



Genome Variation in the Model Halophilic Bacterium *Salinibacter ruber*

Pedro González-Torres^{1,2} and Toni Gabaldón^{2,3,4*}†

¹ Department of Physiology, Genetics and Microbiology, University of Alicante, Alicante, Spain, ² Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Spain, ³ Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Spain, ⁴ Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

OPEN ACCESS

Edited by:

Masahiro Ito,
Toyo University, Japan

Reviewed by:

Francisco Rodríguez-Valera,
Universidad Miguel Hernández
de Elche, Spain
Kian Mau Goh,
University of Technology Malaysia,
Malaysia

*Correspondence:

Toni Gabaldón
tgabaldon@crg.es

†Senior author

Specialty section:

This article was submitted to
Extreme Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 09 March 2018

Accepted: 18 June 2018

Published: 19 July 2018

Citation:

González-Torres P and Gabaldón T
(2018) Genome Variation in the Model
Halophilic Bacterium
Salinibacter ruber.
Front. Microbiol. 9:1499.
doi: 10.3389/fmicb.2018.01499

The halophilic bacterium *Salinibacter ruber* is an abundant and ecologically important member of halophilic communities worldwide. Given its broad distribution and high intraspecific genetic diversity, *S. ruber* is considered one of the main models for ecological and evolutionary studies of bacterial adaptation to hypersaline environments. However, current insights on the genomic diversity of this species is limited to the comparison of the genomes of two co-isolated strains. Here, we present a comparative genomic analysis of eight *S. ruber* strains isolated at two different time points in each of two different Mediterranean solar salterns. Our results show an open pangenome with contrasting evolutionary patterns in the core and accessory genomes. We found that the core genome is shaped by extensive homologous recombination (HR), which results in limited sequence variation within population clusters. In contrast, the accessory genome is modulated by horizontal gene transfer (HGT), with genomic islands and plasmids acting as gateways to the rest of the genome. In addition, both types of genetic exchange are modulated by restriction and modification (RM) or CRISPR-Cas systems. Finally, genes differentially impacted by such processes reveal functional processes potentially relevant for environmental interactions and adaptation to extremophilic conditions. Altogether, our results support scenarios that conciliate “Neutral” and “Constant Diversity” models of bacterial evolution.

Keywords: *Salinibacter ruber*, homologous recombination (HR), core genomes, intraspecific diversity, comparative genomics

INTRODUCTION

The genomic and genetic diversification processes that act on prokaryotic species have been extensively studied (Thomas and Nielsen, 2005; Cohan, 2006; Fraser et al., 2007; Vos and Didelot, 2009; Didelot and Maiden, 2010; Polz, 2013). However, most such studies have historically focused on pathogenic bacteria (Hiller et al., 2010; Morelli et al., 2010; Reeves et al., 2011). In recent years, both whole genome comparisons of isolated strains and metagenomics analyses of environmental samples have opened the way to assess genome variation in free-living species (Mira et al., 2010). The metagenomics approach includes sequences from different individuals of a species within a population, which is used to determine genomic diversity (Coleman et al., 2006; Tully et al., 2015) or the variation of structure and content of genomes across populations, as in the case of *Prochlorococcus* sp. (Luo and Konstantinidis, 2011), marine Crenarchaea

(Konstantinidis and DeLong, 2008) and hypersaline environments (Podell et al., 2013). Some of these latter studies have revealed that the genetic diversity of microbial species is often organized in the form of cohesive clusters, each composed of strains with genetic similarity above a given threshold, which are differentiated from other clusters by much larger genetic distances. This cluster structure has been suggested to result from the action of homologous recombination (HR), which would constrain levels of genomic sequence divergence within a population cluster. Such structure is in agreement with the “Neutral Model” (Fraser et al., 2007), which underscores the role of HR as a neutral, passive mechanism that suffices to drive patterns of genetic divergence and diversity. Under this model HR would constrain sequence divergence within a recombining population while increasing the number of transitory unique genetic entities (referred to as the metaclonal scenario).

On the other hand, genome comparisons of multiple strains from the same species, although still limited by the number of available genomes, have allowed exploring the genomic structural diversity and the emerging temporal and biogeographic patterns thereof (Caro-Quintero et al., 2011; Cadillo-Quiroz et al., 2012; Shapiro et al., 2012; Tully et al., 2015). Although scarce, some of these studies have suggested that the environmental genetic pool, involving the flexible genome, is accessible through mechanisms of horizontal gene transfer (HGT) and site-specific recombination. In contrast, the core genome tends to evolve through mechanisms of vertical transmission and HR (Polz, 2013; López-Pérez and Rodríguez-Valera, 2016). In prokaryotes, a considerable fraction of the accessory genome is typically located within genomic islands (GIs), plasmids, and indels (i.e., insertions and deletions). Based on its architecture and evolution, GIs have been recently classified into one of two types: (i) flexible genomic islands (fGI), when they appear in equivalent genomic locations across strains of the same genus or species and (ii) HGT-GIs, when they are strain or clade specific, typically contain viral and transposable elements, and are flanked by tRNAs (López-Pérez and Rodríguez-Valera, 2016). Additionally, fGI were further divided into two categories: additive fGIs, that act as site directed recombination hotspots for the integration of gene cassettes that vary in number and nature, and replacement fGIs, encoded within different gene clusters in each strain, but coding for similar functional structures, generally related to cell envelopes (Rodríguez-Valera et al., 2016).

