus and family (B), and by genus, site, and family (C). The first two heat maps (A and B) were generated with the heatmap R package, while the last heat map (C) was generated with ggplot2.

Figure 4: Number of CRISPR spacers detected in each metagenome either with the *de novo* detection method (Crass) or the reference-guided method (and corresponding maximum edit distances) described in Materials and Methods that used a query of 29 halobacterial CRISPR direct repeat sequences.

Figure 5: Comparison of CRISPR spacers and direct repeats detected with *de novo* (Crass) and reference-guided methods. All reference-guided spacers compared here with Crass spacers were detected with a maximum allowed edit distance of 3 (insertions, deletions, or mismatches with query direct repeat sequence). CRISPR spacers detected in SS13, SS19, SS33, and SS37

metagenomes with *de novo* (A) and reference-guided (B) methods were compared among sites (not methods) to determine identical spacers among sets. Across IC21 (C) and Cahuil/C34 (D) data sets, spacers detected were compared across the two detection methods. Direct repeats detected with Crass were compared amongst Cahuil (C34), combined Chula Vista, and combined Santa Pola and Isla Cristina metagenomes (E).

Figure 6: Map of virus-host interactions generated by aligning Crass spacers detected in the combined Santa Pola and Isla Cristina metagenome against a library of haloviral genomes (see Supplementary Table S3). Viruses are marked in orange and spacers in blue. Visualization was performed with Cytoscape.

A **B** **C** **D** **E** **F**

Figure 7: Map of virus-host interactions generated by aligning spacers detected with reference-guided methods in SS13 (A), SS19 (B), SS33 (C), SS37 (D), IC21 (E), and Cahuil/C34 (F) metagenomes against a library of haloviral genomes (see Supplementary Table S3). Viruses are marked in orange and spacers in blue. Visualization was performed with Cytoscape.

Figure 8: Measures of difference in dimer and codon usages between host and virus genomes matched with CRISPR spacers. The differences between each possible dimer and codon usage was subtracted, squared, and summed across all possible dimers or codons to generate a measure of dimer or codon difference. The average dimer usage and codon usage was calculated for all combinations of viruses with a particular host (A and B, respectively) and all virus-host combinations detected for a particular metagenome (C and D, respectively). All averages are reported with a 95% confidence interval (within two standard error) in black.

A

B

Figure 9: Detection of *cas* operons in metagenomes and affiliation of detected genes to representative taxa. Three metagenomes (Cahuil/C34, combined Chula Vista, and combined Santa Pola and Isla Cristina) were assembled and searched for contigs matching *cas* genes. Detection results are reported for five halobacterial *cas* operons (A). *cas* genes detected in no metagenomes are shown in red; genes detected in one metagenome are shown in blue; genes detected in two metagenomes are shown in green (A). Detected contigs assembled with Newbler and Velvet were further aligned against a library of *cas* genes from three halobacterial orders (Halobacteriales, Haloferacales, and Natrialbales). The taxonomic affiliations of the best hits for each contig were tabulated into profiles for each order and assembly method (B).

Supplementary Information

Figure S1: Percentage of reads mapping to reference library in taxonomic profiling steps (MetaPhyler)

Figure S2: Percentage of reads mapping to reference library in functional profiling steps (ShotMAP)

Figure S3: Number of CRISPR spacers detected by *de novo* and reference guided methods, normalized by millions of reads in each metagenome

Figure S4: Network of virus-host interactions generated from *Haloquadratum walsbyi* strain C23 spacer alignments against the haloviral genome library used in this study. Viruses are marked in blue, while spacers are in orange.

Figure S5: Measures of the number of contig BLAST hits to *cas* genes extracted from three different halobacterial orders (Halobacteriales, Haloferacales, and Natrialbales)

Figure S6: Taxonomic affiliation of all BLAST hits of *cas* genes against contigs (not only best hits)

Table S1: Additional information about the Santa Pola, Isla Cristina, and Cahuil solar saltern metagenomes provided in the NCBI Sequence Read Archive (SRA) and associated genome announcements (21–23). Site names (as listed in the Figures) are listed in bold. NCBI SRA accession numbers (SRX for experiment and SRR for run) are included in the table.

Site	Salt content	SRX number	SRR number	Sequencing technology	Number of reads (millions)	Average read length (bp)	Total length (Mbp)
SS13	13%	SRX328504	SRR944625	454	1.5	479	722.1
SS19	19%	SRX024859	SRR316684	454	0.293	504	148.9
IC21	21%	SRX352042	SRR988245	454	1.2	508	627
SS33	33%	SRX347883	SRR979792	454	0.971	429	420.1
SS37	37%	SRX090229	SRR328983	454	0.741	511	381.6
Cahuil (C34)	34%	SRX680116	SRR1549536	454	0.222	523	117.0

Table S2: Additional information about the Chula Vista saltern metagenomes obtained from the iMicrobe Collaborative (55) and combined into three metagenomes based on salt concentration. Site names (as listed in the Figures) are in bold. The datasets can be found at this web address: <http://data.imicrobe.us/project/view/58>.

Salt content (and label in figures in bold)	Dates of included metagenomes	Sequencing technology	Number of reads (millions)	Average read length (bp)	Total length (Mbp)
Low (6-8%) (CV6-8)	07/2004 11/10/2005 11/28/2005	454	0.352	95	33.4
Medium (12-14%) (CV12-14)	11/10/2005 11/11/2005 11/16/2005 11/28/2005	454	0.223	99	22.1
High (27-30%) (CV27-30)	07/2004 11/28/2005	454	0.618	102	63.2

Table S3: The library of haloviral genomes examined included 326 contigs more than 500 bp in length assembled from Chula Vista metaviromes (<http://data.imicrobe.us/project/view/58>) and the following haloviral genomes obtained from NCBI GenBank.

GI	GB	Description
82492114	DQ238866.1	Environmental halophage 1 AAJ-2005, partial genome
383398351	JQ807263.1	Environmental Halophage eHP-E5, partial genome
383398308	JQ807262.1	Environmental Halophage eHP-D7, partial genome
383398280	JQ807261.1	Environmental Halophage eHP-42, partial genome
383398256	JQ807260.1	Environmental Halophage eHP-41, partial genome
383398200	JQ807259.1	Environmental Halophage eHP-40, partial genome
383398172	JQ807258.1	Environmental Halophage eHP-39, partial genome
383398140	JQ807257.1	Environmental Halophage eHP-38, partial genome
383398101	JQ807256.1	Environmental Halophage eHP-37, partial genome
383398056	JQ807255.1	Environmental Halophage eHP-36, partial genome
383398012	JQ807254.1	Environmental Halophage eHP-35, partial genome
383397960	JQ807253.1	Environmental Halophage eHP-34, partial genome
383397929	JQ807252.1	Environmental Halophage eHP-33, partial genome
383397879	JQ807251.1	Environmental Halophage eHP-32, partial genome
383397836	JQ807250.1	Environmental Halophage eHP-31, partial genome
383397780	JQ807249.1	Environmental Halophage eHP-30, partial genome
383397746	JQ807248.1	Environmental Halophage eHP-29, partial genome
383397713	JQ807247.1	Environmental Halophage eHP-28, partial genome
383397672	JQ807246.1	Environmental Halophage eHP-27, partial genome
383397640	JQ807245.1	Environmental Halophage eHP-25, partial genome
383397593	JQ807244.1	Environmental Halophage eHP-24, partial genome
383397546	JQ807243.1	Environmental Halophage eHP-23, partial genome
383397498	JQ807242.1	Environmental Halophage eHP-22, partial genome
383397450	JQ807241.1	Environmental Halophage eHP-20, partial genome
383397419	JQ807240.1	Environmental Halophage eHP-19, partial genome
383397370	JQ807239.1	Environmental Halophage eHP-18, partial genome
383397340	JQ807238.1	Environmental Halophage eHP-17, partial genome
383397303	JQ807237.1	Environmental Halophage eHP-16, partial genome
383397240	JQ807236.1	Environmental Halophage eHP-15, partial genome
383397189	JQ807235.1	Environmental Halophage eHP-14, partial genome
383397132	JQ807234.1	Environmental Halophage eHP-13, partial genome
383397082	JQ807233.1	Environmental Halophage eHP-12, partial genome
383397053	JQ807232.1	Environmental Halophage eHP-11, partial genome
383397015	JQ807231.1	Environmental Halophage eHP-10, partial genome
383396977	JQ807230.1	Environmental Halophage eHP-9, partial genome
383396921	JQ807229.1	Environmental Halophage eHP-8, partial genome
383396883	JQ807228.1	Environmental Halophage eHP-7, partial genome

383396834	JQ807227.1	Environmental Halophage eHP-6, partial genome
383396790	JQ807226.1	Environmental Halophage eHP-5, partial genome
383396741	JQ807225.1	Environmental Halophage eHP-4, partial genome
383396711	JQ807224.1	Environmental Halophage eHP-3, partial genome
383396670	JQ807223.1	Environmental Halophage eHP-2, partial genome
383396626	JQ807222.1	Environmental Halophage eHP-1, partial genome
18138397	NC_003345.1	Halorubrum phage HF2, complete genome
18000338	AF222060.1	Halorubrum phage HF2, complete genome
505833397	KC292029.1	Halovirus HCTV-1, complete genome
505833276	KC292028.1	Halovirus HCTV-2, complete genome
505833087	KC292027.1	Halovirus HCTV-5, complete genome
32346393	AY190604.1	Halovirus HF1, complete genome
505832621	KC292026.1	Halovirus HGTV-1, complete genome
505832536	KC292025.1	Halovirus HHTV-1, complete genome
505832447	KC292024.1	Halovirus HHTV-2, complete genome
505832373	KC292023.1	Halovirus HRTV-4, complete genome
505832254	KC292022.1	Halovirus HRTV-5, complete genome
505832148	KC292021.1	Halovirus HRTV-7, complete genome
505832020	KC292020.1	Halovirus HRTV-8, complete genome
441462254	KC117378.1	Halovirus HSTV-1, complete genome
441461976	KC117376.1	Halovirus HSTV-2, complete genome
441462080	KC117377.1	Halovirus HVTV-1, complete genome
442736110	KC252997.1	Halovirus PH1, complete genome
660731908	KF771641.1	Halovirus VNH-1 genomic sequence
33338237	AF191796.1	His1 virus, complete genome
78172403	AF191797.1	His2 virus, complete genome
751367590	AKVG01000001.1	Hypersaline lake metagenome LTV1, whole genome shotgun sequence
751367589	AKVG01000002.1	Hypersaline lake metagenome LTV2, whole genome shotgun sequence
751367588	AKVG01000003.1	Hypersaline lake metagenome LTVLE3, whole genome shotgun sequence
751367587	AKVG01000004.1	Hypersaline lake metagenome LTV4, whole genome shotgun sequence
751367586	AKVG01000005.1	Hypersaline lake metagenome LTV5, whole genome shotgun sequence
751367585	AKVG01000006.1	Hypersaline lake metagenome LTV6, whole genome shotgun sequence
751367584	AKVG01000007.1	Hypersaline lake metagenome LTV7, whole genome shotgun sequence
751367583	AKVG01000008.1	Hypersaline lake metagenome LTVcontig54590, whole genome shotgun sequence
751367582	AKVG01000009.1	Hypersaline lake metagenome LTVcontig02397, whole genome shotgun sequence
751367581	AKVG01000010.1	Hypersaline lake metagenome LTVcontig822605, whole genome shotgun sequence

751367580	AKVG01000011.1	Hypersaline lake metagenome LTVcontig823649, whole genome shotgun sequence
751367579	AKVG01000012.1	Hypersaline lake metagenome LTVcontig823798, whole genome shotgun sequence
751367578	AKVG01000013.1	Hypersaline lake metagenome LTVcontig824981, whole genome shotgun sequence
751367577	AKVG01000014.1	Hypersaline lake metagenome LTVcontig3968333, whole genome shotgun sequence
751367576	AKVG01000015.1	Hypersaline lake metagenome LTVcontig3969208, whole genome shotgun sequence
751367575	AKVG01000016.1	Hypersaline lake metagenome LTVcontig999000, whole genome shotgun sequence
751367574	AKVG01000017.1	Hypersaline lake metagenome LTVcontig999004, whole genome shotgun sequence
751367573	AKVG01000018.1	Hypersaline lake metagenome LTVcontig998975, whole genome shotgun sequence
751367572	AKVG01000019.1	Hypersaline lake metagenome LTVcontig999006, whole genome shotgun sequence
751367571	AKVG01000020.1	Hypersaline lake metagenome LTVcontig1100059, whole genome shotgun sequence
751367570	AKVG01000021.1	Hypersaline lake metagenome LTVcontig1100028, whole genome shotgun sequence
751367569	AKVG01000022.1	Hypersaline lake metagenome LTVcontig1013670, whole genome shotgun sequence
751367568	AKVG01000023.1	Hypersaline lake metagenome LTVcontig1013956, whole genome shotgun sequence
751367567	AKVG01000024.1	Hypersaline lake metagenome LTVcontig1013803, whole genome shotgun sequence
751367566	AKVG01000025.1	Hypersaline lake metagenome LTVcontig1013342, whole genome shotgun sequence
751367565	AKVG01000026.1	Hypersaline lake metagenome LTVcontig1013519, whole genome shotgun sequence
751367564	AKVG01000027.1	Hypersaline lake metagenome LTVcontig224955, whole genome shotgun sequence
751367563	AKVG01000028.1	Hypersaline lake metagenome LTVcontig1012758, whole genome shotgun sequence
751367562	AKVG01000029.1	Hypersaline lake metagenome LTVcontig1012403, whole genome shotgun sequence
751367561	AKVG01000030.1	Hypersaline lake metagenome LTVcontig761242, whole genome shotgun sequence
751367560	AKVG01000031.1	Hypersaline lake metagenome LTVcontig1014414, whole genome shotgun sequence
751367559	AKVG01000032.1	Hypersaline lake metagenome LTVcontig1015495, whole genome shotgun sequence
751367558	AKVG01000033.1	Hypersaline lake metagenome LTVcontig1015890, whole genome shotgun sequence
751367557	AKVG01000034.1	Hypersaline lake metagenome LTVcontig1013394, whole genome shotgun sequence
751367556	AKVG01000035.1	Hypersaline lake metagenome LTVcontig1015952, whole genome shotgun sequence

Table S4: Sequence assembly statistics

	Cahuil		Combined Santa Pola		Combined Chula Vista	
Assembler	Newbler	Velvet	Newbler	Velvet	Newbler	Velvet
N50 value	881*	205	996*	106	882*	102
Average contig length (100 bp cutoff)	687	239	709	200	315	152
Number of contigs (100 bp cutoff)	8,167	140,071	156,214	1,639,903	18,519	75,046

Table S5: Alignments of combined Santa Pola CRISPR direct repeats identified with Crass against a library of taxonomically annotated direct repeat sequences obtained from CRISPRdb

Crass DR	Taxonomic affiliation of aligned DR	Percent identity	Alignment length	E-value	Score
G2DR1	Natronomonas pharaonis DSM 2160	100	28	2.00E-11	56
G2DR1	Natronomonas pharaonis DSM 2160	100	28	2.00E-11	56
G2DR1	Haloferax volcanii DS2	100	28	2.00E-11	56
G2DR1	Halogeometricum borinquense DSM 11551	100	26	2.00E-10	52
G2DR1	Natrinema sp. J7-2	96.43	28	4.00E-09	48.1
G2DR1	Halophilic archaeon DL31	96.43	28	4.00E-09	48.1
G2DR1	Haloarcula hispanica ATCC 33960	96.43	28	4.00E-09	48.1
G2DR1	Haloarcula hispanica N601	96.43	28	4.00E-09	48.1
G2DR1	Natronobacterium gregoryi SP2	96.43	28	4.00E-09	48.1
G2DR1	Haloferax volcanii DS2	96.43	28	4.00E-09	48.1
G2DR1	Haloferax volcanii DS2	96.43	28	4.00E-09	48.1
G2DR1	Haloarcula marismortui ATCC 43049	96.43	28	4.00E-09	48.1
G2DR1	Halomicrobium mukohataei DSM 12286	92.86	28	9.00E-07	40.1
G2DR1	Halomicrobium mukohataei DSM 12286	92.86	28	9.00E-07	40.1
G2DR1	Halorhabdus tiamatea SARL4B	92.86	28	9.00E-07	40.1
G2DR1	Halorhabdus utahensis DSM 12940	92.86	28	9.00E-07	40.1
G7DR1	Dinoroseobacter shibae DFL 12	96.55	29	1.00E-09	50.1
G7DR1	Methylococcus capsulatus str. Bath	96.43	28	4.00E-09	48.1
G7DR1	Rhodothermus marinus DSM 4252	100	20	9.00E-07	40.1
G7DR1	Alkalilimnicola ehrlichii MLHE-1	92.86	28	9.00E-07	40.1
G37DR1	Haloarcula hispanica ATCC 33960	93.33	30	6.00E-08	44.1
G37DR1	Haloarcula hispanica N601	93.33	30	6.00E-08	44.1
G37DR1	Natronobacterium gregoryi SP2	93.33	30	6.00E-08	44.1
G38DR1	Halorubrum lacusprofundi ATCC 49239	100	25	1.00E-09	50.1

G38DR1	Halophilic archaeon DL31	96	25	4.00E-07	42.1
G40DR1	Haloarcula hispanica ATCC 33960	93.33	30	7.00E-08	44.1
G40DR1	Haloarcula hispanica N601	93.33	30	7.00E-08	44.1
G40DR1	Natronobacterium gregoryi SP2	93.33	30	7.00E-08	44.1
G42DR1	Natronomonas pharaonis DSM 2160	100	30	1.00E-12	60
G42DR1	Natronomonas pharaonis DSM 2160	100	30	1.00E-12	60
G42DR1	Haloferax volcanii DS2	100	30	1.00E-12	60
G42DR1	Natrinema sp. J7-2	96.67	30	3.00E-10	52
G42DR1	Halophilic archaeon DL31	96.67	30	3.00E-10	52
G42DR1	Haloarcula hispanica ATCC 33960	96.67	30	3.00E-10	52
G42DR1	Haloarcula hispanica N601	96.67	30	3.00E-10	52
G42DR1	Natronobacterium gregoryi SP2	96.67	30	3.00E-10	52
G42DR1	Halogeometricum borinquense DSM 11551	100	26	3.00E-10	52
G42DR1	Haloferax volcanii DS2	96.67	30	3.00E-10	52
G42DR1	Haloferax volcanii DS2	96.67	30	3.00E-10	52
G42DR1	Haloarcula marismortui ATCC 43049	96.67	30	3.00E-10	52
G42DR1	Halomicrobium mukohataei DSM 12286	93.33	30	6.00E-08	44.1
G42DR1	Halomicrobium mukohataei DSM 12286	93.33	30	6.00E-08	44.1
G42DR1	Halorhabdus tiamatea SARL4B	93.33	30	6.00E-08	44.1
G42DR1	Halorhabdus utahensis DSM 12940	93.33	30	6.00E-08	44.1
G42DR1	Natrialba magadii ATCC 43099	93.1	29	3.00E-07	42.1
G43DR1	Natronomonas pharaonis DSM 2160	93.33	30	6.00E-08	44.1
G43DR1	Natronomonas pharaonis DSM 2160	93.33	30	6.00E-08	44.1
G43DR1	Haloferax volcanii DS2	93.33	30	6.00E-08	44.1
G57DR1	Saccharomonospora viridis DSM 43017	100	23	1.00E-08	46.1
G57DR1	Methylococcus capsulatus str. Bath	100	23	1.00E-08	46.1
G59DR1	Geobacter uraniireducens Rf4	100	25	1.00E-09	50.1
G59DR1	Komagataeibacter medellinensis NBRC 3288	96	25	3.00E-07	42.1
G59DR1	Gluconacetobacter diazotrophicus PA1 5	96	25	3.00E-07	42.1
G59DR1	Gluconacetobacter diazotrophicus PA1 5	96	25	3.00E-07	42.1
G82DR1	Natrinema sp. J7-2	96.67	30	3.00E-10	52
G82DR1	Natrinema sp. J7-2	100	26	3.00E-10	52
G82DR1	Haloarcula marismortui ATCC 43049	96.67	30	3.00E-10	52
G82DR1	Halophilic archaeon DL31	93.33	30	7.00E-08	44.1
G82DR1	Halomicrobium mukohataei DSM 12286	100	22	7.00E-08	44.1
G82DR1	Halomicrobium mukohataei DSM 12286	100	22	7.00E-08	44.1
G82DR1	Halorhabdus tiamatea SARL4B	93.33	30	7.00E-08	44.1
G82DR1	Halorhabdus utahensis DSM 12940	93.33	30	7.00E-08	44.1
G82DR1	Natronomonas pharaonis DSM 2160	93.33	30	7.00E-08	44.1
G82DR1	Natronomonas pharaonis DSM 2160	93.33	30	7.00E-08	44.1
G82DR1	Haloferax volcanii DS2	93.33	30	7.00E-08	44.1

G88DR1	Natrialba magadii ATCC 43099	93.33	30	6.00E-08	44.1
G88DR1	Halomicrobium mukohataei DSM 12286	100	20	1.00E-06	40.1
G88DR1	Halomicrobium mukohataei DSM 12286	100	20	1.00E-06	40.1
G110DR1	Saccharomonospora viridis DSM 43017	95.83	24	1.00E-06	40.1
G110DR1	Rhodospirillum rubrum ATCC 11170	95.83	24	1.00E-06	40.1
G110DR1	Rhodospirillum rubrum F11	95.83	24	1.00E-06	40.1
G121DR1	Natrinema sp. J7-2	96.67	30	3.00E-10	52
G121DR1	Halogeometricum borinquense DSM 11551	96.67	30	3.00E-10	52
G121DR1	Haloarcula marismortui ATCC 43049	96.67	30	3.00E-10	52
G121DR1	Haloarcula marismortui ATCC 43049	100	26	3.00E-10	52
G121DR1	Haloferax mediterranei ATCC 33500	93.33	30	7.00E-08	44.1
G121DR1	Halophilic archaeon DL31	96.15	26	7.00E-08	44.1
G121DR1	Halomicrobium mukohataei DSM 12286	96.15	26	7.00E-08	44.1
G121DR1	Halomicrobium mukohataei DSM 12286	96.15	26	7.00E-08	44.1
G121DR1	Halorhabdus tiamatea SARL4B	96.15	26	7.00E-08	44.1
G121DR1	Halorhabdus utahensis DSM 12940	96.15	26	7.00E-08	44.1
G121DR1	Natronomonas pharaonis DSM 2160	96.15	26	7.00E-08	44.1
G121DR1	Natronomonas pharaonis DSM 2160	96.15	26	7.00E-08	44.1
G121DR1	Haloferax volcanii DS2	96.15	26	7.00E-08	44.1
G467DR1	Haloquadratum walsbyi C23	100	30	1.00E-12	60
G467DR1	Haloquadratum walsbyi C23	96.67	30	3.00E-10	52
G467DR1	Haloquadratum walsbyi DSM 16790	96	25	3.00E-07	42.1

Table S6: Query sequences used for reference-guided CRISPR search and associated archaeal hosts

Pattern (Query) Number	Query DR Taxonomic Affiliation	Query Sequence
0	Haloferax volcanii DS2	GTTTCAGACGAACCCCTTGTGGGGTTGAAGC
1	Haloquadratum walsbyi DSM 16790	GTTTCAGATGAACCCCTTGATGGGTT
2	Halorubrum lacusprofundi ATCC 49239	TCGCCGTTCTCCTCTTCGCCGTTCTC
3	Halorubrum lacusprofundi ATCC 49239	GCTTCAAACCCACGAGGGGTTGCGCTGAAAG
4	Halorubrum lacusprofundi ATCC 49239	ACTACCCCACTATTTCGCTGTTAG
5	Halorubrum lacusprofundi ATCC 49239	GTCGAGACACTCAGAAAACCCAGCACGGGATTGAAAC
6	Haloferax volcanii DS2	GTTTCAGACGAACCCCTTGTGGGATTGAAGC
7	Haloferax volcanii DS2	GGTTTCAGACGAACCCCTTGTGGGTTTGAAGC
8	Halogeometricum borinquense DSM 11551	GCTTCAACCCCAAGGGTTCGTCTGTAAC
9	Haloquadratum walsbyi C23	GTTGCAACGAAGAGAAAACCCGCTAAGGGATTGAAAC
10	Haloquadratum walsbyi C23	GTTTCAGATGAACCCCTTGTGGGTTGAAGT
11	Haloquadratum walsbyi C23	GTTTCAGATGAACCCCTTGATGGGTTGAAGT
12	Haloferax mediterranei ATCC 33500	GGTTACAGACGGACCCTCGTTGGGTTGAAG

13	<i>Haloferax mediterranei</i> ATCC 33500	GTTACAGACGAACCCTAGTTGGGTTGAAGC
14	<i>Haloferax mediterranei</i> ATCC 33500	GTTACAGACGAACCCTAGTTGGGTTGAAGC
15	<i>Haloferax mediterranei</i> ATCC 33500	GCTTCAACCCAATTAGGGTTCGTCTGTAAC
16	<i>Haloarcula marismortui</i> ATCC 43049	GCTTCAACCCCACAAGGGTCCGTCTGAAAC
17	<i>Haloarcula marismortui</i> ATCC 43049	GTTACAGACGGACCCTCGTGGGGTTGAAGC
18	<i>Haloarcula marismortui</i> ATCC 43049	GTTTCAGACGGACCCTTGGGCGGTTGAAGT
19	<i>Natronomonas pharaonis</i> DSM 2160	GTTTCAGACGAACCCTTGTGGGGTTGAAGC
20	<i>Natronomonas pharaonis</i> DSM 2160	GTTTCAGACGAACCCTTGTGGGGTTGAAGC
21	<i>Natronomonas pharaonis</i> DSM 2160	GTCGAGACGGACTGAAAACCCAGAACGGGATTGAAAC
22	<i>Natronomonas pharaonis</i> DSM 2160	GTCGAGACGGACTGAAAACCCAGAACGGGATTGAAAC
23	<i>Halorhabdus utahensis</i> DSM 12940	GCTTCAACCCCACGAGGGTCCGTCTGAAAC
24	<i>Halorhabdus tiamatea</i> SARL4B	GCTTCAACCCCACGAGGGTCCGTCTGAAAC
25	<i>Halomicrobium mukohataei</i> DSM 12286	GCTTCAATCCCACAAGGGTCCGTCTGAAAC
26	<i>Halomicrobium mukohataei</i> DSM 12286	GCTTCAATCCCACAAGGGTCCGTCTGAAAC
27	<i>Haloarcula hispanica</i> N601	GCTTCAACCCCACGAGGGTTCGTCTGAAAC
28	<i>Haloarcula hispanica</i> ATCC 33960	GCTTCAACCCCACGAGGGTTCGTCTGAAAC

Table S7: Summary of CRISPR virus-host pairings in Cahuil (C34) metagenome; CRISPR spacers were detected with the reference-guided method

Putative host	Matched virus
<i>Haloquadratum walsbyi</i>	CVcontig00192
<i>Haloquadratum walsbyi</i>	LTVcontig998975
<i>Haloquadratum walsbyi</i>	eHP-11
<i>Haloquadratum walsbyi</i>	CVcontig00213
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-31
<i>Haloarcula marismortui</i>	CVcontig00192
<i>Haloarcula marismortui</i>	eHP-11

Table S8: Summary of CRISPR virus-host pairings in SS13 metagenome; CRISPR spacers were detected with the reference-guided method

Putative host	Matched virus
<i>Haloquadratum walsbyi</i>	CVcontig00305

Table S9: Summary of CRISPR virus-host pairings in SS19 metagenome; CRISPR spacers were detected with the reference-guided method

Putative host	Matched virus
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	CVcontig00024
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-37
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-38
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-41
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-42
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-D7
<i>Halorhabdus utahensis</i> (or <i>Haloarcula hispanica</i>)	eHP-E5
<i>Haloquadratum walsbyi</i>	eHP-2
<i>Haloquadratum walsbyi</i>	eHP-42
<i>Haloquadratum walsbyi</i>	eHP-5
<i>Haloquadratum walsbyi</i>	eHP-29

Haloquadratum walsbyi	eHP-E5
Haloquadratum walsbyi	eHP-22
Haloquadratum walsbyi	eHP-24
Haloquadratum walsbyi	eHP-37
Haloquadratum walsbyi	eHP-38
Haloquadratum walsbyi	eHP-39
Haloquadratum walsbyi	eHP-41
Haloquadratum walsbyi	eHP-D7

Table S10: Summary of CRISPR virus-host pairings in IC21 metagenome; CRISPR spacers were detected with the reference-guided method

Putative host	Matched virus
Haloferax volcanii	eHP-10
Haloferax volcanii	HRTV-4
Haloarcula marismortui	eHP-15
Haloarcula marismortui	eHP-1
Haloarcula marismortui	eHP-19
Haloarcula marismortui	eHP-34
Haloquadratum walsbyi	CVcontig00020
Haloquadratum walsbyi	eHP-32
Haloquadratum walsbyi	CVcontig00081

Table S11: Summary of CRISPR virus-host pairings in SS33 metagenome; CRISPR spacers were detected with the reference-guided method

Putative host	Matched virus
Haloquadratum walsbyi	eHP-2
Haloquadratum walsbyi	eHP-42
Haloquadratum walsbyi	eHP-5
Haloquadratum walsbyi	eHP-29
Haloquadratum walsbyi	eHP-E5
Haloquadratum walsbyi	eHP-22
Haloquadratum walsbyi	eHP-24
Haloquadratum walsbyi	eHP-37
Haloquadratum walsbyi	eHP-38
Haloquadratum walsbyi	eHP-39
Haloquadratum walsbyi	eHP-41
Haloquadratum walsbyi	eHP-D7
Haloquadratum walsbyi	eHP-40
Haloarcula hispanica (or Halorhabdus utahensis)	eHP-4
Haloarcula hispanica (or Halorhabdus utahensis)	eHP-8
Haloarcula hispanica (or Halorhabdus utahensis)	eHP-13

Table S12: Summary of CRISPR virus-host pairings in SS37 metagenome; CRISPR spacers were detected with the reference-guided method

Putative host	Matched virus
Haloquadratum walsbyi	eHP-2
Haloquadratum walsbyi	eHP-42
Haloquadratum walsbyi	eHP-5
Haloquadratum walsbyi	eHP-29
Haloquadratum walsbyi	eHP-E5
Haloquadratum walsbyi	eHP-22
Haloquadratum walsbyi	eHP-24
Haloquadratum walsbyi	eHP-37
Haloquadratum walsbyi	eHP-38
Haloquadratum walsbyi	eHP-39
Haloquadratum walsbyi	eHP-41
Haloquadratum walsbyi	eHP-D7
Halorhabdus utahensis (or Haloarcula hispanica)	CVcontig00024

Halorhabdus utahensis (or Haloarcula hispanica)	eHP-37
Halorhabdus utahensis (or Haloarcula hispanica)	eHP-38
Halorhabdus utahensis (or Haloarcula hispanica)	eHP-41
Halorhabdus utahensis (or Haloarcula hispanica)	eHP-42
Halorhabdus utahensis (or Haloarcula hispanica)	eHP-D7
Halorhabdus utahensis (or Haloarcula hispanica)	eHP-E5

Table S13: The library of *cas* genes identified in Halobacteriales genomes that were aligned against Newbler contigs assembled from the Cahuil/C34, combined Chula Vista, and combined Santa Pola/Isla Cristina metagenomes

GI	GB	Description
749702013	AJF26807.1	CRISPR-associated protein [Haloarcula sp. CBA1115]
445788732	EMA39438.1	CRISPR-associated protein, partial [Halococcus morrhuae DSM 1307]
320548909	EFW90576.1	CRISPR-associated protein [Haladaptatus paucihalophilus DX253]
55228999	AAV44422.1	CRISPR-associated protein (plasmid) [Haloarcula marismortui ATCC 43049]
491183000	WP_005041362.1	crispr-associated protein [Halococcus salifodinae]
445804137	EMA54400.1	crispr-associated protein [Halococcus salifodinae DSM 8989]
910232731	WP_050037906.1	CRISPR-associated protein [Haloarcula sp. CBA1115]
910017809	WP_049972516.1	CRISPR-associated protein [Haladaptatus cibarius]
257171180	ACV48939.1	CRISPR-associated protein, Csh2 family [Halomicrobium mukohataei DSM 12286]
256691292	ACV11629.1	CRISPR-associated protein, Csh2 family [Halorhabdus utahensis DSM 12940]
257171185	ACV48944.1	CRISPR-associated protein Cas2 [Halomicrobium mukohataei DSM 12286]
257171184	ACV48943.1	CRISPR-associated protein Cas1 [Halomicrobium mukohataei DSM 12286]
257171181	ACV48940.1	CRISPR-associated protein Cas5, Hmari subtype [Halomicrobium mukohataei DSM 12286]
256691291	ACV11628.1	CRISPR-associated protein Cas5, Hmari subtype [Halorhabdus utahensis DSM 12940]
256691288	ACV11625.1	CRISPR-associated protein Cas1 [Halorhabdus utahensis DSM 12940]
256691287	ACV11624.1	CRISPR-associated protein Cas2 [Halorhabdus utahensis DSM 12940]
257171183	ACV48942.1	CRISPR-associated protein Cas4 [Halomicrobium mukohataei DSM 12286]
256691289	ACV11626.1	CRISPR-associated protein Cas4 [Halorhabdus utahensis DSM 12940]
256691294	ACV11631.1	CRISPR-associated protein Cas6 [Halorhabdus utahensis DSM 12940]
257171179	ACV48938.1	CRISPR-associated protein, Csh1 family [Halomicrobium mukohataei DSM 12286]
257171178	ACV48937.1	CRISPR-associated protein Cas6 [Halomicrobium mukohataei DSM 12286]
256691293	ACV11630.1	CRISPR-associated protein, Csh1 family [Halorhabdus utahensis DSM 12940]
497448090	WP_009762288.1	CRISPR-associated HD domain protein [Halobacterium sp. DL1]