During the last decade, hypersaline environments have emerged as important models to study prokaryotic ecology, evolution, and adaptive strategies. These environments, defined by salt concentrations above that of sea water, account for 50% of continental waters (Shiklomanov, 1998; Villamor et al., 2017). In addition to these physical–chemical traits, hypersaline environments show particular biotic compositions, including a relatively low number but microdiverse halophilic prokaryotic species, defined as microorganisms whose optimal growth is set over salt concentration of 50 g/l (0.85 M NaCl), or higher and that tolerate 100 g/l salt (1.7 M) (Oren and Mana, 2003; Oren, 2008). In addition, hypersaline environments host the highest concentrations of virus-like particles among aquatic environments (Santos et al., 2012). Therefore they constitute

a privileged scenario to study virus–host interactions and adaptations. The “Kill the Winner” model (Thingstad and Lignell, 1997; Thingstad, 2000) is an adaptation of Lotka–Volterra model to understand population fluctuations derived from virus–bacteria interactions in aquatic environments. Complementary to this, the effect of virus–host interaction on the genomic diversity of bacterial populations is the focus of the “Constant Diversity” model (Rodríguez-Valera et al., 2009), which considers the existence of metaclonal populations whose diversity is maintained at equilibrium by means of phage predation.

Among the microbes inhabiting hypersaline environments, the Bacteroidetes *Salinibacter ruber* offers a particularly excellent model for microdiversity studies (Oren, 2013). This species is present in most hypersaline waters worldwide, including salter pond crystallizers, forming part of the core bank, usually as the most abundant bacterial species (Oren, 2013). Furthermore, *S. ruber* has a high intraspecific genomic and functional diversity, at both transcriptomic and metabolomic levels (Rosselló-Mora et al., 2008; González-Torres et al., 2015), even when co-occurring strains are considered (Pasić et al., 2009; Peña et al., 2010; Antón et al., 2013; González-Torres et al., 2015). This makes *S. ruber* an excellent model for microevolutionary studies, particularly with regards to its adaptation to an extreme environment (Oren, 2008, 2013).

Here we set out to broaden our insights into the genomic diversity in *S. ruber*, by comparing eight fully sequenced genomes. Our results define a highly variable accessory genome and an extremely conserved core genome, and highlight the differential impacts that HGT and HR processes have in these two fractions of the genome. Finally, we explore relationships between dynamic regions in the core and accessory genome and adaptation to environmental conditions, including phage predation. Our genomic comparisons, and earlier genomic and experimental studies, underscore the potential value of *S. ruber* as a model for evolutionary genomics studies of bacterial adaptation to hypersaline environments.

MATERIALS AND METHODS

Sample Collection and DNA Extraction

We used in this study eight strains previously isolated from Santa Pola and Campos saltern ponds (**Supplementary Table S1**). Details about the isolation and origin of the *S. ruber* M8, M31, M1, SP38, P13, P18, SP73, and RM158 strains used in this study are provided in **Supplementary Table S1** and in a previous study (Antón et al., 2013). Pre-inocula from individual colonies were set up into 10 ml tubes with 1 ml SW 25% supplemented with 2 g/l yeast extract and grown for at least 10 days at 37°C in an orbital shaker (170 rpm). Finally 500 ml 25% salt medium supplemented with 0.2% yeast extract was inoculated at 5%. Cultures were grown at 37°C in an orbital shaker (170 rpm) to exponential mid-point ($DO_{600} = 0,5$). Screening and selection of *Salinibacter isolates* were performed as previously described (Antón et al., 2000; Peña et al., 2005). Template DNA was obtained by using the DNeasy Blood and Tissue kit (QIAGEN, Netherlands), according to the manufacturer's

recommendations. The amount and quality of the extracted DNA was checked by spectrophotometry (Nanodrop ND-1000) 1% agarose (Seakem LE FMC Bioproducts) gel and a BioAnalyzer instrument (Agilent) prior to sequencing.

Sequencing, Assembly and Genome Annotation

Six *S. ruber* genomes (P13, P18, SP38, SP73, RM158, and M1) were assembled *de novo* in this study. Libraries were performed with an Illumina Sample Prep kit according to manufacturer's recommendations. Sequencing was performed using Illumina HiSeq 2000 (100 pb paired-end read) at the Center for Genomic Regulation (CRG) core facility, Barcelona Biomedical Research Park (PRBB). The generated FASTQ files were first processed to remove poor-quality reads as follows. Reads with a PHRED quality score of <10 were trimmed at the first base and, subsequently, read pairs with any read of length <31 bases were discarded. The remaining reads were assembled using SOAPdenovo 2.04v (Luo et al., 2012), IDBA-UD 1.1.0v (Peng et al., 2012) and JR-Assembler 1.0.4v (Chu et al., 2013) assemblers with varying parameters. For SOAPdenovo, optimal *k*-mer size was set between 41 and 81 bp. For IDBA-UD optimal *mink* ranged from 40 to 60, *min_count* from 12 to 20, and *min_support* from 6 to 20. Gaps were filled with GapCloser from SOAPdenovo2 2.04v (Luo et al., 2012), and the best resulting assemblies (based on N50 and N90 statistics) were combined. Resulting scaffolds were mapped and redirected using progressiveMauve from MAUVE 2.3.1v (Darling et al., 2010) using M31 and M8 *S. ruber* genomes as references (Peña et al., 2010). Formed gaps were closed connecting the scaffolds by using Geneious Read Mapper implemented in Geneious 7.1.7v (Kearse et al., 2012). First, the extremes of resulting scaffolds were used as a reference to drive the assembly toward the gaps. As previously described (Martínez-García et al., 2014), a single file with the paired-end reads was set and used as a query for the mapping driving assembly of the gaps using stringent parameters (no allowed gap between two matched reads, 99% of minimum overlap identity, 1% maximum mismatches per read). A total of 25–100 mapping and extension iterations were carried out until reaching a sequence overlap between the extended extremes of two consecutive scaffolds (i.e., closing the gap). Manual inspection of the overlapping extremes and uniqueness correspondence of the connections were checked by Nucleotide Blast (Altschul et al., 1990). The identity and completion of plasmid sequences was assessed by checking their circularity, the presence of low copy replicative origin (ParA/RepB), and by confirming their lower GC content and higher coverage as compared to the genome average. Additionally plasmidSPAdes pipeline (Antipov et al., 2016) from SPAdes 3.11.1v assembler (Bankevich et al., 2012) confirmed that these contigs conformed plasmidic elements. Strains P13 and P18 were used as negative controls as they lacked plasmids. The raw sequencing data have been submitted to the European Nucleotide Archive (ENA) and Short Read Archive (SRA) and are available under the project accession identifiers PRJEB25455 and SRP151445, respectively. The genome sequences have been deposited at NCBI

database under the bioproject accession number PRJNA474853. Biosample accessions are indicated in **Supplementary Table S5**.

De novo gene, CRISPR-Cas, and RM systems prediction and annotation were performed using the Integrated Microbial Genomes (IMG) system (Joint Genome Institute) (Markowitz et al., 2012), selecting the algorithm implemented for isolated genomes (IMG/ER). All genomes were annotated with a common platform, including M8 and M31, to reduce possible bias, as it is recommended in comparative genomic studies (Kimes et al., 2014). Functional terms from Clusters of Orthologous Groups (COG) (Tatusov et al., 2000) and Gene Ontology (GO) (Ashburner et al., 2000) databases were retrieved using the JGI platform and the Blast2GO software (B2G) 3.1v (Conesa and Götz, 2008; Götz et al., 2008), respectively, (**Supplementary Tables S2, S3**). In the latter, hits with an *E*-value lower than 1×10^{-20} and amino acid sequence identity higher than 55% were considered (Ogata et al., 1999). A further characterization of CRISPR-Cas systems and its associated spacers and proteins was performed using CRISPRfinder (Grissa et al., 2007) and Nucleotide Blast (Altschul et al., 1990) against the nr nucleotide collection and specific metavirome and metagenomes from several halophilic environments (**Supplementary Table S4**), including San Diego saltern ponds (EEUU) (Rodríguez-Brito et al., 2010), Tuz Lake (Turkey) (Boujelben et al., 2012), Santa Pola saltern pond CR30 (Spain) (Santos et al., 2010; Ghai et al., 2012; Fernandez et al., 2013), Tyrrell Lake (Australia) (Emerson et al., 2013).

Genome Comparison Analysis

Genomes of the six strains sequenced in this study and two previously available genomes (Peña et al., 2010) were compared. First, genomes were aligned using the progressive Mauve function from Mauve 2.3.1v (default parameters) (Darling et al., 2010). The output consisted of the alignment showing the Locally Collinear Blocks and the main rearrangements, the sequences contained in the aligned blocks in eXtended Multi-FastA (xmfa format), and the positional orthologs therein (*ort files). The output file was parsed with a Python script which selected the core genome (genes shared between all the strains) and accessory genomes (genes missing in some strains) (Mira et al., 2010). The genome sequences within each species were compared in a pairwise manner using the Nucmer function from the MUMmer package v3.0 (Kurtz et al., 2004). The resulting coordinate files (coords files) were parsed with a python script to calculate the percentage of identity for the aligned sequences. For this, fragments that overlapped over 10% of their length were joined, and those contained within a larger fragment were filtered out. An average nucleotide identity value (ANiB) was obtained for each comparison as the average of each aligned fragment weighted by its fragment length.

Detection and Functional Characterization of Recombination Events

RDP4 v4.15 (Martin et al., 2010) was used to detect and characterize recombination events. Genomic alignments in

*XMFA format obtained with MAUVE 2.3.1v (Darling et al., 2010) were used as input. Recombination events were predicted in a two step procedure. In a first exploratory phase, four different methods were run: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), MaxCHI (Smith, 1992), Chimera (Posada and Crandall, 2001). In a second phase, the program re-scanned more thoroughly every detected event with seven methods: BootScan (Salminen et al., 1995), SiScan (Gibbs et al., 2000), RDP, GENECONV, MaxCHI, Chimera and 3Seq (Boni et al., 2007). Only recombination signals considered by at least three of the seven methods were considered, using a cutoff value $p < 0.001$ in all cases. For RDP, window size was set to 90 nucleotides and for MaxCHI the variable sites per window was set to 210. For the remaining methods, default parameters were applied. The predicted recombination events were manually curated, and their breakpoints were inferred with MaxCHI method (considered the most accurate breakpoint detection method among the five non-parametric methods in RDP3) (Martin et al., 2010). From these data, the following recombination variables were inferred: recombined fraction of the genome, number of recombination events per strain, and distribution of sizes of recombination events (i.e., length of the recombined region). Nucleotide sequences of the predicted recombined fragments were retrieved from the genome FASTA files using the positional coordinates provided by RDP4 4.15v (Martin et al., 2010). For these sequences, *de novo* gene prediction and annotation was performed as indicated above.