497448087	WP_009762285.1	CRISPR-associated protein Csh1 domain protein [Halobacterium sp. DL1]
543101072	ERJ06791.1	CRISPR-associated protein Csh2 family protein [Halorhabdus tiamatea SARL4B]
543101071	ERJ06790.1	CRISPR-associated protein Csh1 family protein [Halorhabdus tiamatea SARL4B]
499642613	WP_011323347.1	CRISPR-associated protein Cas3 [Natronomonas pharaonis]
445756790	EMA08148.1	CRISPR-associated protein Cas6 [Haloarcula sinaiiensis ATCC 33800]
749702016	AJF26810.1	CRISPR-associated protein Cas2 [Haloarcula sp. CBA1115]
749702015	AJF26809.1	CRISPR-associated protein Cas1 [Haloarcula sp. CBA1115]
749702014	AJF26808.1	CRISPR-associated protein Cas4 [Haloarcula sp. CBA1115]
749702011	AJF26805.1	CRISPR-associated protein Csc2 [Haloarcula sp. CBA1115]
749702010	AJF26804.1	CRISPR-associated protein Cas6 [Haloarcula sp. CBA1115]
573487046	AHG05451.1	CRISPR-associated protein Cas2 (plasmid) [Halobacterium sp. DL1]
573487045	AHG05450.1	CRISPR-associated protein Cas1 (plasmid) [Halobacterium sp. DL1]
573487042	AHG05447.1	CRISPR-associated protein Cas5 (plasmid) [Halobacterium sp. DL1]
573487041	AHG05446.1	CRISPR-associated protein Csh2 (plasmid) [Halobacterium sp. DL1]
573487040	AHG05445.1	CRISPR-associated protein Csh1 (plasmid) [Halobacterium sp. DL1]
573487039	AHG05444.1	CRISPR-associated protein Cas6 (plasmid) [Halobacterium sp. DL1]
491106984	WP_004966752.1	CRISPR-associated protein Cas6 [Haloarcula sinaiiensis]
573487044	AHG05449.1	CRISPR-associated protein Cas4 (plasmid) [Halobacterium sp. DL1]
556695528	ESP86785.1	CRISPR-associated protein Cas2 [Candidatus Halobonum tyrrellensis G22]
556695527	ESP86784.1	CRISPR-associated protein Cas1 [Candidatus Halobonum tyrrellensis G22]
556695524	ESP86781.1	CRISPR-associated protein Cas5, Hmari subtype [Candidatus Halobonum tyrrellensis G22]
556695523	ESP86780.1	CRISPR-associated protein Cas7/Csh2, subtype I-B/HMARI [Candidatus Halobonum tyrrellensis G22]
445806733	EMA56844.1	CRISPR-associated protein Cas5 [Halococcus thailandensis JCM 13552]
445806731	EMA56842.1	Csh2 family CRISPR-associated protein [Halococcus thailandensis JCM 13552]
445806729	EMA56840.1	CRISPR-associated protein Cas6 [Halococcus thailandensis JCM 13552]
445806440	EMA56569.1	CRISPR-associated Cas1 family protein [Halococcus thailandensis JCM 13552]
445806439	EMA56568.1	CRISPR-associated protein Cas2 [Halococcus thailandensis JCM 13552]
445804135	EMA54398.1	CRISPR-associated Cas5h family protein [Halococcus salifodinae DSM 8989]
445788730	EMA39436.1	CRISPR-associated protein Cas6 [Halococcus morrhuae DSM 1307]

445774126	EMA25148.1	CRISPR-associated protein Cas6 [Haloarcula argentinensis DSM 12282]
445774124	EMA25146.1	CRISPR-associated protein, Csh2 family [Haloarcula argentinensis DSM 12282]
445774123	EMA25145.1	CRISPR-associated protein Cas5, Hmari subtype [Haloarcula argentinensis DSM 12282]
445774120	EMA25142.1	CRISPR-associated protein Cas1 [Haloarcula argentinensis DSM 12282]
445774119	EMA25141.1	CRISPR-associated protein Cas2 [Haloarcula argentinensis DSM 12282]
445767708	EMA18802.1	Csh2 family CRISPR-associated protein [Haloarcula californiae ATCC 33799]
445767707	EMA18801.1	CRISPR-associated protein Cas5, Hmari subtype [Haloarcula californiae ATCC 33799]
445756797	EMA08155.1	CRISPR-associated protein Cas2 [Haloarcula sinaiensis ATCC 33800]
445756796	EMA08154.1	CRISPR-associated protein Cas1 [Haloarcula sinaiensis ATCC 33800]
445756792	EMA08150.1	CRISPR-associated protein, Csh2 family [Haloarcula sinaiensis ATCC 33800]
320548914	EFW90581.1	CRISPR-associated protein Cas2 [Haladaptatus paucihalophilus DX253]
320548913	EFW90580.1	CRISPR-associated protein Cas1 [Haladaptatus paucihalophilus DX253]
320548907	EFW90574.1	CRISPR-associated protein Cas6 [Haladaptatus paucihalophilus DX253]
497448092	WP_009762290.1	CRISPR-associated protein Cas2 [Halobacterium sp. DL1]
497448091	WP_009762289.1	CRISPR-associated protein Cas1 [Halobacterium sp. DL1]
497448089	WP_009762287.1	CRISPR-associated protein Cas5 [Halobacterium sp. DL1]
497448088	WP_009762286.1	CRISPR-associated protein Csh2 [Halobacterium sp. DL1]
494345545	WP_007188931.1	CRISPR-associated protein Cas5 [Haloarcula californiae]
491677582	WP_005533724.1	CRISPR-associated protein Cas5 [Haloarcula argentinensis]
491182996	WP_005041358.1	CRISPR-associated protein Cas5 [Halococcus salifodinae]
543416120	GAD52992.1	CRISPR-associated protein, Cas5h family [Halarchaeum acidiphilum MH1-52-1]
543414359	GAD53621.1	CRISPR-associated protein Csc2 [Halarchaeum acidiphilum MH1-52-1]
543414355	GAD53617.1	CRISPR-associated protein Cas1 [Halarchaeum acidiphilum MH1-52-1]
543414354	GAD53616.1	CRISPR-associated protein Cas2 [Halarchaeum acidiphilum MH1-52-1]
543413946	GAD53845.1	CRISPR-associated protein Cas1 [Halarchaeum acidiphilum MH1-52-1]
543413945	GAD53844.1	CRISPR-associated protein Cas2 [Halarchaeum acidiphilum MH1-52-1]
769188913	WP_044957276.1	CRISPR-associated protein Cas2 [Halarchaeum acidiphilum]
769188805	WP_044957195.1	CRISPR-associated protein Cas2 [Halarchaeum acidiphilum]
769188798	WP_044957190.1	CRISPR-associated protein Cas6 [Halarchaeum acidiphilum]
769182643	WP_044952751.1	CRISPR-associated protein Csh1 [Haloarcula hispanica]
769182640	WP_044952748.1	CRISPR-associated protein Cas6 [Haloarcula hispanica]

746356220	WP_039400886.1	CRISPR-associated protein Csh1 [haloarchaeon 3A1_DGR]
746356216	WP_039400882.1	CRISPR-associated protein Cas6 [haloarchaeon 3A1_DGR]
76558138	CAI49725.1	CRISPR-associated protein Csc1 [Natronomonas pharaonis DSM 2160]
76558137	CAI49724.1	CRISPR-associated protein Cas7 [Natronomonas pharaonis DSM 2160]
76558136	CAI49723.1	CRISPR-associated protein Cas10d [Natronomonas pharaonis DSM 2160]
76558134	CAI49721.1	CRISPR-associated DNA-binding protein Csm6 [Natronomonas pharaonis DSM 2160]
648486250	WP_026178001.1	CRISPR-associated protein Csh1 [Haladaptatus paucihalophilus]
648486249	WP_026178000.1	CRISPR-associated protein Cas5 [Haladaptatus paucihalophilus]
648486247	WP_026177998.1	CRISPR-associated protein Cas1 [Haladaptatus paucihalophilus]
528526430	CCQ33719.1	CRISPR-associated protein, Cas2 [Halorhabdus tiamatea SARL4B]
528526429	CCQ33718.1	CRISPR-associated protein, Cas1 [Halorhabdus tiamatea SARL4B]
528526426	CCQ33715.1	CRISPR-associated protein, Cas5h [Halorhabdus tiamatea SARL4B]
528526425	CCQ33714.1	CRISPR-associated protein, Csh2 [Halorhabdus tiamatea SARL4B]
528526424	CCQ33713.1	CRISPR-associated protein, Csh1 [Halorhabdus tiamatea SARL4B]
528526423	CCQ33712.1	CRISPR-associated protein, Cas6 [Halorhabdus tiamatea SARL4B]
564124632	AHB67952.1	CRISPR-associated protein Cas2 (plasmid) [Haloarcula hispanica N601]
564124631	AHB67951.1	CRISPR-associated protein Cas1 (plasmid) [Haloarcula hispanica N601]
564124628	AHB67948.1	CRISPR-associated protein Cas5 (plasmid) [Haloarcula hispanica N601]
564124627	AHB67947.1	CRISPR-associated protein Csh2 (plasmid) [Haloarcula hispanica N601]
564124626	AHB67946.1	CRISPR-associated protein Csh1 (plasmid) [Haloarcula hispanica N601]
564124625	AHB67945.1	CRISPR-associated protein Cas6 (plasmid) [Haloarcula hispanica N601]
543101077	ERJ06796.1	CRISPR-associated endoribonuclease Cas2 2 protein [Halorhabdus tiamatea SARL4B]
543101076	ERJ06795.1	CRISPR-associated endonuclease Cas1 2 protein [Halorhabdus tiamatea SARL4B]
543101074	ERJ06793.1	CRISPR-associated helicase Cas3 family protein [Halorhabdus tiamatea SARL4B]
543101073	ERJ06792.1	CRISPR-associated protein Cas5 Hmari subtype [Halorhabdus tiamatea SARL4B]
543101070	ERJ06789.1	CRISPR-associated protein Cas6 [Halorhabdus tiamatea SARL4B]
55229004	AAV44427.1	CRISPR-associated protein Cas2 (plasmid) [Haloarcula marismortui ATCC 43049]
55229003	AAV44426.1	CRISPR-associated protein Cas1 (plasmid) [Haloarcula marismortui ATCC 43049]
343784867	AEM58843.1	CRISPR-associated protein Cas6 [Haloarcula hispanica ATCC 33960]

343784866	AEM58842.1	CRISPR-associated protein, Csh1 family [Haloarcula hispanica ATCC 33960]
343784865	AEM58841.1	CRISPR-associated protein, Csh2 family [Haloarcula hispanica ATCC 33960]
343784864	AEM58840.1	CRISPR-associated protein Cas5, Hmari subtype [Haloarcula hispanica ATCC 33960]
343784861	AEM58837.1	CRISPR-associated protein Cas1 [Haloarcula hispanica ATCC 33960]
343784860	AEM58836.1	CRISPR-associated protein Cas2 [Haloarcula hispanica ATCC 33960]
557375040	WP_023396045.1	CRISPR-associated protein Cas1 [Candidatus Halobonum tyrrellensis]
557375036	WP_023396041.1	CRISPR-associated protein Cas7/Csh2, subtype I-B/HMARI [Candidatus Halobonum tyrrellensis]
557375035	WP_023396040.1	CRISPR-associated protein, Csh1 family [Candidatus Halobonum tyrrellensis]
556695522	ESP86779.1	CRISPR-associated protein, Csh1 family [Candidatus Halobonum tyrrellensis G22]
545903686	WP_021780686.1	CRISPR-associated protein Cas1 [Halarchaeum acidiphilum]
544640163	WP_021074704.1	CRISPR-associated protein Csh2 [haloarchaeon 3A1_DGR]
544640159	WP_021074700.1	CRISPR-associated protein Cas1 [haloarchaeon 3A1_DGR]
541202045	ERH10110.1	CRISPR-associated protein Cas5, Hmari subtype [halophilic archaeon J07HX64]
491106992	WP_004966758.1	CRISPR-associated protein Cas1 [Haloarcula sinaiensis]
491106986	WP_004966754.1	CRISPR-associated protein, Csh2 family [Haloarcula sinaiensis]
519065063	WP_020220938.1	CRISPR-associated protein Cas1 [Halarchaeum acidiphilum]
519065059	WP_020220934.1	CRISPR-associated protein Csh2 [Halarchaeum acidiphilum]
517069344	WP_018258162.1	CRISPR-associated protein Cas6 [Halomicrobium katesii]
517069343	WP_018258161.1	CRISPR-associated protein Csh1 [Halomicrobium katesii]
517069342	WP_018258160.1	CRISPR-associated protein Csh2 [Halomicrobium katesii]
517069341	WP_018258159.1	CRISPR-associated protein Cas5 [Halomicrobium katesii]
517069338	WP_018258156.1	CRISPR-associated protein Cas1 [Halomicrobium katesii]
495010973	WP_007736985.1	CRISPR-associated protein Csc2 [Halococcus thailandensis]
495010970	WP_007736982.1	CRISPR-associated protein Cas4 [Halococcus thailandensis]
495010465	WP_007736477.1	CRISPR-associated protein Cas5 [Halococcus thailandensis]
495010461	WP_007736473.1	CRISPR-associated protein, Csh1 family [Halococcus thailandensis]
494345549	WP_007188933.1	CRISPR-associated Csh1 family protein [Haloarcula californiae]
491677585	WP_005533727.1	CRISPR-associated protein Cas6 [Haloarcula argentinensis]
491677584	WP_005533726.1	CRISPR-associated protein, Csh1 family [Haloarcula argentinensis]
491677583	WP_005533725.1	CRISPR-associated protein, Csh2 family [Haloarcula argentinensis]
491677579	WP_005533721.1	CRISPR-associated protein Cas1 [Haloarcula argentinensis]
503796723	WP_014030717.1	CRISPR-associated protein Csh2 [Haloarcula hispanica]
503796722	WP_014030716.1	CRISPR-associated protein Cas5 [Haloarcula hispanica]
495800278	WP_008524857.1	CRISPR-associated protein Cas6 [Halorhabdus tiamatea]
495800276	WP_008524855.1	CRISPR-associated protein Csh1 [Halorhabdus tiamatea]

495800274	WP_008524853.1	CRISPR-associated protein Csh2 [Halorhabdus tiamatea]
495800273	WP_008524852.1	CRISPR-associated protein Cas5 [Halorhabdus tiamatea]
495800270	WP_008524849.1	CRISPR-associated protein Cas1 [Halorhabdus tiamatea]
495257574	WP_007982329.1	CRISPR-associated protein Csh2 [Haladaptatus paucihalophilus]
506269429	WP_015789204.1	CRISPR-associated protein Csh1 [Halorhabdus utahensis]
506269428	WP_015789203.1	CRISPR-associated protein Csh2 [Halorhabdus utahensis]
506269427	WP_015789202.1	CRISPR-associated protein Cas5 [Halorhabdus utahensis]
506244006	WP_015763781.1	CRISPR-associated protein Csh2 [Halomicrobium mukohataei]
506244005	WP_015763780.1	CRISPR-associated protein Csh1 [Halomicrobium mukohataei]
499642615	WP_011323349.1	CRISPR-associated protein Cas1 [Natronomonas pharaonis]
499642614	WP_011323348.1	CRISPR-associated protein Cas4 [Natronomonas pharaonis]
499642612	WP_011323346.1	CRISPR-associated protein Csc1 [Natronomonas pharaonis]
499642611	WP_011323345.1	CRISPR-associated protein Csc2 [Natronomonas pharaonis]
499642610	WP_011323344.1	CRISPR-associated protein Csc3 [Natronomonas pharaonis]
499642609	WP_011323343.1	CRISPR-associated protein Cas6 [Natronomonas pharaonis]
499541488	WP_011222271.1	CRISPR-associated protein Cas1 [Haloarcula marismortui]
499541485	WP_011222268.1	CRISPR-associated protein Cas5 [Haloarcula marismortui]
499541484	WP_011222267.1	CRISPR-associated protein Csh2 [Haloarcula marismortui]
499541483	WP_011222266.1	CRISPR-associated protein Csh1 [Haloarcula marismortui]
499541482	WP_011222265.1	CRISPR-associated protein Cas6 [Haloarcula marismortui]
506269430	WP_015789205.1	CRISPR-associated protein Cas6 [Halorhabdus utahensis]
506269424	WP_015789199.1	CRISPR-associated protein Cas1 [Halorhabdus utahensis]
506244010	WP_015763785.1	CRISPR-associated protein Cas1 [Halomicrobium mukohataei]
506244007	WP_015763782.1	CRISPR-associated protein Cas5 [Halomicrobium mukohataei]
506244004	WP_015763779.1	CRISPR-associated protein Cas6 [Halomicrobium mukohataei]
503796719	WP_014030713.1	CRISPR-associated protein Cas1 [Haloarcula hispanica]
445806730	EMA56841.1	CRISPR-associated protein, Csh1 family [Halococcus thailandensis JCM 13552]
445806444	EMA56573.1	CRISPR-associated protein Csc2 [Halococcus thailandensis JCM 13552]
445806441	EMA56570.1	CRISPR-associated protein Cas4 [Halococcus thailandensis JCM 13552]
445804136	EMA54399.1	CRISPR-associated protein Csh2 [Halococcus salifodinae DSM 8989]
445774125	EMA25147.1	CRISPR-associated protein, Csh1 family [Haloarcula argentinensis DSM 12282]
445767709	EMA18803.1	CRISPR-associated Csh1 family protein [Haloarcula californiae ATCC 33799]
445767704	EMA18798.1	CRISPR-associated Cas2 family protein [Haloarcula californiae ATCC 33799]
445756793	EMA08151.1	CRISPR-associated protein Cas5, Hmari subtype [Haloarcula sinaiensis ATCC 33800]
320548910	EFW90577.1	CRISPR-associated protein Cas5, Hmari subtype [Haladaptatus paucihalophilus DX253]
910243469	WP_050048135.1	CRISPR-associated protein Cas2 [Halanaeroarchaeum sulfireducens]