Evolutionary and Phylogenomics Analysis

We reconstructed the clonal genealogy of the analyzed strains using ClonalFrame v1.2 (Didelot and Falush, 2007), a Bayesian inference method that, uses the genome alignment for the eight strains as input. Only aligned regions containing all genomes and longer than 500 bp were used. Three independent ClonalFrame runs, each consisting of 40,000 iterations were run, and the first 20,000 iterations for each run were discarded as burn-in. The convergence of the tree runs was checked, making sure that they produced consistent estimates of the clonal genealogy, and the global parameters r/m (where r = rate of recombination and m = rate of mutation), ρ/θ (σ/θ) (Vos and Didelot, 2009) (where σ and θ are the rates of occurrence of recombination and mutation, respectively) with 95% credibility (Vos and Didelot, 2009). The ratio σ/θ is a measure of the frequency at which recombination occurs relative to mutation, whereas r/m correspond to the ratio of rates at which nucleotides become substituted as a result of recombination or mutation, thus measuring the relative impact of HR on the observed population genetic diversification.

The amino acid sequences for the positional orthologous genes were retrieved from the GenBank files and pairwise coding sequence alignments between selected strains were performed with MUSCLE 3.8v (Edgar, 2004). The resulting alignments were reverse-translated to codon-based nucleotide alignments using trimAL 1.3v and the corresponding coding sequences (Capella-Gutiérrez et al., 2009). dN/dS values were obtained

using the CodeML function, (parwise mode with model 1 NS sites 0 parameters), from PaML package 4.4v (Yang, 2007). A low ratio (dN/dS <1) indicates purifying selection, whereas a high ratio (dN/dS >1) is suggestive of diversifying selection. Finally, pairwise rRNA 16S ANI values were calculated from pairwise alignments using SILVA Incremental Aligner (SINA) 1.2.11v (Pruesse et al., 2012).

Statistical Analysis

Tests for functional enrichment of genes contained in recombination fragments and GIs vs. the genomic background were performed using the Fisher Exact Test (FT) and the annotated COG and GO terms. Fisher's test were performed for each of the functional categories in each species, in the case of COG, by applying a FDR (False Discovery Rate) correction and considering a p -value < 0.05 as threshold for over or under-representations, compared to the whole genome as implemented in Gossip from Blast2Go (B2G) 3.1v (Conesa and Götz, 2008).

RESULTS AND DISCUSSION

Genome Sequences

In order to gain further insights into *S. ruber* genome diversity we analyzed the complete genomes of eight different strains, of which six (*S. ruber* M1, SP38, SP73, P13, P18, and RM158) have been sequenced and assembled for this study (**Supplementary Table S5**), and the other two (*S. ruber* M8 and M31), including the reference strain for the species, were obtained from previous studies (Mongodin et al., 2005; Peña et al., 2010). The set includes strains isolated from two mediterranean saltern ponds in Spain: Campos de Mallorca (RM158, M1, M31, and M8) and Bras del Port crystallizer CR30 Santa Pola (P13, P18, SP38 y SP73) (**Supplementary Table S1**). These two locations are situated 320 Km apart from each other and their ecological parameters and microbial communities are well-characterized from previous studies (Ventosa et al., 2014). The strains correspond to serial isolations of the two sites over the course of 10 years. Our assembly strategy produced a single chromosome for each of the strains, and for 18 plasmids present in the various strains (**Supplementary Figure S1** and **Supplementary Table S6**). We next reconstructed whole genome alignments of the eight strains. All strains shared a high level of sequence identity over aligned regions (Tetranucleotides >99.9%, ANIb >98%, and 16S rRNA ANI 98%) (**Supplementary Tables S6, S7**), indicating that they belong to the same genospecies (Konstantinidis and Tiedje, 2005). The absence of a clear biogeographical structure, in terms of the genetic divergence between strains from the two sites, represented also by clonal frame (CF) distributions (detailed below), is in line with earlier results based on Pulse Field Gel Electrophoresis (PFGE) patterns (Peña et al., 2005), and Multi-Locus Sequence Analysis (MLSA) (Rosselló-Mora et al., 2008). To avoid biases, all genomes, including the two from other studies, were annotated (or re-annotated) using a common framework (see section "Materials and Methods"). Thanks to the use of updated databases and methods, the annotation of the previously available strains, M31 and M8, improved with respect to the

available version. For instance, for M31 our annotation detects 296 additional genes, and reduces the number of hypothetical proteins from 35 to 24%.

Pangenome Distribution, Architecture and Overall Conservation of Synteny

The core genome comprised large syntenic blocks that were extremely conserved at the sequence level, and were interrupted by inter-spersed GIs and indels, which together with plasmids comprise the flexible genome (Figure 1 and Supplementary Figure S2). This is in line with observations from other free-living genomic models of aquatic bacteria such as *Prochlorococcus* sp. (Coleman et al., 2006), *Pelagibacter ubique* (Vergin et al., 2007) or *Alteromonas macleodii* (López-Pérez and Rodríguez-Valera, 2016). All vs. all genome comparisons of the eight strains defined a core genome of 2,389 genes, which on average represented 73.53% of the genome. This indicates that *S. ruber* entails higher microdiversity than the bacterial average (Lapierre and Gogarten, 2009), and closer to that of pathogenic bacteria than other free-living extremophiles such as *S. islandicus* (86%) (Cádiz-Quiroz et al., 2012) (Figure 2). Our reconstructed *S. ruber* pangenome comprises 5,792 genes (2,389 core and 3,403 accessory genes). Remarkably, most of the expansion of the flexible genome occurs within strains of the same CF, and even between co-isolated strains. This fact is also reflected in the high proportion of strain-specific genes, with each newly sequenced genome contributing about 300 new genes. This shows that *S. ruber* has an open pangenome (Figure 2), very similar to what has been reported for *Escherichia coli* (Mira et al., 2010) or *Haemophilus influenzae* (Hogg et al., 2007), which are pathogens with large, diverse, and highly dynamic flexible genomes resulting from their pathogenic lifestyle (Roberts and Kreth, 2014), or some free-living marine species such as *A. macleodii* (López-Pérez and Rodríguez-Valera, 2016). We also found that the plasmid content varied among the strains, both in terms of plasmid type and copy number (inferred from mean relative coverage) (Supplementary Table S5).

Evolution of the Core Genome

We next set out to investigate possible evolutionary mechanisms underlying the above described genomic variability in both the core and the accessory genome. In other aquatic microorganisms such as *Shewanella baltica* or *A. macleodii*, it has been shown that HR shapes the diversity of the core genome (Fernández-Gómez et al., 2012; López-Pérez and Rodríguez-Valera, 2016). In order to explore the impact of HR in the evolution of *S. ruber* core genomes, we estimated the clonal genealogy, as well as r/m and ρ/θ (σ/θ) ratios (Vos and Didelot, 2009) (see section “Materials and Methods”). Our results suggest that the eight analyzed sequences belong to five distinct CFs. CF is a term that refers to the smallest unit of differentiation in prokaryotes, (also referred to as strain biotypes or serotypes) and equivalent to the concept of “epidemic clone” used in the clinical field (Cui et al., 2015; López-Pérez and Rodríguez-Valera, 2016; Osaka et al., 2018). Strains within the same CF showed lower ANI_b and dN/dS pairwise distances as compared to strains from different CFs (Figure 3), which is in line with previous studies in which HR

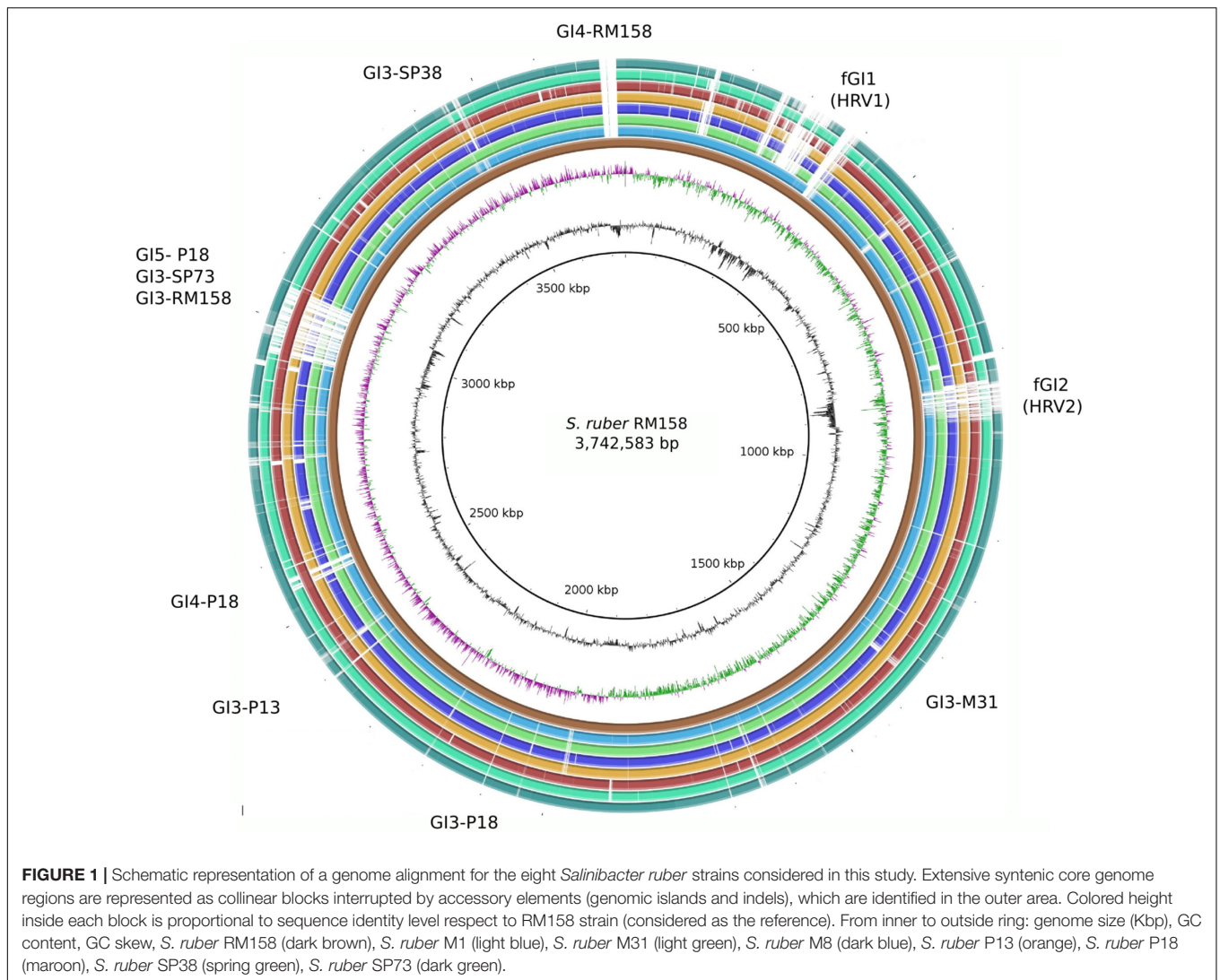
was shown to act as the main evolutionary mechanism in highly recombinogenic bacteria (Castillo-Ramírez et al., 2011; Cádiz-Quiroz et al., 2012). Recombination ratios (r/m) and (σ/θ) values for *S. ruber* were 1.52 and 0.29, respectively, higher than those reported for many other species with different lifestyles (Vos and Didelot, 2009), but similar to other extremophiles, where HR is thought to be one of the main evolutionary forces (Papke et al., 2004; Whitaker, 2005).