910243468	WP_050048134.1	CRISPR-associated protein Cas1 [Halanaeroarchaeum sulfurreducens]
910243467	WP_050048133.1	CRISPR-associated protein Cas4 [Halanaeroarchaeum sulfurreducens]
910243465	WP_050048131.1	CRISPR-associated protein Cas5 [Halanaeroarchaeum sulfurreducens]
910243464	WP_050048130.1	CRISPR-associated protein Csh2 [Halanaeroarchaeum sulfurreducens]
910243463	WP_050048129.1	CRISPR-associated protein Csh1 [Halanaeroarchaeum sulfurreducens]
910243462	WP_050048128.1	CRISPR-associated protein Cas6 [Halanaeroarchaeum sulfurreducens]
910232734	WP_050037909.1	CRISPR-associated protein Cas2 [Haloarcula sp. CBA1115]
910232733	WP_050037908.1	CRISPR-associated protein Cas1 [Haloarcula sp. CBA1115]
910232732	WP_050037907.1	CRISPR-associated protein Cas4 [Haloarcula sp. CBA1115]
910232729	WP_050037904.1	CRISPR-associated protein Csc2 [Haloarcula sp. CBA1115]
910232728	WP_050037903.1	CRISPR-associated protein Cas6 [Haloarcula sp. CBA1115]
497448086	WP_009762284.1	CRISPR-associated protein Cas6 [Halobacterium sp. DL1]
910017932	WP_049972639.1	CRISPR-associated protein Csh1 [Haladaptatus cibarius]
910017856	WP_049972563.1	CRISPR-associated protein Cas2 [Haladaptatus cibarius]
910017855	WP_049972562.1	CRISPR-associated protein Cas1 [Haladaptatus cibarius]
910017854	WP_049972561.1	CRISPR-associated protein Cas4 [Haladaptatus cibarius]
910017852	WP_049972559.1	CRISPR-associated protein Cas5 [Haladaptatus cibarius]
910017851	WP_049972558.1	CRISPR-associated protein Csh2 [Haladaptatus cibarius]
910017850	WP_049972557.1	CRISPR-associated protein Cas6 [Haladaptatus cibarius]
910017813	WP_049972520.1	CRISPR-associated protein Cas6 [Haladaptatus cibarius]
910017811	WP_049972518.1	CRISPR-associated protein Csc2 [Haladaptatus cibarius]
910017808	WP_049972515.1	CRISPR-associated protein Cas4 [Haladaptatus cibarius]
910017807	WP_049972514.1	CRISPR-associated protein Cas1 [Haladaptatus cibarius]
910017806	WP_049972513.1	CRISPR-associated protein Cas2 [Haladaptatus cibarius]
909704826	WP_049948027.1	CRISPR-associated protein Cas4 [Candidatus Halobonum tyrrellensis]
909704824	WP_049948025.1	CRISPR-associated protein Cas5 [Candidatus Halobonum tyrrellensis]
909704823	WP_049948024.1	CRISPR-associated protein Cas6 [Candidatus Halobonum tyrrellensis]
909701387	WP_049944624.1	CRISPR-associated protein Cas2 [Haloarcula californiae]
909697750	WP_049940987.1	CRISPR-associated protein Csh1 [Haladaptatus paucihalophilus]
909696352	WP_049939589.1	CRISPR-associated protein Cas2 [Natronomonas pharaonis]
909676135	WP_049919489.1	CRISPR-associated protein Cas4 [Haloarcula sinaiensis]
909676134	WP_049919488.1	CRISPR-associated protein Cas5 [Haloarcula sinaiensis]
909673035	WP_049916389.1	CRISPR-associated protein Cas6 [Halococcus morrhuae]
909669322	WP_049912676.1	CRISPR-associated protein Cas1 [Halococcus thailandensis]
909669240	WP_049912594.1	CRISPR-associated protein Cas6 [Halococcus thailandensis]
909660037	WP_049903450.1	CRISPR-associated protein Cas6 [Halococcus sp. 197A]
909660035	WP_049903448.1	CRISPR-associated protein Csh1 [Halococcus sp. 197A]

909660032	WP_049903445.1	CRISPR-associated protein Csh2 [Halococcus sp. 197A]
909660029	WP_049903442.1	CRISPR-associated protein Cas5 [Halococcus sp. 197A]
909660024	WP_049903437.1	CRISPR-associated protein Cas4 [Halococcus sp. 197A]
909660021	WP_049903434.1	CRISPR-associated protein Cas1 [Halococcus sp. 197A]
909660019	WP_049903432.1	CRISPR-associated protein Cas2 [Halococcus sp. 197A]
908685970	WP_049817745.1	CRISPR-associated protein Csh1 [haloarchaeon 3A1_DGR]
822661539	AKH97363.1	CRISPR-associated protein Cas2 [Halanaeroarchaeum sulfureducens]
822661538	AKH97362.1	CRISPR-associated protein Cas1 [Halanaeroarchaeum sulfureducens]
822661537	AKH97361.1	CRISPR-associated protein Cas4 [Halanaeroarchaeum sulfureducens]
822661536	AKH97360.1	CRISPR-associated protein Cas3 [Halanaeroarchaeum sulfureducens]
822661535	AKH97359.1	CRISPR-associated protein Cas5 [Halanaeroarchaeum sulfureducens]
822661534	AKH97358.1	CRISPR-associated protein Cas7/Csh2 [Halanaeroarchaeum sulfureducens]
822661533	AKH97357.1	CRISPR-associated protein Cas8b/Csh1 [Halanaeroarchaeum sulfureducens]
822661532	AKH97356.1	CRISPR-associated protein Cas6 [Halanaeroarchaeum sulfureducens]
769188803	WP_044957193.1	CRISPR-associated protein Cas4 [Halarchaeum acidiphilum]
746356226	WP_039400892.1	CRISPR-associated protein Cas4 [haloarchaeon 3A1_DGR]
648486248	WP_026177999.1	CRISPR-associated protein Cas4 [Haladaptatus paucihalophilus]
528526428	CCQ33717.1	CRISPR-associated protein, Cas4a [Halorhabdus tiamatea SARL4B]
564124630	AHB67950.1	CRISPR-associated protein Cas4 (plasmid) [Haloarcula hispanica N601]
543101075	ERJ06794.1	CRISPR-associated protein Cas4 [Halorhabdus tiamatea SARL4B]
55229002	AAV44425.1	CRISPR-associated protein Cas4 (plasmid) [Haloarcula marismortui ATCC 43049]
343784862	AEM58838.1	CRISPR-associated protein Cas4 [Haloarcula hispanica ATCC 33960]
556695526	ESP86783.1	CRISPR-associated protein Cas4 [Candidatus Halobonum tyrrellensis G22]
517069339	WP_018258157.1	CRISPR-associated protein Cas4 [Halomicrobium katesii]
491677580	WP_005533722.1	CRISPR-associated protein Cas4 [Haloarcula argentinensis]
491182998	WP_005041360.1	CRISPR-associated protein Csh2 [Halococcus salifodinae]
495800271	WP_008524850.1	CRISPR-associated protein Cas4 [Halorhabdus tiamatea]
499541487	WP_011222270.1	CRISPR-associated protein Cas4 [Haloarcula marismortui]
506269425	WP_015789200.1	CRISPR-associated protein Cas4 [Halorhabdus utahensis]
506244009	WP_015763784.1	CRISPR-associated protein Cas4 [Halomicrobium mukohataei]
503796720	WP_014030714.1	CRISPR-associated protein Cas4 [Haloarcula hispanica]
445774121	EMA25143.1	CRISPR-associated protein Cas4 [Haloarcula argentinensis DSM 12282]
445756795	EMA08153.1	CRISPR-associated protein Cas4 [Haloarcula sinaiensis ATCC 33800]

320548912	EFW90579.1	CRISPR-associated Cas4 family protein [Haladaptatus paucihalophilus DX253]
543414360	GAD53622.1	CRISPR associated protein Csc3 [Halarchaeum acidiphilum MH1-52-1]
257168932	ACV46691.1	conserved hypothetical protein [Halomicrobium mukohataei DSM 12286]
76558139	CAI49726.1	CRISPR-associated helicase Cas3 [Natronomonas pharaonis DSM 2160]
343784863	AEM58839.1	CRISPR-associated helicase Cas3 [Haloarcula hispanica ATCC 33960]
519065061	WP_020220936.1	hypothetical protein [Halarchaeum acidiphilum]
516847548	WP_018129105.1	hypothetical protein [Haladaptatus paucihalophilus]
493940704	WP_006884856.1	hypothetical protein [Halosimplex carlsbadense]
491677581	WP_005533723.1	CRISPR-associated helicase Cas3 [Haloarcula argentinensis]
445804138	EMA54401.1	hypothetical protein C450_06210 [Halococcus salifodinae DSM 8989]
445774122	EMA25144.1	CRISPR-associated helicase Cas3 [Haloarcula argentinensis DSM 12282]
445670557	ELZ23156.1	hypothetical protein C475_15944 [Halosimplex carlsbadense 2-9-1]
320549835	EFW91492.1	hypothetical protein ZOD2009_14031 [Haladaptatus paucihalophilus DX253]
543414357	GAD53619.1	CRISPR-associated helicase Cas3 [Halarchaeum acidiphilum MH1-52-1]
556695525	ESP86782.1	CRISPR-associated helicase Cas3 [Candidatus Halobonum tyrrellensis G22]
519065260	WP_020221135.1	hypothetical protein [Halarchaeum acidiphilum]
491183003	WP_005041365.1	hypothetical protein [Halococcus salifodinae]
529147355	WP_020936219.1	CRISPR-associated helicase Cas3 [Halorhabdus tiamatea]
648486246	WP_026177997.1	CRISPR-associated endonuclease Cas2 [Haladaptatus paucihalophilus]
557375041	WP_023396046.1	CRISPR-associated endonuclease Cas2 [Candidatus Halobonum tyrrellensis]
544640158	WP_021074699.1	CRISPR-associated endonuclease Cas2 [haloarchaeon 3A1_DGR]
543935980	WP_021029443.1	CRISPR-associated endonuclease Cas2 [Halorhabdus tiamatea]
519065263	WP_020221138.1	CRISPR-associated endonuclease Cas2 [Halarchaeum acidiphilum]
506269423	WP_015789198.1	CRISPR-associated endonuclease Cas2 [Halorhabdus utahensis]
506244011	WP_015763786.1	MULTISPECIES: CRISPR-associated endonuclease Cas2 [Halomicrobium]
503796718	WP_014030712.1	CRISPR-associated endonuclease Cas2 [Haloarcula hispanica]
499541489	WP_011222272.1	CRISPR-associated endonuclease Cas2 [Haloarcula marismortui]
495010967	WP_007736979.1	CRISPR-associated endonuclease Cas2 [Halococcus thailandensis]
491677576	WP_005533718.1	CRISPR-associated endonuclease Cas2 [Haloarcula argentinensis]
491106993	WP_004966759.1	CRISPR-associated endonuclease Cas2 [Haloarcula sinaiensis]
543414356	GAD53618.1	CRISPR-associated RecB family exonuclease Cas4b [Halarchaeum acidiphilum MH1-52-1]

76558142	CAI49729.1	CRISPR-associated endonuclease Cas2 [Natronomonas pharaonis DSM 2160]
76558141	CAI49728.1	CRISPR-associated endonuclease Cas1 [Natronomonas pharaonis DSM 2160]
76558140	CAI49727.1	CRISPR-associated exonuclease Cas4 [Natronomonas pharaonis DSM 2160]
76558135	CAI49722.1	CRISPR-associated endoribonuclease Cas6 [Natronomonas pharaonis DSM 2160]
528526427	CCQ33716.1	CRISPR-associated helicase, Cas3 [Halorhabdus tiamatea SARL4B]
556695521	ESP86778.1	CRISPR-associated endoribonuclease Cas6 [Candidatus Halobonum tyrrellensis G22]
495010977	WP_007736989.1	CRISPR-associated DNA-binding Csa3 [Halococcus thailandensis]
491182628	WP_005040990.1	CRISPR-associated DNA-binding Csa3 [Halococcus salifodinae]
445806447	EMA56576.1	CRISPR-associated DNA-binding Csa3 [Halococcus thailandensis JCM 13552]
445804444	EMA54700.1	CRISPR-associated DNA-binding Csa3 [Halococcus salifodinae DSM 8989]
543413948	GAD53847.1	CRISPR-associated helicase Cas3 [Halarchaeum acidiphilum MH1-52-1]
543413947	GAD53846.1	CRISPR-associated RecB family exonuclease Cas4a [Halarchaeum acidiphilum MH1-52-1]
495010971	WP_007736983.1	CRISPR-associated helicase [Halococcus thailandensis]
445806442	EMA56571.1	CRISPR-associated helicase [Halococcus thailandensis JCM 13552]
573487043	AHG05448.1	helicase (plasmid) [Halobacterium sp. DL1]
746356223	WP_039400889.1	helicase [haloarchaeon 3A1_DGR]
564124629	AHB67949.1	helicase (plasmid) [Haloarcula hispanica N601]
55229001	AAV44424.1	putative helicase (plasmid) [Haloarcula marismortui ATCC 43049]
257171182	ACV48941.1	putative helicase [Halomicrobium mukohataei DSM 12286]
256691290	ACV11627.1	putative helicase [Halorhabdus utahensis DSM 12940]
517069340	WP_018258158.1	helicase [Halomicrobium katesii]
503796721	WP_014030715.1	helicase [Haloarcula hispanica]
495257577	WP_007982332.1	helicase [Haladaptatus paucihalophilus]
506269426	WP_015789201.1	helicase [Halorhabdus utahensis]
506244008	WP_015763783.1	helicase [Halomicrobium mukohataei]
445756794	EMA08152.1	putative helicase [Haloarcula sinaiensis ATCC 33800]
320548911	EFW90578.1	putative helicase [Haladaptatus paucihalophilus DX253]
494345543	WP_007188930.1	helicase [Haloarcula californiae]
445806735	EMA56846.1	helicase [Halococcus thailandensis JCM 13552]
445767706	EMA18800.1	helicase [Haloarcula californiae ATCC 33799]

Table S14: The library of *cas* genes identified in Haloferacales genomes that were aligned against Newbler contigs assembled from the Cahuil/C34, combined Chula Vista, and combined Santa Pola/Isla Cristina metagenomes