Metagenome recruitment analysis (Supplementary Figure S3) revealed discontinuities around 95% of ANI, as described by previous studies in marine microorganisms (Luo and Konstantinidis, 2011; Caro-Quintero and Konstantinidis, 2012). The maximum distribution peak is around 98% ANI, well above 95%, considered the threshold for species with cohesive population structures (Cuadros-Orellana et al., 2007; Tully et al., 2015) maintained by HR (Fraser et al., 2007). This 95% identity threshold corresponds with a 70% DNA-DNA hybridization value, above which two strains are considered to be from the same species and under which HR efficiency decreases rapidly (Konstantinidis and Tiedje, 2005; Caro-Quintero and Konstantinidis, 2012; Yarza et al., 2014). Moreover (r/m) ratio value in *S. ruber*, 1.52, well above the reference value of 0.25 (r/m), considered the minimum for which HR can act as a cohesive force on the population emerging clusters (Fraser et al., 2007).

Altogether r/m measures, ANI_b and dN/dS among CF support a predominant role of HR in maintaining genome similarity (both at the sequence and the gene order level) in the genomic core regions of *S. ruber*. This high impact of HR would explain previously reported phylogenetic inconsistencies in MLST studies (Rosselló-Mora et al., 2008) and the absence of clear biogeographic patterns (Peña et al., 2005). Similarly, we observed high levels of phylogenetic inconsistency when the rRNA gene tree was compared to phylogenetic reconstructions of randomly selected core genome regions (Figure 3). Despite this general trend, some previous studies have detected biogeographical patterns when analyzing metabolomic and HGT events profiles (Rosselló-Mora et al., 2008; Antón et al., 2013; Peña et al., 2014), probably reflecting recently acquired elements from the accessory genome pool and the effect of alternative evolutionary mechanisms acting over this fraction.

Core Genome Components and Functions Enriched in Recombined Regions

High-salinity environments generally are concomitant with high levels of temperatures, heavy metals, and UV radiation. In these conditions microorganisms have developed strategies that promote the repair of DNA. These include repair and recombination systems (García-González et al., 2013; van Wolferen et al., 2013), competence capabilities, as well as defense systems such as restriction modification (RM) (Oliveira et al., 2016) and CRISPR-Cas (Mojica et al., 2005), the last two acting as secondary barriers of DNA integration by means of HR (Thomas and Nielsen, 2005). So far no competence system has been described for *S. ruber*. We detected comEA (*rec2*) in the



core genome. This gene is involved in DNA internalization in competent Gram-negative bacteria (Takeno et al., 2011). Internalized DNA can also be exploited as a carbon source, especially in environments such as halophilic media with high concentrations of DNA and viral lysis (Thomas and Nielsen, 2005; Rodríguez-Valera et al., 2009). In addition, we detected the RecFOR repair and recombination system which, similarly to the RecBCD and SOS systems, is able to differentiate between own and foreign DNA by means of *Chi* sequences (Halpern et al., 2007), and which additionally acts as HR-dependent DNA repair system. We detected a low abundance of RM and CRISPR-Cas systems in the core genome. This could be compensated by the presence of an adapted and optimized RecFOR system, as it has been noted in other species (Handa et al., 2009; Vasu and Nagaraja, 2013). RecFOR system could be contributing to the maintenance of species identity, acting as a primary barrier against foreign DNA integration (Jeltsch, 2003).

Genes under positive selection or involved in HGT have been implicated in adaptation to the environment (Yahara et al., 2016).

We here explored whether core genome regions involved in HR events are enriched in certain functions (see section “Materials and Methods”). Altogether, we detected 958 HR events (**Supplementary Figure S4**), which collectively covered a 31% of the genome. The number of detected events in pairwise comparisons was congruent with the clonal frame tree topology (**Figure 3**), and the results of ANI, TETRA, and phylogenetic analyses (**Supplementary Tables S6, S7**). Only 233 events (25%) were longer than 10 kb, and 28 (3%) were longer than 50 kb. Four out of the six events longer than 100 kb covered fGI-1, in agreement with previous observations (López-Pérez et al., 2014b; López-Pérez and Rodríguez-Valera, 2016), and with the Constant Diversity model (Rodríguez-Valera et al., 2009) as further discussed below. Functional analysis revealed an under-representation of 10 different COG categories (Fisher’s test, $p < 0.05$; FDR < 0.05) in genomic regions involved in HR: “Ribosomal structure” (COG J), “Cellular motility” (COG N), “Postranscriptional modification” (COG O) and most of the categories related to cellular metabolism (**Supplementary Table S8**). This

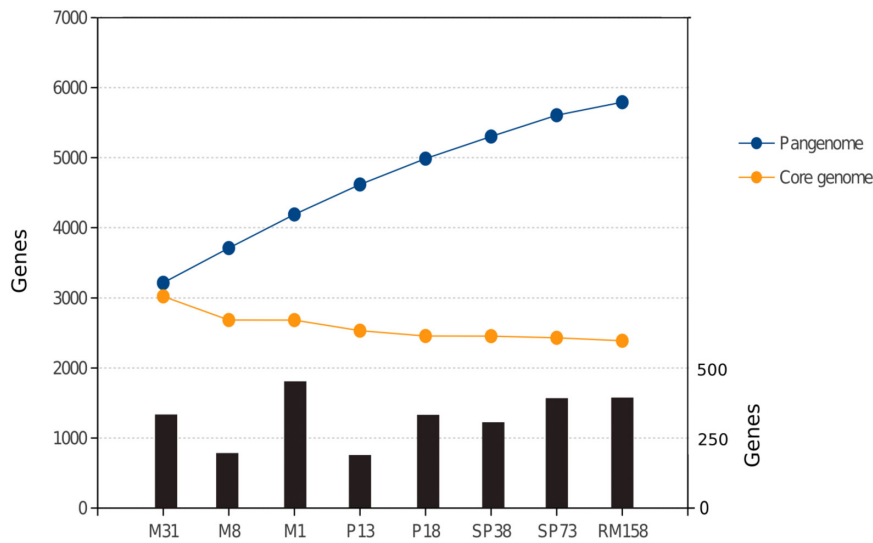


FIGURE 2 | Gene content in the pangenome (blue dots) and stable core genome (orange dots) as *S. ruber* strains analyzed in this study are sequentially considered in a random order. Black bars indicate the number of new genes contributed to the pangenome by each strain (secondary y-axis in the right).

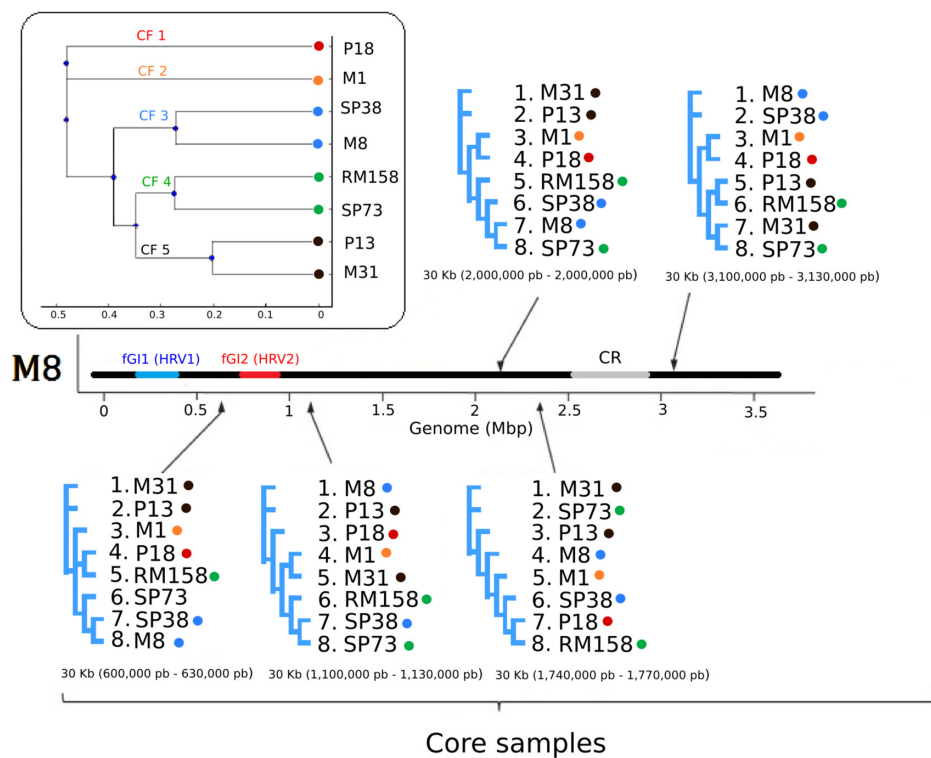


FIGURE 3 | Maximum likelihood trees of the eight *S. ruber* strains included in this study. Members of the five CFs identified (box) were considered and labeled with color-coded dots. The consensus tree based on the whole aligned core genome is shown in the box. The trees surrounding the genome were calculated based on alignments of randomly selected 30 Kb core genome regions, whose relative positions in *S. ruber* M8 are indicated in brackets. In all the cases *S. ruber* M8 was used as a reference genome to locate the position of the sequence used to generate the trees.

underrepresentation of metabolic genes in regions involved in HR is in stark contrast with their preponderance among regions involved in bacterial HGT (Popa et al., 2011). Among