GI	GB	Description
544612413	WP_021050764.1	CRISPR-associated protein [Haloquadratum walsbyi]
222454589	ACM58852.1	CRISPR-associated protein, Csh2 family (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454588	ACM58851.1	CRISPR-associated protein Cas5, Hmari subtype (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454585	ACM58848.1	CRISPR-associated protein Cas1 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454584	ACM58847.1	CRISPR-associated protein Cas2 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454586	ACM58849.1	CRISPR-associated protein Cas4 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454823	ACM59086.1	CRISPR-associated protein Cas2 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454822	ACM59085.1	CRISPR-associated protein Cas1 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454821	ACM59084.1	CRISPR-associated protein Cas4 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454819	ACM59082.1	CRISPR-associated protein Csc1 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454818	ACM59081.1	CRISPR-associated protein Csc2 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454817	ACM59080.1	CRISPR-associated protein Csc3 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454591	ACM58854.1	CRISPR-associated protein Cas6 (plasmid) [Halorubrum lacusprofundi ATCC 49239]
222454590	ACM58853.1	CRISPR-associated protein, Csh1 family (plasmid) [Halorubrum lacusprofundi ATCC 49239]
756974975	WP_042663236.1	CRISPR-associated protein Cas3 [Haloferax sp. ATB1]
388246497	AFK21440.1	CRISPR-associated Cas3 family protein (plasmid) [Haloferax mediterranei ATCC 33500]
495599996	WP_008324575.1	CRISPR-associated Cas3 family protein [Haloferax elongans]
445745777	ELZ97243.1	CRISPR-associated Cas3 family protein [Haloferax mediterranei ATCC 33500]
445742165	ELZ93661.1	CRISPR-associated Cas3 family protein [Haloferax mucosum ATCC BAA-1512]
445733001	ELZ84576.1	CRISPR-associated Cas3 family protein [Haloferax elongans ATCC BAA-1513]
631806802	AHZ24495.1	CRISPR-associated protein Cas6 (plasmid) [Haloferax mediterranei ATCC 33500]
631806801	AHZ24494.1	CRISPR-associated protein Csh1 (plasmid) [Haloferax mediterranei ATCC 33500]
631806800	AHZ24493.1	CRISPR-associated protein Csh2 (plasmid) [Haloferax mediterranei ATCC 33500]
631806799	AHZ24492.1	CRISPR-associated protein Cas5 (plasmid) [Haloferax mediterranei ATCC 33500]
631806796	AHZ24489.1	CRISPR-associated protein Cas1 (plasmid) [Haloferax mediterranei ATCC 33500]

631806795	AHZ24488.1	CRISPR-associated protein Cas2 (plasmid) [Haloferax mediterranei ATCC 33500]
291369447	ADE01675.1	CRISPR-associated protein Cas6 (plasmid) [Haloferax volcanii DS2]
495369681	WP_008094394.1	CRISPR-associated protein Cas6 [Haloferax prahovense]
504368112	WP_014555214.1	CRISPR-associated protein Cas3 [Haloquadratum walsbyi]
502800038	WP_013035014.1	CRISPR-associated protein Cas6 [Haloferax volcanii]
490145995	WP_004046325.1	CRISPR-associated Cas3 family protein [Halorubrum saccharovorum]
445717150	ELZ68871.1	CRISPR-associated protein Cas6 [Haloferax prahovense DSM 18310]
445690551	ELZ42761.1	CRISPR-associated Cas3 family protein [Halorubrum saccharovorum DSM 1137]
445577659	ELY32091.1	CRISPR-associated protein Cas6 [Haloferax volcanii DS2]
432197975	ELK54311.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax sp. BAB2207]
432197974	ELK54310.1	Csh2 family CRISPR-associated protein [Haloferax sp. BAB2207]
432197972	ELK54308.1	CRISPR-associated protein Cas6 [Haloferax sp. BAB2207]
631806797	AHZ24490.1	CRISPR-associated protein Cas4 (plasmid) [Haloferax mediterranei ATCC 33500]
504368108	WP_014555210.1	CRISPR-associated protein Cas6 [Haloquadratum walsbyi]
445819508	EMA69349.1	CRISPR-associated protein Cas2 [Halorubrum kocurii JCM 14978]
445819507	EMA69348.1	CRISPR-associated protein Cas1 [Halorubrum kocurii JCM 14978]
445819504	EMA69345.1	CRISPR-associated protein Cas5, Hmari subtype [Halorubrum kocurii JCM 14978]
445819503	EMA69344.1	CRISPR-associated protein, Csh2 family [Halorubrum kocurii JCM 14978]
445819501	EMA69342.1	CRISPR-associated protein Cas6 [Halorubrum kocurii JCM 14978]
445748297	ELZ99745.1	CRISPR-associated protein Cas2 [Haloferax denitrificans ATCC 35960]
445748296	ELZ99744.1	CRISPR-associated protein Cas1 [Haloferax denitrificans ATCC 35960]
445748293	ELZ99741.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax denitrificans ATCC 35960]
445748292	ELZ99740.1	CRISPR-associated protein, Csh2 family [Haloferax denitrificans ATCC 35960]
445748290	ELZ99738.1	CRISPR-associated protein Cas6 [Haloferax denitrificans ATCC 35960]
445745781	ELZ97247.1	CRISPR-associated Cas6 family protein [Haloferax mediterranei ATCC 33500]
445745779	ELZ97245.1	CRISPR-associated Csh2 family protein [Haloferax mediterranei ATCC 33500]
445745778	ELZ97244.1	CRISPR-associated Cas5h family protein [Haloferax mediterranei ATCC 33500]
445745775	ELZ97241.1	CRISPR-associated protein Cas1 [Haloferax mediterranei ATCC 33500]
445745774	ELZ97240.1	CRISPR-associated Cas2 family protein [Haloferax mediterranei ATCC 33500]
445744554	ELZ96029.1	CRISPR-associated protein Cas2 [Haloferax sulfurifontis ATCC BAA-897]

445744553	ELZ96028.1	CRISPR-associated Cas1 family protein [Haloferax sulfurifontis ATCC BAA-897]
445742970	ELZ94458.1	CRISPR-associated Cas1 family protein [Haloferax alexandrinus JCM 10717]
445742969	ELZ94457.1	CRISPR-associated protein Cas2 [Haloferax alexandrinus JCM 10717]
445742168	ELZ93664.1	CRISPR-associated Cas2 family protein [Haloferax mucosum ATCC BAA-1512]
445742167	ELZ93663.1	CRISPR-associated protein Cas1 [Haloferax mucosum ATCC BAA-1512]
445742163	ELZ93659.1	CRISPR-associated Csh2 family protein [Haloferax mucosum ATCC BAA-1512]
445742161	ELZ93657.1	CRISPR-associated Cas6 family protein [Haloferax mucosum ATCC BAA-1512]
445737820	ELZ89351.1	CRISPR-associated protein, Csh2 family [Haloferax alexandrinus JCM 10717]
445737819	ELZ89350.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax alexandrinus JCM 10717]
445737816	ELZ89347.1	CRISPR-associated protein Cas1 [Haloferax alexandrinus JCM 10717]
445737815	ELZ89346.1	CRISPR-associated protein Cas2 [Haloferax alexandrinus JCM 10717]
445737434	ELZ88968.1	CRISPR-associated protein Cas6 [Haloferax alexandrinus JCM 10717]
445733005	ELZ84580.1	CRISPR-associated Cas6 family protein [Haloferax elongans ATCC BAA-1513]
445733003	ELZ84578.1	CRISPR-associated Csh2 family protein [Haloferax elongans ATCC BAA-1513]
445732999	ELZ84574.1	CRISPR-associated protein Cas1 [Haloferax elongans ATCC BAA-1513]
445732998	ELZ84573.1	CRISPR-associated Cas2 family protein [Haloferax elongans ATCC BAA-1513]
445732888	ELZ84466.1	CRISPR-associated Cas5h family protein [Haloferax gibbonsii ATCC 33959]
445732886	ELZ84464.1	CRISPR-associated protein Cas1 [Haloferax gibbonsii ATCC 33959]
445732885	ELZ84463.1	CRISPR-associated protein Cas2 [Haloferax gibbonsii ATCC 33959]
445732882	ELZ84461.1	CRISPR-associated protein Cas6 [Haloferax gibbonsii ATCC 33959]
445724435	ELZ76067.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax lucentense DSM 14919]
445724434	ELZ76066.1	Csh2 family CRISPR-associated protein [Haloferax lucentense DSM 14919]
445724432	ELZ76064.1	CRISPR-associated protein Cas6 [Haloferax lucentense DSM 14919]
445720591	ELZ72264.1	CRISPR-associated protein Cas6 [Haloferax sp. ATCC BAA-644]
445720589	ELZ72262.1	CRISPR-associated protein, Csh2 family [Haloferax sp. ATCC BAA-644]
445720588	ELZ72261.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax sp. ATCC BAA-644]
445720585	ELZ72258.1	CRISPR-associated protein Cas1 [Haloferax sp. ATCC BAA-644]
445720584	ELZ72257.1	CRISPR-associated protein Cas2 [Haloferax sp. ATCC BAA-644]

445720377	ELZ72051.1	CRISPR-associated protein Cas6 [Haloferax lucentense DSM 14919]
445720188	ELZ71864.1	CRISPR-associated protein Cas2 [Haloferax lucentense DSM 14919]
445720187	ELZ71863.1	CRISPR-associated protein Cas1 [Haloferax lucentense DSM 14919]
445720184	ELZ71860.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax lucentense DSM 14919]
445720183	ELZ71859.1	CRISPR-associated protein, Csh2 family [Haloferax lucentense DSM 14919]
445717152	ELZ68873.1	CRISPR-associated Csh2 family protein [Haloferax prahovense DSM 18310]
445717148	ELZ68869.1	CRISPR-associated protein Cas1 [Haloferax prahovense DSM 18310]
445717147	ELZ68868.1	CRISPR-associated Cas2 family protein [Haloferax prahovense DSM 18310]
445712009	ELZ63794.1	CRISPR-associated protein Cas6 [Haloferax sp. ATCC BAA-644]
445712007	ELZ63792.1	Csh2 family CRISPR-associated protein [Haloferax sp. ATCC BAA-644]
445712006	ELZ63791.1	CRISPR-associated protein Cas5 [Haloferax sp. ATCC BAA-644]
445710854	ELZ62650.1	CRISPR-associated protein Cas5 [Haloferax sp. ATCC BAA-645]
445710853	ELZ62649.1	Csh2 family CRISPR-associated protein [Haloferax sp. ATCC BAA-645]
445710851	ELZ62647.1	CRISPR-associated protein Cas6 [Haloferax sp. ATCC BAA-645]
445708593	ELZ60432.1	CRISPR-associated protein Cas2 [Haloferax sp. ATCC BAA-645]
445708592	ELZ60431.1	CRISPR-associated protein Cas1 [Haloferax sp. ATCC BAA-645]
445708589	ELZ60428.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax sp. ATCC BAA-645]
445708588	ELZ60427.1	CRISPR-associated protein, Csh2 family [Haloferax sp. ATCC BAA-645]
445708586	ELZ60425.1	CRISPR-associated protein Cas6 [Haloferax sp. ATCC BAA-645]
445707690	ELZ59543.1	CRISPR-associated protein Cas2 [Haloferax sp. ATCC BAA-646]
445707689	ELZ59542.1	CRISPR-associated protein Cas1 [Haloferax sp. ATCC BAA-646]
445707686	ELZ59539.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax sp. ATCC BAA-646]
445707685	ELZ59538.1	CRISPR-associated protein, Csh2 family [Haloferax sp. ATCC BAA-646]
445707683	ELZ59536.1	CRISPR-associated protein Cas6 [Haloferax sp. ATCC BAA-646]
445705979	ELZ57866.1	CRISPR-associated protein Cas5 [Haloferax sp. ATCC BAA-646]
445705978	ELZ57865.1	Csh2 family CRISPR-associated protein [Haloferax sp. ATCC BAA-646]
445705976	ELZ57863.1	CRISPR-associated protein Cas6 [Haloferax sp. ATCC BAA-646]
445691207	ELZ43399.1	CRISPR-associated protein Cas5, Hmari subtype [Halorubrum coriense DSM 10284]
445691206	ELZ43398.1	CRISPR-associated protein, Csh2 family [Halorubrum coriense DSM 10284]
445691203	ELZ43395.1	CRISPR-associated protein Cas6 [Halorubrum coriense DSM 10284]
445690555	ELZ42765.1	CRISPR-associated protein Cas6 [Halorubrum saccharovorum DSM 1137]

445690553	ELZ42763.1	Csh2 family CRISPR-associated protein [Halorubrum saccharovorum DSM 1137]
445690549	ELZ42759.1	CRISPR-associated protein Cas1 [Halorubrum saccharovorum DSM 1137]
445690548	ELZ42758.1	CRISPR-associated protein Cas2 [Halorubrum saccharovorum DSM 1137]
445688495	ELZ40752.1	CRISPR-associated protein Cas6 [Halorubrum californiensis DSM 19288]
445688493	ELZ40750.1	CRISPR-associated protein, Csh2 family [Halorubrum californiensis DSM 19288]
445688492	ELZ40749.1	CRISPR-associated protein Cas5, Hmari subtype [Halorubrum californiensis DSM 19288]
445688489	ELZ40746.1	CRISPR-associated protein Cas1 [Halorubrum californiensis DSM 19288]
445688488	ELZ40745.1	CRISPR-associated protein Cas2 [Halorubrum californiensis DSM 19288]
445577667	ELY32099.1	CRISPR-associated protein Cas2 [Haloferax volcanii DS2]
445577666	ELY32098.1	CRISPR-associated protein Cas1 [Haloferax volcanii DS2]
445577662	ELY32094.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax volcanii DS2]
445577661	ELY32093.1	CRISPR-associated protein, Csh2 family [Haloferax volcanii DS2]
496122562	WP_008847069.1	CRISPR-associated protein Cas5 [Halorubrum kocurii]
495847873	WP_008572452.1	MULTISPECIES: CRISPR-associated protein Cas5 [Haloferax]
495720289	WP_008444868.1	CRISPR-associated protein Cas5 [Halorubrum californiense]
495369687	WP_008094400.1	CRISPR-associated protein Cas5 [Haloferax prahovense]
493057808	WP_006114886.1	CRISPR-associated protein Cas5 [Halorubrum coriense]
491114036	WP_004972493.1	CRISPR-associated protein Cas5 [Haloferax gibbonsii]
491112827	WP_004971286.1	CRISPR-associated protein Cas5 [Haloferax denitrificans]
490166027	WP_004064674.1	MULTISPECIES: CRISPR-associated protein Cas5 [Haloferax]
490163813	WP_004062466.1	MULTISPECIES: CRISPR-associated protein Cas5 [Haloferax]
490145996	WP_004046326.1	CRISPR-associated protein Cas5 [Halorubrum saccharovorum]
811258927	CQR53165.1	CRISPR-associated protein (Cas_Cas5) [Haloferax sp. Arc-Hr]
811258925	CQR53160.1	CRISPR-associated protein (cas_TM1802) [Haloferax sp. Arc-Hr]
339729023	CCC40224.1	CRISPR-associated protein Cas7 [Haloquadratum walsbyi C23]
339729022	CCC40223.1	CRISPR-associated protein Cas8b [Haloquadratum walsbyi C23]
339728196	CCC39329.1	CRISPR-associated protein Csc1 [Haloquadratum walsbyi C23]
339728195	CCC39328.1	CRISPR-associated protein Csc2 [Haloquadratum walsbyi C23]
339728194	CCC39327.1	CRISPR-associated protein Cas10d [Haloquadratum walsbyi C23]
339728192	CCC39325.1	CRISPR-associated DNA-binding protein Csm6 [Haloquadratum walsbyi C23]
756974974	WP_042663235.1	CRISPR-associated protein Cas5 [Haloferax sp. ATB1]
756974962	WP_042663224.1	CRISPR-associated protein Csh2 [Haloferax sp. ATB1]
756974958	WP_042663221.1	CRISPR-associated protein Cas6 [Haloferax sp. ATB1]
756973645	WP_042661972.1	CRISPR-associated protein Csh2 [Haloferax sp. ATB1]
756973643	WP_042661970.1	CRISPR-associated protein Cas6 [Haloferax sp. ATB1]
756973641	WP_042661968.1	CRISPR-associated protein Cas1 [Haloferax sp. ATB1]

388246500	AFK21443.1	CRISPR-associated Cas2 family protein (plasmid) [Haloferax mediterranei ATCC 33500]
388246499	AFK21442.1	CRISPR-associated Cas1 family protein (plasmid) [Haloferax mediterranei ATCC 33500]
388246496	AFK21439.1	CRISPR-associated Cas5h family protein (plasmid) [Haloferax mediterranei ATCC 33500]
388246495	AFK21438.1	CRISPR-associated Csh2 family protein (plasmid) [Haloferax mediterranei ATCC 33500]
388246494	AFK21437.1	CRISPR-associated Csh1 family protein (plasmid) [Haloferax mediterranei ATCC 33500]
388246493	AFK21436.1	CRISPR-associated Cas6 family protein (plasmid) [Haloferax mediterranei ATCC 33500]
291369979	ADE02207.1	CRISPR-associated protein Cas2 (plasmid) [Haloferax volcanii DS2]
291369715	ADE01943.1	CRISPR-associated protein Cas5, Hmari subtype (plasmid) [Haloferax volcanii DS2]
291369627	ADE01855.1	CRISPR-associated protein, Csh2 family (plasmid) [Haloferax volcanii DS2]
291369516	ADE01744.1	CRISPR-associated protein Cas1 (plasmid) [Haloferax volcanii DS2]
544612414	WP_021050765.1	CRISPR-associated protein Cas7/Csh2, subtype I-B/HMARI [Haloquadratum walsbyi]
541182738	ERG91248.1	CRISPR-associated protein Cas5, subtype I-B/HMARI [Haloquadratum walsbyi J07Hqw1]
541182737	ERG91247.1	CRISPR-associated protein Cas7/Csh2, subtype I-B/HMARI [Haloquadratum walsbyi J07Hqw1]
541182736	ERG91246.1	CRISPR-associated protein TM1802 (cas_TM1802) [Haloquadratum walsbyi J07Hqw1]
491114040	WP_004972497.1	CRISPR-associated protein Cas6 [Haloferax gibbonsii]
491112830	WP_004971289.1	CRISPR-associated protein Cas1 [Haloferax denitrificans]
491112826	WP_004971285.1	CRISPR-associated protein, Csh2 family [Haloferax denitrificans]
491112824	WP_004971283.1	CRISPR-associated protein, Csh1 family [Haloferax denitrificans]
491112822	WP_004971281.1	CRISPR-associated protein Cas6 [Haloferax denitrificans]
495852334	WP_008576913.1	MULTISPECIES: CRISPR-associated protein Cas5 [Haloferax]
495847879	WP_008572458.1	MULTISPECIES: CRISPR-associated protein Cas6 [Haloferax]
495847877	WP_008572456.1	MULTISPECIES: CRISPR-associated protein, Csh1 family [Haloferax]
495847875	WP_008572454.1	MULTISPECIES: CRISPR-associated protein, Csh2 family [Haloferax]
495847867	WP_008572446.1	MULTISPECIES: CRISPR-associated protein Cas1 [Haloferax]
493648623	WP_006600194.1	MULTISPECIES: CRISPR-associated protein Cas4 [Haloferax]
490166014	WP_004064661.1	MULTISPECIES: CRISPR-associated protein Cas6 [Haloferax]
490163810	WP_004062463.1	MULTISPECIES: CRISPR-associated protein Cas6 [Haloferax]
496122561	WP_008847068.1	CRISPR-associated protein, Csh2 family [Halorubrum kocurii]
496122560	WP_008847067.1	CRISPR-associated protein, Csh1 family [Halorubrum kocurii]
496122559	WP_008847066.1	CRISPR-associated protein Cas6 [Halorubrum kocurii]
495720296	WP_008444875.1	CRISPR-associated protein Cas6 [Halorubrum californiense]
495720294	WP_008444873.1	CRISPR-associated protein, Csh1 family [Halorubrum californiense]

495720292	WP_008444871.1	CRISPR-associated protein, Csh2 family [Halorubrum californiense]
495720281	WP_008444860.1	CRISPR-associated protein Cas1 [Halorubrum californiense]
495599998	WP_008324577.1	CRISPR-associated Csh2 family protein [Haloferax elongans]
495599994	WP_008324573.1	CRISPR-associated protein Cas1 [Haloferax elongans]
495596107	WP_008320686.1	CRISPR-associated protein Cas1 [Haloferax mucosum]
495596102	WP_008320681.1	CRISPR-associated Csh2 family protein [Haloferax mucosum]
495596101	WP_008320680.1	CRISPR-associated Csh1 family protein [Haloferax mucosum]
495369685	WP_008094398.1	CRISPR-associated Csh2 family protein [Haloferax prahovense]
495369683	WP_008094396.1	CRISPR-associated protein Csh1 [Haloferax prahovense]
494484860	WP_007274333.1	CRISPR-associated Cas1 family protein [Haloferax sulfurifontis]
494484857	WP_007274330.1	CRISPR-associated protein Csc2 [Haloferax sulfurifontis]
493649497	WP_006601038.1	CRISPR-associated protein, Csh1 family, partial [Haloferax alexandrinus]
493649496	WP_006601037.1	CRISPR-associated protein, Csh2 family [Haloferax alexandrinus]
493649495	WP_006601036.1	CRISPR-associated protein Cas1 [Haloferax alexandrinus]
493648626	WP_006600197.1	CRISPR-associated protein Csc2 [Haloferax alexandrinus]
493648622	WP_006600193.1	CRISPR-associated Cas1 family protein [Haloferax alexandrinus]
493057806	WP_006114885.1	CRISPR-associated protein, Csh2 family [Halorubrum coriense]
504545715	WP_014732817.1	CRISPR-associated protein Cas6 [Haloferax mediterranei]
502800142	WP_013035118.1	CRISPR-associated protein Cas5 [Haloferax volcanii]
490162157	WP_004060812.1	CRISPR-associated protein Csh1 [Haloferax mediterranei]
490162156	WP_004060811.1	CRISPR-associated protein Csh2 [Haloferax mediterranei]
490162155	WP_004060810.1	CRISPR-associated protein Cas5 [Haloferax mediterranei]
490142785	WP_004043127.1	CRISPR-associated protein Csh2 [Haloferax volcanii]
501864844	WP_012660271.1	CRISPR-associated protein Cas6 [Halorubrum lacusprofundi]
501864461	WP_012660066.1	CRISPR-associated protein Csh2 [Halorubrum lacusprofundi]
501864457	WP_012660065.1	CRISPR-associated protein Cas5 [Halorubrum lacusprofundi]
504368806	WP_014555908.1	CRISPR-associated protein Cas1 [Haloquadratum walsbyi]
504368802	WP_014555904.1	CRISPR-associated protein Csh2 [Haloquadratum walsbyi]
504368801	WP_014555903.1	CRISPR-associated protein Csh1 [Haloquadratum walsbyi]
504368114	WP_014555216.1	CRISPR-associated protein Cas1 [Haloquadratum walsbyi]
504368113	WP_014555215.1	CRISPR-associated protein Cas4 [Haloquadratum walsbyi]
504368111	WP_014555213.1	CRISPR-associated protein Csc1 [Haloquadratum walsbyi]
504368110	WP_014555212.1	CRISPR-associated protein Csc2 [Haloquadratum walsbyi]
504368109	WP_014555211.1	CRISPR-associated protein Csc3 [Haloquadratum walsbyi]
502800067	WP_013035043.1	CRISPR-associated protein Cas1 [Haloferax volcanii]
501864850	WP_012660277.1	CRISPR-associated protein Cas1 [Halorubrum lacusprofundi]
501864846	WP_012660273.1	CRISPR-associated protein Csc2 [Halorubrum lacusprofundi]
501864845	WP_012660272.1	CRISPR-associated protein Csc3 [Halorubrum lacusprofundi]
501864444	WP_012660062.1	CRISPR-associated protein Cas1 [Halorubrum lacusprofundi]
490166030	WP_004064677.1	CRISPR-associated protein Cas1 [Haloferax lucentense]

490166026	WP_004064673.1	CRISPR-associated protein, Csh2 family [Haloferax lucentense]
490166015	WP_004064662.1	CRISPR-associated protein, Csh1 family, partial [Haloferax lucentense]
490162152	WP_004060807.1	CRISPR-associated protein Cas1 [Haloferax mediterranei]
490145999	WP_004046329.1	CRISPR-associated protein Cas6 [Halorubrum saccharovorum]
490145998	WP_004046328.1	CRISPR-associated protein, Csh1 family [Halorubrum saccharovorum]
490145993	WP_004046323.1	CRISPR-associated protein Cas1 [Halorubrum saccharovorum]
445819502	EMA69343.1	CRISPR-associated protein, Csh1 family [Halorubrum kocurii JCM 14978]
445748291	ELZ99739.1	CRISPR-associated protein, Csh1 family [Haloferax denitrificans ATCC 35960]
445745780	ELZ97246.1	CRISPR-associated Csh1 family protein [Haloferax mediterranei ATCC 33500]
445744552	ELZ96027.1	CRISPR-associated protein Cas4 [Haloferax sulfurifontis ATCC BAA-897]
445744549	ELZ96024.1	CRISPR-associated protein Csc2 [Haloferax sulfurifontis ATCC BAA-897]
445742974	ELZ94462.1	CRISPR-associated protein Csc2 [Haloferax alexandrinus JCM 10717]
445742971	ELZ94459.1	CRISPR-associated protein Cas4 [Haloferax alexandrinus JCM 10717]
445742164	ELZ93660.1	CRISPR-associated Cas5h family protein [Haloferax mucosum ATCC BAA-1512]
445742162	ELZ93658.1	CRISPR-associated Csh1 family protein [Haloferax mucosum ATCC BAA-1512]
445737433	ELZ88967.1	CRISPR-associated protein, Csh1 family, partial [Haloferax alexandrinus JCM 10717]
445733004	ELZ84579.1	CRISPR-associated Csh1 family protein [Haloferax elongans ATCC BAA-1513]
445733002	ELZ84577.1	CRISPR-associated Cas5h family protein [Haloferax elongans ATCC BAA-1513]
445732889	ELZ84467.1	CRISPR-associated protein, Csh2 family [Haloferax gibbonsii ATCC 33959]
445720590	ELZ72263.1	CRISPR-associated protein, Csh1 family [Haloferax sp. ATCC BAA-644]
445720378	ELZ72052.1	CRISPR-associated protein, Csh1 family, partial [Haloferax lucentense DSM 14919]
445717153	ELZ68874.1	CRISPR-associated protein Cas5, Hmari subtype [Haloferax prahovense DSM 18310]
445717151	ELZ68872.1	CRISPR-associated protein Csh1 [Haloferax prahovense DSM 18310]
445708587	ELZ60426.1	CRISPR-associated protein, Csh1 family [Haloferax sp. ATCC BAA-645]
445707684	ELZ59537.1	CRISPR-associated protein, Csh1 family [Haloferax sp. ATCC BAA-646]
445690554	ELZ42764.1	CRISPR-associated protein, Csh1 family [Halorubrum saccharovorum DSM 1137]
445690552	ELZ42762.1	CRISPR-associated protein Cas5, Hmari subtype [Halorubrum saccharovorum DSM 1137]
445688494	ELZ40751.1	CRISPR-associated protein, Csh1 family [Halorubrum californiensis DSM 19288]

491114037	WP_004972494.1	MULTISPECIES: CRISPR-associated protein, Csh2 family [Haloferax]
910013508	WP_049968221.1	CRISPR-associated protein Cas4 [Haloferax alexandrinus]
910013507	WP_049968220.1	CRISPR-associated protein Cas5 [Haloferax alexandrinus]
910013503	WP_049968216.1	CRISPR-associated protein Cas6 [Haloferax alexandrinus]
909691407	WP_049934655.1	CRISPR-associated protein Cas2 [Haloplanus natans]
909691405	WP_049934653.1	CRISPR-associated protein Cas1 [Haloplanus natans]
909691404	WP_049934652.1	CRISPR-associated protein Cas4 [Haloplanus natans]
909691402	WP_049934650.1	CRISPR-associated protein Cas5 [Haloplanus natans]
909691400	WP_049934648.1	CRISPR-associated protein Csh2 [Haloplanus natans]
909691398	WP_049934646.1	CRISPR-associated protein Csh1 [Haloplanus natans]
909691396	WP_049934644.1	CRISPR-associated protein Cas6 [Haloplanus natans]
909690778	WP_049934026.1	CRISPR-associated protein Cas4 [Halorubrum lacusprofundi]
909690768	WP_049934016.1	CRISPR-associated protein Csh1 [Halorubrum lacusprofundi]
909690668	WP_049933916.1	CRISPR-associated protein Cas6 [Halorubrum lacusprofundi]
909689303	WP_049932648.1	CRISPR-associated protein Cas3 [Halorubrum ezzemoulense]
909689301	WP_049932646.1	CRISPR-associated protein Csh2 [Halorubrum ezzemoulense]
909689161	WP_049932506.1	CRISPR-associated protein Cas5 [Halorubrum ezzemoulense]
909689158	WP_049932503.1	CRISPR-associated protein Cas6 [Halorubrum ezzemoulense]
909678745	WP_049922099.1	CRISPR-associated protein Csa3 [Haloferax sp. BAB2207]
909674905	WP_049918259.1	CRISPR-associated protein Cas4 [Haloferax denitrificans]
909674211	WP_049917565.1	CRISPR-associated protein Cas4 [Haloferax mediterranei]
909674210	WP_049917564.1	CRISPR-associated protein Cas2 [Haloferax mediterranei]
909672422	WP_049915776.1	CRISPR-associated protein Cas2 [Haloferax mucosum]
909672417	WP_049915771.1	CRISPR-associated protein Cas5 [Haloferax mucosum]
909672415	WP_049915769.1	CRISPR-associated protein Cas6 [Haloferax mucosum]
909670166	WP_049913520.1	CRISPR-associated protein Cas4 [Haloferax prahovense]
909670152	WP_049913506.1	CRISPR-associated protein Cas1 [Haloferax prahovense]
909668482	WP_049911836.1	CRISPR-associated protein Cas1 [Halorubrum kocurii]
909665569	WP_049908923.1	CRISPR-associated protein Cas2 [Halorubrum saccharovorum]
909664318	WP_049907672.1	CRISPR-associated protein Csh1 [Haloferax elongans]
909664317	WP_049907671.1	CRISPR-associated protein Cas2 [Haloferax elongans]
909664289	WP_049907643.1	CRISPR-associated protein Cas6 [Haloferax elongans]
909664288	WP_049907642.1	CRISPR-associated protein Cas5 [Haloferax elongans]
909661582	WP_049904936.1	CRISPR-associated protein Cas1 [Haloferax gibbonsii]
909661490	WP_049904844.1	CRISPR-associated protein Csa3 [Haloferax gibbonsii]
909653668	WP_049897081.1	MULTISPECIES: CRISPR-associated protein Cas6 [Haloferax]
909648857	WP_049892270.1	CRISPR-associated protein Cas5 [Haloquadratum walsbyi]
909648358	WP_049891771.1	CRISPR-associated protein Cas6 [Haloquadratum walsbyi]
756973642	WP_042661969.1	CRISPR-associated protein Cas4 [Haloferax sp. ATB1]
388246498	AFK21441.1	CRISPR-associated Cas4 family protein (plasmid) [Haloferax mediterranei ATCC 33500]

291369910	ADE02138.1	CRISPR-associated protein Cas4 (plasmid) [Haloferax volcanii DS2]
495847869	WP_008572448.1	MULTISPECIES: CRISPR-associated protein Cas4 [Haloferax]
490166029	WP_004064676.1	MULTISPECIES: CRISPR-associated protein Cas4 [Haloferax]
496122564	WP_008847071.1	CRISPR-associated protein Cas4 [Halorubrum kocurii]
495599995	WP_008324574.1	CRISPR-associated Cas4 family protein [Haloferax elongans]
495596105	WP_008320684.1	CRISPR-associated Cas4 family protein [Haloferax mucosum]
490162153	WP_004060808.1	CRISPR-associated protein Cas4 [Haloferax mediterranei]
504368805	WP_014555907.1	CRISPR-associated protein Cas4 [Haloquadratum walsbyi]
502800211	WP_013035187.1	CRISPR-associated protein Cas4 [Haloferax volcanii]
501864445	WP_012660063.1	CRISPR-associated protein Cas4 [Halorubrum lacusprofundi]
490145994	WP_004046324.1	CRISPR-associated protein Cas4 [Halorubrum saccharovorum]
445819506	EMA69347.1	CRISPR-associated protein Cas4 [Halorubrum kocurii JCM 14978]
445748295	ELZ99743.1	CRISPR-associated protein Cas4 [Haloferax denitrificans ATCC 35960]
445745776	ELZ97242.1	CRISPR-associated Cas4 family protein [Haloferax mediterranei ATCC 33500]
445742166	ELZ93662.1	CRISPR-associated Cas4 family protein [Haloferax mucosum ATCC BAA-1512]
445737817	ELZ89348.1	CRISPR-associated protein Cas4 [Haloferax alexandrinus JCM 10717]
445733000	ELZ84575.1	CRISPR-associated Cas4 family protein [Haloferax elongans ATCC BAA-1513]
445720586	ELZ72259.1	CRISPR-associated protein Cas4 [Haloferax sp. ATCC BAA-644]
445720186	ELZ71862.1	CRISPR-associated protein Cas4 [Haloferax lucentense DSM 14919]
445720182	ELZ71858.1	CRISPR-associated protein, Csh1 family, partial [Haloferax lucentense DSM 14919]
445717149	ELZ68870.1	CRISPR-associated protein Cas4 [Haloferax prahovense DSM 18310]
445708591	ELZ60430.1	CRISPR-associated protein Cas4 [Haloferax sp. ATCC BAA-645]
445707688	ELZ59541.1	CRISPR-associated protein Cas4 [Haloferax sp. ATCC BAA-646]
445690550	ELZ42760.1	CRISPR-associated protein Cas4 [Halorubrum saccharovorum DSM 1137]
445688490	ELZ40747.1	CRISPR-associated protein Cas4 [Halorubrum californiensis DSM 19288]
445577664	ELY32096.1	CRISPR-associated protein Cas4 [Haloferax volcanii DS2]
811258924	CQR53158.1	CRISPR associated protein Cas6 [Haloferax sp. Arc-Hr]
765680567	D4GQN5.1	RecName: Full=CRISPR-associated protein Cas8b
765680566	D4GQN6.1	RecName: Full=CRISPR-associated protein Cas7
765680563	D4GQN7.1	RecName: Full=CRISPR-associated protein Cas5
222454820	ACM59083.1	CRISPR-associated helicase, Cyano-type (plasmid) [Halorubrum lacusprofundi ATCC 49239]
490736648	WP_004598973.1	hypothetical protein [Halorubrum distributum]
339729025	CCC40226.1	CRISPR-associated helicase Cas3 [Haloquadratum walsbyi C23]
291370021	ADE02249.1	CRISPR-associated helicase Cas3 (plasmid) [Haloferax volcanii DS2]
291369495	ADE01723.1	conserved hypothetical protein (plasmid) [Haloferax volcanii DS2]

491112828	WP_004971287.1	CRISPR-associated helicase Cas3 [Haloferax denitrificans]
495852331	WP_008576910.1	MULTISPECIES: CRISPR-associated helicase Cas3 [Haloferax]
495847871	WP_008572450.1	MULTISPECIES: CRISPR-associated helicase Cas3 [Haloferax]
490166028	WP_004064675.1	MULTISPECIES: CRISPR-associated helicase Cas3 [Haloferax]
504368804	WP_014555906.1	CRISPR-associated helicase Cas3 [Haloquadratum walsbyi]
490142784	WP_004043126.1	hypothetical protein [Haloferax volcanii]
445818153	EMA68020.1	hypothetical protein C462_15150 [Halorubrum arcis JCM 13916]
445807291	EMA57376.1	hypothetical protein C470_13958 [Halorubrum litoreum JCM 13561]
445748294	ELZ99742.1	CRISPR-associated helicase Cas3 [Haloferax denitrificans ATCC 35960]
445737818	ELZ89349.1	CRISPR-associated helicase Cas3 [Haloferax alexandrinus JCM 10717]
445720587	ELZ72260.1	CRISPR-associated helicase Cas3 [Haloferax sp. ATCC BAA-644]
445720185	ELZ71861.1	CRISPR-associated helicase Cas3 [Haloferax lucentense DSM 14919]
445712005	ELZ63790.1	CRISPR-associated helicase Cas3 [Haloferax sp. ATCC BAA-644]
445710855	ELZ62651.1	CRISPR-associated helicase Cas3 [Haloferax sp. ATCC BAA-645]
445708590	ELZ60429.1	CRISPR-associated helicase Cas3 [Haloferax sp. ATCC BAA-645]
445707687	ELZ59540.1	CRISPR-associated helicase Cas3 [Haloferax sp. ATCC BAA-646]
445705980	ELZ57867.1	CRISPR-associated helicase Cas3 [Haloferax sp. ATCC BAA-646]
445704182	ELZ56100.1	hypothetical protein C466_04619 [Halorubrum distributum JCM 10118]
445693246	ELZ45405.1	hypothetical protein C465_14140 [Halorubrum distributum JCM 9100]
445577663	ELY32095.1	CRISPR-associated helicase Cas3 [Haloferax volcanii DS2]
445577660	ELY32092.1	hypothetical protein C498_09706 [Haloferax volcanii DS2]
494484859	WP_007274332.1	CRISPR-associated helicase Cyano-type [Haloferax sulfurifontis]
493648624	WP_006600195.1	CRISPR-associated helicase Cyano-type [Haloferax alexandrinus]
339728197	CCC39330.1	CRISPR-associated helicase Cas3 [Haloquadratum walsbyi C23]
541184072	ERG92582.1	DEAD/DEAH box helicase domain protein, partial [Haloquadratum walsbyi J07HQW1]
445744551	ELZ96026.1	CRISPR-associated helicase, Cyano-type [Haloferax sulfurifontis ATCC BAA-897]
445742972	ELZ94460.1	CRISPR-associated helicase, Cyano-type [Haloferax alexandrinus JCM 10717]
339728193	CCC39326.1	CRISPR-associated endoribonuclease Cas6 [Haloquadratum walsbyi C23]
756974963	WP_042663225.1	CRISPR-associated endonuclease Cas2 [Haloferax sp. ATB1]
756973640	WP_042661967.1	CRISPR-associated endonuclease Cas2 [Haloferax sp. ATB1]
504368807	WP_014555909.1	CRISPR-associated endonuclease Cas2 [Haloquadratum walsbyi]
504368115	WP_014555217.1	CRISPR-associated endonuclease Cas2 [Haloquadratum walsbyi]
501864443	WP_012660061.1	CRISPR-associated endonuclease Cas2 [Halorubrum lacusprofundi]
496122566	WP_008847073.1	CRISPR-associated endonuclease Cas2 [Halorubrum kocurii]
495720279	WP_008444858.1	CRISPR-associated endonuclease Cas2 [Halorubrum californiense]

495369676	WP_008094389.1	CRISPR-associated endonuclease Cas2 [Haloferax prahovense]
491114032	WP_004972490.1	CRISPR-associated endonuclease Cas2 [Haloferax gibbonsii]
491112831	WP_004971290.1	CRISPR-associated endonuclease Cas2 [Haloferax denitrificans]
490166032	WP_004064679.1	MULTISPECIES: CRISPR-associated endonuclease Cas2 [Haloferax]
490162151	WP_004060806.1	CRISPR-associated endonuclease Cas2 [Haloferax mediterranei]
490142791	WP_004043133.1	CRISPR-associated endonuclease Cas2 [Haloferax volcanii]
339729028	CCC40229.1	CRISPR-associated endonuclease Cas2 [Haloquadratum walsbyi C23]
339729027	CCC40228.1	CRISPR-associated endonuclease Cas1 [Haloquadratum walsbyi C23]
339729024	CCC40225.1	CRISPR-associated endoribonuclease Cas5, Hmari subtype [Haloquadratum walsbyi C23]
339728200	CCC39333.1	CRISPR-associated endonuclease Cas2 [Haloquadratum walsbyi C23]
339728199	CCC39332.1	CRISPR-associated endonuclease Cas1 [Haloquadratum walsbyi C23]
339728198	CCC39331.1	CRISPR-associated exonuclease Cas4 [Haloquadratum walsbyi C23]
544612416	WP_021050767.1	CRISPR-associated endonuclease Cas3-HD [Haloquadratum walsbyi]
541182739	ERG91249.1	CRISPR-associated endonuclease Cas3-HD [Haloquadratum walsbyi J07HQW1]
490733198	WP_004595537.1	CRISPR-associated DNA-binding Csa3 [Halorubrum distributum]
504368107	WP_014555209.1	CRISPR-associated DNA-binding Csa3 [Haloquadratum walsbyi]
445707019	ELZ58886.1	CRISPR-associated DNA-binding Csa3 [Halorubrum distributum JCM 10118]
445700343	ELZ52346.1	CRISPR-associated DNA-binding Csa3 [Halorubrum distributum JCM 9100]
432192782	ELK49604.1	CRISPR-associated DNA-binding Csa3 [Haloferax sp. BAB2207]
811258931	CQR53172.1	CRISPR-associated endoribonuclease Cas2 [Haloferax sp. Arc-Hr]
811258930	CQR53171.1	CRISPR-associated endonuclease Cas1 [Haloferax sp. Arc-Hr]
339729026	CCC40227.1	CRISPR-associated exonuclease Cas4 [Haloquadratum walsbyi C23]
445735348	ELZ86898.1	CRISPR-associated DNA-binding Csa3 [Haloferax gibbonsii ATCC 33959]
765680561	D4GQN8.1	RecName: Full=CRISPR-associated nuclease/helicase Cas3
765680565	D4GQN4.1	RecName: Full=CRISPR-associated endoribonuclease Cas6
765680564	I3RB26.1	RecName: Full=CRISPR-associated endoribonuclease Cas6
765680560	D4GQP1.1	RecName: Full=CRISPR-associated endoribonuclease Cas2
765680562	D4GQN9.1	RecName: Full=CRISPR-associated exonuclease Cas4
765680559	D4GQP0.1	RecName: Full=CRISPR-associated endonuclease Cas1
631806798	AHZ24491.1	helicase (plasmid) [Haloferax mediterranei ATCC 33500]
222454587	ACM58850.1	putative helicase (plasmid) [Halorubrum lacusprofundi ATCC 49239]
493057810	WP_006114887.1	helicase [Halorubrum coriense]
445691208	ELZ43400.1	helicase [Halorubrum coriense DSM 10284]
496122563	WP_008847070.1	helicase [Halorubrum kocurii]

495720287	WP_008444866.1	helicase [Halorubrum californiense]
495369689	WP_008094402.1	helicase [Haloferax prahovense]
490162154	WP_004060809.1	helicase [Haloferax mediterranei]
490142787	WP_004043129.1	helicase [Haloferax volcanii]
445819505	EMA69346.1	helicase [Halorubrum kocurii JCM 14978]
445717154	ELZ68875.1	helicase [Haloferax prahovense DSM 18310]
445688491	ELZ40748.1	helicase [Halorubrum californiense DSM 19288]
544601826	WP_021040304.1	superfamily II helicase archaea-specific [Halonotus sp. J07HN4]
541198251	ERH06414.1	superfamily II helicase, archaea-specific [Halonotus sp. J07HN4]
490163814	WP_004062467.1	helicase [Haloferax lucentense]
445724436	ELZ76068.1	helicase [Haloferax lucentense DSM 14919]

Table S15: The library of *cas* genes identified in Natrionalbales genomes that were aligned against Newbler contigs assembled from the Cahuil/C34, combined Chula Vista, and combined Santa Pola/Isla Cristina metagenomes

GI	GB	Description
445783277	EMA34110.1	CRISPR-associated protein [Halobiforma nitratireducens JCM 10879]
445652087	ELZ04988.1	CRISPR-associated protein [Natrialba chahannaoensis JCM 10990]
445638626	ELY91753.1	CRISPR-associated protein [Natrialba hulunbeirensis JCM 10989]
445619053	ELY72600.1	CRISPR-associated protein [Natronobacterium gregoryi SP2]
633918735	KDD77846.1	CRISPR-associated protein, partial [Haloterrigena sp. H13]
493724240	WP_006673618.1	CRISPR-associated protein [Halobiforma nitratireducens]
493703361	WP_006653095.1	CRISPR-associated protein [Natrialba hulunbeirensis]
493157409	WP_006165795.1	CRISPR-associated protein [Natrialba chahannaoensis]
909685432	WP_049928780.1	CRISPR-associated protein [Halopiger sp. IH3]
429137566	AFZ74577.1	CRISPR-associated protein Cas5, subtype I-B/HMARI [Natronobacterium gregoryi SP2]
429137565	AFZ74576.1	CRISPR-associated protein Cas7/Csh2, subtype I-B/HMARI [Natronobacterium gregoryi SP2]
429137564	AFZ74575.1	CRISPR-associated protein Cas8b/Csh1, subtype I-B/HMARI [Natronobacterium gregoryi SP2]
429137568	AFZ74579.1	CRISPR-associated protein Cas4 [Natronobacterium gregoryi SP2]
445643588	ELY96633.1	CRISPR-associated protein Cas6 [Natrialba chahannaoensis JCM 10990]
493161517	WP_006168575.1	CRISPR-associated protein Cas6 [Natrialba chahannaoensis]
445783278	EMA34111.1	CRISPR-associated protein Cas5, Hmari subtype [Halobiforma nitratireducens JCM 10879]
445783275	EMA34108.1	CRISPR-associated protein Cas6 [Halobiforma nitratireducens JCM 10879]
445779236	EMA30171.1	CRISPR-associated protein Cas6 [Halobiforma lacisalsi AJ5]
445779234	EMA30169.1	CRISPR-associated protein, Csh2 family [Halobiforma lacisalsi AJ5]

445779233	EMA30168.1	CRISPR-associated protein Cas5, Hmari subtype [Halobiforma lacisalsi AJ5]
445779230	EMA30165.1	CRISPR-associated protein Cas1 [Halobiforma lacisalsi AJ5]
445779229	EMA30164.1	CRISPR-associated protein Cas2 [Halobiforma lacisalsi AJ5]
445664215	ELZ16934.1	CRISPR-associated protein Cas6 [Haloterrigena thermotolerans DSM 11522]
445664213	ELZ16932.1	Csh2 family CRISPR-associated protein [Haloterrigena thermotolerans DSM 11522]
445664212	ELZ16931.1	CRISPR-associated protein Cas5 [Haloterrigena thermotolerans DSM 11522]
445664209	ELZ16928.1	CRISPR-associated protein Cas1 [Haloterrigena thermotolerans DSM 11522]
445664208	ELZ16927.1	CRISPR-associated protein Cas2 [Haloterrigena thermotolerans DSM 11522]
445652092	ELZ04993.1	CRISPR-associated protein Cas2 [Natrialba chahannaoensis JCM 10990]
445652091	ELZ04992.1	CRISPR-associated protein Cas1 [Natrialba chahannaoensis JCM 10990]
445652088	ELZ04989.1	CRISPR-associated protein Cas5, Hmari subtype [Natrialba chahannaoensis JCM 10990]
445652085	ELZ04986.1	CRISPR-associated protein Cas6 [Natrialba chahannaoensis JCM 10990]
445644086	ELY97118.1	CRISPR-associated protein Csc2 [Natrialba asiatica DSM 12278]
445644082	ELY97114.1	CRISPR-associated Cas1 family protein [Natrialba asiatica DSM 12278]
445644081	ELY97113.1	CRISPR-associated protein Cas2 [Natrialba asiatica DSM 12278]
445643586	ELY96631.1	CRISPR-associated protein, Csh2 family [Natrialba chahannaoensis JCM 10990]
445643585	ELY96630.1	CRISPR-associated protein Cas5, Hmari subtype [Natrialba chahannaoensis JCM 10990]
445643029	ELY96085.1	CRISPR-associated protein Cas2 [Natrialba hulunbeirensis JCM 10989]
445643028	ELY96084.1	CRISPR-associated Cas1 family protein [Natrialba hulunbeirensis JCM 10989]
445638628	ELY91755.1	CRISPR-associated protein Cas6 [Natrialba hulunbeirensis JCM 10989]
445638625	ELY91752.1	CRISPR-associated protein Cas5, Hmari subtype [Natrialba hulunbeirensis JCM 10989]
445638622	ELY91749.1	CRISPR-associated protein Cas1 [Natrialba hulunbeirensis JCM 10989]
445638621	ELY91748.1	CRISPR-associated protein Cas2 [Natrialba hulunbeirensis JCM 10989]
445630635	ELY83896.1	CRISPR-associated protein Cas2 [Natriinema gari JCM 14663]
445630634	ELY83895.1	CRISPR-associated protein Cas1 [Natriinema gari JCM 14663]
445630631	ELY83892.1	CRISPR-associated protein Cas5 [Natriinema gari JCM 14663]
445630630	ELY83891.1	CRISPR-associated protein, Csh2 family [Natriinema gari JCM 14663]
445630628	ELY83889.1	CRISPR-associated protein Cas6 [Natriinema gari JCM 14663]
445628549	ELY81855.1	CRISPR-associated protein Cas2 [Natriinema pallidum DSM 3751]
445628548	ELY81854.1	CRISPR-associated protein Cas1 [Natriinema pallidum DSM 3751]

445620369	ELY73870.1	CRISPR-associated protein Cas6, partial [Natrinema pallidum DSM 3751]
445620327	ELY73832.1	CRISPR-associated protein Cas5 [Natrinema pallidum DSM 3751]
445619055	ELY72602.1	CRISPR-associated protein Cas6 [Natronobacterium gregoryi SP2]
445619052	ELY72599.1	CRISPR-associated protein Cas5, Hmari subtype [Natronobacterium gregoryi SP2]
445619049	ELY72596.1	CRISPR-associated protein Cas1 [Natronobacterium gregoryi SP2]
445619048	ELY72595.1	CRISPR-associated protein Cas2 [Natronobacterium gregoryi SP2]
445602000	ELY55980.1	CRISPR-associated regulatory protein, DevR family [Natronococcus amylolyticus DSM 10524]
445601999	ELY55979.1	CRISPR-associated protein Cas5 [Natronococcus amylolyticus DSM 10524]
445587219	ELY41483.1	CRISPR-associated protein Cas6 [Natronorubrum sulfidifaciens JCM 14089]
445587217	ELY41481.1	CRISPR-associated protein, Csh2 family [Natronorubrum sulfidifaciens JCM 14089]
445587216	ELY41480.1	CRISPR-associated protein Cas5, Hmari subtype [Natronorubrum sulfidifaciens JCM 14089]
445587213	ELY41477.1	CRISPR-associated protein Cas1 [Natronorubrum sulfidifaciens JCM 14089]
445587212	ELY41476.1	CRISPR-associated protein Cas2 [Natronorubrum sulfidifaciens JCM 14089]
504676159	WP_014863261.1	MULTISPECIES: CRISPR-associated protein Cas1 [Natrinema]
495727693	WP_008452272.1	MULTISPECIES: CRISPR-associated protein, Csh2 family [Natrinema]
493698384	WP_006648202.1	MULTISPECIES: CRISPR-associated protein Cas1 [Haloterrigena]
635282939	KDE59361.1	CRISPR-associated protein Cas5 [Halostagnicola sp. A56]
635282938	KDE59360.1	CRISPR-associated protein Csh2 [Halostagnicola sp. A56]
635282937	KDE59359.1	CRISPR-associated protein Csh1 [Halostagnicola sp. A56]
635279523	KDE56713.1	CRISPR-associated protein Cas2 [Halostagnicola sp. A56]
495440267	WP_008164961.1	CRISPR-associated protein Cas5 [Natronorubrum sulfidifaciens]
493724241	WP_006673619.1	CRISPR-associated protein Cas5 [Halobiforma nitratreducens]
493703360	WP_006653094.1	CRISPR-associated protein Cas5 [Natrialba hulunbeirensis]
493161511	WP_006168572.1	CRISPR-associated protein Cas5 [Natrialba chahannaoensis]
493157411	WP_006165797.1	CRISPR-associated protein Cas5 [Natrialba chahannaoensis]
397681801	AFO56178.1	CRISPR-associated protein Cas2 [Natrinema sp. J7-2]
397681800	AFO56177.1	CRISPR-associated protein Cas1 [Natrinema sp. J7-2]
397681797	AFO56174.1	CRISPR-associated protein Cas5, Hmari subtype [Natrinema sp. J7-2]
397681796	AFO56173.1	CRISPR-associated protein, Csh2 family [Natrinema sp. J7-2]
397681795	AFO56172.1	CRISPR-associated protein, Csh1 family [Natrinema sp. J7-2]
397681794	AFO56171.1	CRISPR-associated protein Cas6 [Natrinema sp. J7-2]
495727701	WP_008452280.1	CRISPR-associated protein Cas1 [Natrinema gari]
495727694	WP_008452273.1	CRISPR-associated protein Cas5 [Natrinema gari]
495727689	WP_008452268.1	CRISPR-associated protein, Csh1 family [Natrinema gari]

495727687	WP_008452266.1	CRISPR-associated protein Cas6 [Natrinema gari]
495701826	WP_008426405.1	CRISPR-associated regulatory protein, DevR family [Natronococcus jeotgali]
495697739	WP_008422318.1	CRISPR-associated regulatory protein, DevR family [Natronococcus jeotgali]
495440272	WP_008164966.1	CRISPR-associated protein Cas6 [Natronorubrum sulfidifaciens]
495440270	WP_008164964.1	CRISPR-associated protein, Csh1 family [Natronorubrum sulfidifaciens]
495440269	WP_008164963.1	CRISPR-associated protein, Csh2 family [Natronorubrum sulfidifaciens]
495440262	WP_008164956.1	CRISPR-associated protein Cas1 [Natronorubrum sulfidifaciens]
493724239	WP_006673617.1	CRISPR-associated protein, Csh1 family [Halobiforma nitratireducens]
493703363	WP_006653097.1	CRISPR-associated protein Cas6 [Natrialba hulunbeirensis]
493703362	WP_006653096.1	CRISPR-associated protein, Csh1 family [Natrialba hulunbeirensis]
493703358	WP_006653092.1	CRISPR-associated protein Cas1 [Natrialba hulunbeirensis]
493701604	WP_006651368.1	CRISPR-associated protein Cas4 [Natrialba hulunbeirensis]
493701601	WP_006651365.1	CRISPR-associated protein Csc2 [Natrialba hulunbeirensis]
493698389	WP_006648207.1	CRISPR-associated protein, Csh1 family [Haloterrigena thermotolerans]
493698387	WP_006648205.1	CRISPR-associated protein Cas5 [Haloterrigena thermotolerans]
493196772	WP_006186929.1	CRISPR-associated protein Cas5 [Natrinema pallidum]
493196770	WP_006186927.1	CRISPR-associated protein, Csh2 family, partial [Natrinema pallidum]
493196767	WP_006186924.1	CRISPR-associated protein Cas6, partial [Natrinema pallidum]
493161515	WP_006168574.1	CRISPR-associated protein, Csh1 family [Natrialba chahannaoensis]
493161513	WP_006168573.1	CRISPR-associated protein, Csh2 family [Natrialba chahannaoensis]
493157416	WP_006165802.1	CRISPR-associated protein Cas1 [Natrialba chahannaoensis]
493157408	WP_006165794.1	CRISPR-associated protein, Csh1 family [Natrialba chahannaoensis]
493157406	WP_006165792.1	CRISPR-associated protein Cas6 [Natrialba chahannaoensis]
493051977	WP_006111345.1	CRISPR-associated protein Csc2 [Natrialba asiatica]
493051973	WP_006111341.1	CRISPR-associated protein Cas4 [Natrialba asiatica]
493051972	WP_006111340.1	CRISPR-associated Cas1 family protein [Natrialba asiatica]
491743546	WP_005576679.1	CRISPR-associated protein Cas6 [Natronobacterium gregoryi]
491711242	WP_005557132.1	CRISPR-associated regulatory protein, DevR family [Natronococcus amylolyticus]
491711241	WP_005557131.1	CRISPR-associated protein Cas5 [Natronococcus amylolyticus]
504676154	WP_014863256.1	CRISPR-associated protein Csh1 [Natrinema sp. J7-2]
494240648	WP_007143042.1	CRISPR-associated protein Csh1 [Halobiforma lacisalsi]
494240646	WP_007143041.1	CRISPR-associated protein Csh2 [Halobiforma lacisalsi]
491743544	WP_005576677.1	CRISPR-associated protein Csh1 [Natronobacterium gregoryi]
491743542	WP_005576675.1	CRISPR-associated protein Csh2 [Natronobacterium gregoryi]
491743541	WP_005576674.1	CRISPR-associated protein Cas5 [Natronobacterium gregoryi]
491743533	WP_005576666.1	CRISPR-associated protein Cas1 [Natronobacterium gregoryi]

445783276	EMA34109.1	CRISPR-associated protein, Csh1 family [Halobiforma nitratireducens JCM 10879]
445664214	ELZ16933.1	CRISPR-associated protein, Csh1 family [Haloterrigena thermotolerans DSM 11522]
445652086	ELZ04987.1	CRISPR-associated protein, Csh1 family [Natrialba chahannaoensis JCM 10990]
445644083	ELY97115.1	CRISPR-associated protein Cas4 [Natrialba asiatica DSM 12278]
445643587	ELY96632.1	CRISPR-associated protein, Csh1 family [Natrialba chahannaoensis JCM 10990]
445643027	ELY96083.1	CRISPR-associated protein Cas4 [Natrialba hulunbeirensis JCM 10989]
445643024	ELY96080.1	CRISPR-associated protein Csc2 [Natrialba hulunbeirensis JCM 10989]
445638627	ELY91754.1	CRISPR-associated protein, Csh1 family [Natrialba hulunbeirensis JCM 10989]
445630629	ELY83890.1	CRISPR-associated protein, Csh1 family [Natrinema gari JCM 14663]
445620326	ELY73831.1	CRISPR-associated protein, Csh2 family, partial [Natrinema pallidum DSM 3751]
445619054	ELY72601.1	CRISPR-associated protein, Csh1 family [Natronobacterium gregoryi SP2]
445608223	ELY62080.1	CRISPR-associated regulatory protein, DevR family [Natronococcus jeotgali DSM 18795]
445601996	ELY55976.1	CRISPR-associated protein Cas4 [Natronococcus amylolyticus DSM 10524]
445598724	ELY52778.1	CRISPR-associated regulatory protein, DevR family [Natronococcus jeotgali DSM 18795]
445587218	ELY41482.1	CRISPR-associated protein, Csh1 family [Natronorubrum sulfidifaciens JCM 14089]
910247548	WP_050052211.1	CRISPR-associated protein Cas2 [Halostagnicola sp. A56]
910246009	WP_050050672.1	CRISPR-associated protein Cas4 [Halostagnicola sp. A56]
910246007	WP_050050670.1	CRISPR-associated protein Cas5 [Halostagnicola sp. A56]
910246006	WP_050050669.1	CRISPR-associated protein Csh2 [Halostagnicola sp. A56]
910246005	WP_050050668.1	CRISPR-associated protein Csh1 [Halostagnicola sp. A56]
504676158	WP_014863260.1	MULTISPECIES: CRISPR-associated protein Cas4 [Natrinema]
910075582	WP_049992249.1	CRISPR-associated protein Cas6 [Halopiger salifodinae]
910075581	WP_049992248.1	CRISPR-associated protein Csh1 [Halopiger salifodinae]
910075580	WP_049992247.1	CRISPR-associated protein Csh2 [Halopiger salifodinae]
910012337	WP_049967071.1	CRISPR-associated protein Csh1 [Haloterrigena jeotgali]
910012336	WP_049967070.1	CRISPR-associated protein Csh2 [Haloterrigena jeotgali]
910012335	WP_049967069.1	CRISPR-associated protein Cas5 [Haloterrigena jeotgali]
910012333	WP_049967067.1	CRISPR-associated protein Cas4 [Haloterrigena jeotgali]
909685435	WP_049928783.1	CRISPR-associated protein Cas2 [Halopiger sp. IIH3]
909685434	WP_049928782.1	CRISPR-associated protein Cas1 [Halopiger sp. IIH3]
909685433	WP_049928781.1	CRISPR-associated protein Cas4 [Halopiger sp. IIH3]
909685430	WP_049928778.1	CRISPR-associated protein Csc2 [Halopiger sp. IIH3]
909685429	WP_049928777.1	CRISPR-associated protein Cas6 [Halopiger sp. IIH3]
909676875	WP_049920229.1	CRISPR-associated protein Cas6 [Halobiforma nitratireducens]

909668993	WP_049912347.1	CRISPR-associated protein Cas1 [Natrialba hulunbeirensis]
909668974	WP_049912328.1	CRISPR-associated protein Cas6 [Natrialba hulunbeirensis]
909664492	WP_049907846.1	CRISPR-associated protein Cas1 [Natrinema pallidum]
909662084	WP_049905438.1	MULTISPECIES: CRISPR-associated protein Cas6 [Haloterrigena]
909656866	WP_049900279.1	CRISPR-associated protein Csh2 [Natrinema sp. J7-1]
909656863	WP_049900276.1	CRISPR-associated protein Csh1 [Natrinema sp. J7-1]
909644697	WP_049888110.1	MULTISPECIES: CRISPR-associated protein Cas5 [Natrinema]
909644696	WP_049888109.1	MULTISPECIES: CRISPR-associated protein Cas6 [Natrinema]
635282941	KDE59363.1	CRISPR-associated protein Cas4 [Halostagnicola sp. A56]
397681799	AFO56176.1	CRISPR-associated protein Cas4 [Natrinema sp. J7-2]
493157414	WP_006165800.1	MULTISPECIES: CRISPR-associated protein Cas4 [Natrialba]
495727698	WP_008452277.1	CRISPR-associated protein Cas4 [Natrinema gari]
495440264	WP_008164958.1	CRISPR-associated protein Cas4 [Natronorubrum sulfidifaciens]
493698385	WP_006648203.1	CRISPR-associated protein Cas4 [Haloterrigena thermotolerans]
493196997	WP_006187151.1	CRISPR-associated protein Cas4 [Natrinema pallidum]
491743536	WP_005576669.1	CRISPR-associated protein Cas4 [Natronobacterium gregoryi]
445779231	EMA30166.1	CRISPR-associated protein Cas4 [Halobiforma lacisalsi AJ5]
445664210	ELZ16929.1	CRISPR-associated protein Cas4 [Haloterrigena thermotolerans DSM 11522]
445652090	ELZ04991.1	CRISPR-associated protein Cas4 [Natrialba chahannaoensis JCM 10990]
445638623	ELY91750.1	CRISPR-associated protein Cas4 [Natrialba hulunbeirensis JCM 10989]
445630633	ELY83894.1	CRISPR-associated protein Cas4 [Natrinema gari JCM 14663]
445619409	ELY72948.1	CRISPR-associated protein Cas4 [Natrinema pallidum DSM 3751]
445619050	ELY72597.1	CRISPR-associated protein Cas4 [Natronobacterium gregoryi SP2]
445587214	ELY41478.1	CRISPR-associated protein Cas4 [Natronorubrum sulfidifaciens JCM 14089]
429137569	AFZ74580.1	CRISPR-associated endonuclease Cas1 [Natronobacterium gregoryi SP2]
429137567	AFZ74578.1	CRISPR-associated endonuclease Cas3-HD [Natronobacterium gregoryi SP2]
429137570	AFZ74581.1	CRISPR-associated endonuclease Cas2 [Natronobacterium gregoryi SP2]
429137563	AFZ74574.1	CRISPR-associated endonuclease Cas6 [Natronobacterium gregoryi SP2]
493182304	WP_006180277.1	hypothetical protein [Natrinema pellirubrum]
433304520	AGB30332.1	hypothetical protein Natpe_0401 [Natrinema pellirubrum DSM 15624]
495440266	WP_008164960.1	CRISPR-associated helicase Cas3 [Natronorubrum sulfidifaciens]
445783279	EMA34112.1	CRISPR-associated helicase Cas3, partial [Halobiforma nitratireducens JCM 10879]
445665560	ELZ18236.1	hypothetical protein C478_01220 [Haloterrigena thermotolerans DSM 11522]
445652089	ELZ04990.1	CRISPR-associated helicase Cas3 [Natrialba chahannaoensis JCM 10990]
445648836	ELZ01784.1	hypothetical protein C481_09697 [Natrialba asiatica DSM 12278]

445625936	ELY79288.1	hypothetical protein C488_04817 [Natrinema pellirubrum DSM 15624]
445619051	ELY72598.1	CRISPR-associated helicase Cas3 [Natronobacterium gregoryi SP2]
445587215	ELY41479.1	CRISPR-associated helicase Cas3 [Natronorubrum sulfidifaciens JCM 14089]
493051974	WP_006111342.1	CRISPR-associated helicase Cyano-type [Natrialba asiatica]
493701603	WP_006651367.1	CRISPR-associated helicase [Natrialba hulunbeirensis]
445644084	ELY97116.1	CRISPR-associated helicase, Cyano-type [Natrialba asiatica DSM 12278]
445643026	ELY96082.1	CRISPR-associated helicase [Natrialba hulunbeirensis JCM 10989]
445615998	ELY69634.1	DEAD/DEAH box helicase domain-containing protein [Natronobacterium gregoryi SP2]
445611132	ELY64892.1	DEAD/DEAH box helicase domain protein [Natrinema versiforme JCM 10478]
445601994	ELY55974.1	CRISPR-associated endoribonuclease Cas2 [Natronococcus amylolyticus DSM 10524]
495727703	WP_008452282.1	MULTISPECIES: CRISPR-associated endonuclease Cas2 [Natrinema]
493698383	WP_006648201.1	MULTISPECIES: CRISPR-associated endonuclease Cas2 [Haloterrigena]
495440261	WP_008164955.1	CRISPR-associated endonuclease Cas2 [Natronorubrum sulfidifaciens]
493701606	WP_006651370.1	CRISPR-associated endonuclease Cas2 [Natrialba hulunbeirensis]
493191532	WP_006184181.1	CRISPR-associated endonuclease Cas2 [Natrinema pallidum]
493157417	WP_006165803.1	MULTISPECIES: CRISPR-associated endonuclease Cas2 [Natrialba]
493051970	WP_006111338.1	CRISPR-associated endonuclease Cas2 [Natrialba asiatica]
491743530	WP_005576663.1	CRISPR-associated endonuclease Cas2 [Natronobacterium gregoryi]
445642052	ELY95123.1	CRISPR-associated DNA-binding Csa3 [Natrialba hulunbeirensis JCM 10989]
445638624	ELY91751.1	CRISPR-associated helicase Cas3 [Natrialba hulunbeirensis JCM 10989]
505046736	WP_015233838.1	helicase [Natronobacterium gregoryi]
493477301	WP_006432253.1	DEAD/DEAH box helicase [Natrinema versiforme]
429135865	AFZ72876.1	DNA/RNA helicase, superfamily II [Natronobacterium gregoryi SP2]
491745336	WP_005578469.1	DEAD/DEAH box helicase [Natronobacterium gregoryi]
445779232	EMA30167.1	putative helicase [Halobiforma lacisalsi AJ5]