“Replication, recombination and repair” (COG L), 58% (29/50) of recombined genes corresponded with “site-specific tyrosine recombinase” (XerD) term, which was the most abundant

term among HR regions. These enrichments support a role of transposable elements and *XerD* in HR processes, directly promoting their own mobilization and likely hitchhiking neighboring genes (Thomas and Nielsen, 2005), thereby contributing to genetic exchange between accessory and core genome fractions (García-González et al., 2013). Amongst COGs involved in “cellular processes and signaling” and “cellular metabolism” we found significant enrichments ($P < 0.05$, $p\text{FDR} < 0.05$, Fisher’s test) in functional terms related to glycosylation, “Cell wall/membrane/envelope biogenesis” (COG M) and “Carbohydrate transport and metabolism” (COG G) (**Supplementary Table S8**), which suggest a role of HR in the dynamics of envelope diversity. This includes modifications of surface elements usually encoded within GIs O-antigen clusters of free-living and pathogenic prokaryotes (Dobrindt et al., 2004; Fernández-Gómez et al., 2012; Rodríguez-Valera et al., 2016) and over syntenic clusters in the case of *S. ruber* fGIs as described below. Complex envelopes with a variable repertoire of glycosylases, together with the presence of the RecFOR system, detailed above, may compensate for the low abundance of RM and CRISPR-Cas systems in *S. ruber*, and could act as major defense systems against viral infection (Martín-Cuadrado et al., 2015; Rodríguez-Valera et al., 2016). O-antigen diversification has been described as an adaptive mechanism in response to environmental viral pressure in free living bacteria (López-Pérez et al., 2013) or to the host immune system in pathogenic bacteria (Schmidt et al., 2003; Wang and Quinn, 2010). Thus, recombination could sustain the high levels of inter-strain diversity observed in previous *S. ruber* metabolomic and transcriptomic studies (Rosselló-Mora et al., 2008; Antón et al., 2013; González-Torres et al., 2015). Additionally, we found enrichments in functional terms related to resistance, pathogenesis and adaptive mechanisms linked to “coenzyme and inorganic ion transport” (COG H and COG P), and “defense mechanisms” (COG V). Among the enriched GO terms, we detected multidrug or antibiotic resistance, transport systems, or beta-lactam resistance (**Supplementary Table S9**), which modulates intraspecific cellular communication or competition in several bacteria, including *S. ruber* (Bernier and Surette, 2013; González-Torres et al., 2015). Over 70% of recombined genes within COG V category were related to multidrug resistance or resistance to antibiotics, which in *S. ruber* are associated with intraspecific cellular communication processes or competition (González-Torres et al., 2015). Among transport systems, we found enrichments in ABC transporters, and Fe and TonB external receptors, which would suggest positive selection on HR events that affect transport capacities, or environmental adaptation under iron-limiting conditions as described in both free-living and pathogenic human bacteria (Hacker and Carniel, 2001; Gonzaga et al., 2012; Yahara et al., 2016).

Flexible Genome: Diversity and Architecture of GIs

We observed a large microdiversity of the accessory genome, exemplified by the variable number (2–5), size (10–200 kb), gene content, and structure of GIs among *S. ruber* genomes

(**Table 1**). The genomic fraction represented by GIs ranged from 6.1 to 10.6%, which is larger than those observed in many other *Bacteroidetes* and bacterial classes (2–5%) (Fernández-Gómez et al., 2012). As usual (López-Pérez and Rodríguez-Valera, 2016), these regions presented lower GC content and Codon Adaptation Index (CAI) values than the genome average, and were under-represented in environmental recruitments (**Supplementary Figure S5**), reported also in the case of *S. ruber* M8 and M31 strains (Pasić et al., 2009; Peña et al., 2010). Here, the availability of a larger number of genomes from different CFs allowed us to characterize these GIs under a functional and structural perspective. We confirm the presence of two types of GIs, as reported for other bacteria (Fernández-Gómez et al., 2012; López-Pérez et al., 2014a). Firstly, fGIs appeared as two regions (fGI-1 and fGI-2) located in the same relative position, corresponding to hypervariable regions (HRV-1 and HRV-2) (Peña et al., 2010). Secondly, a variable number of HGT-GIs, indels, and plasmids, which are mostly strain-specific but sometimes shared among few strains (**Supplementary Figure S5**).

The flexible genomic Islands IGF-1 and IGF-2 contained mainly genes from the accessory genome (**Figure 1** and **Table 1**) and only 13 and 15 core genome genes over syntenic clusters or cassettes (**Figure 4**), respectively, as it is often the case in additive fGIs (López-Pérez et al., 2013). These syntenic cassettes, sometimes shared by all strains, were interspersed by transposable elements, genes of viral origin, TonB transporter genes, and TonB-dependent receptors. Some of these TonB receptors are TonB-linked outer membrane proteins that belong to protein families restricted to the *Bacteroidetes* lineage such as the SusC/RagA (IPR023996) family. Transporter complexes including these outer membrane proteins are involved in the transport of outer membrane proteins and carbohydrates (glycans) in association with SusD proteins, that bind glycans to the cell surface and facilitate shuttling to the TonB-dependent porin (Bolam and Koropatkin, 2012). Additionally, we found functionally conserved blocks with different gene content, usually located in the same areas within the replacement fGIs-2 (**Supplementary Figure S6**). These latter clusters shared only one common protein, as exemplified by a bacteriocin cluster or an O-antigen cluster (**Supplementary Figure S6**). Amongst fGIs we detected an enrichment ($p < 0.05$, $\text{FDR} < 0.05$, Fisher’s test) in genes annotated as “cell wall and membrane envelope biogenesis” (COG M) and “Mobilome: prophages, transposons” (COG X), corresponding mainly to transposable elements and hypothetical proteins (58%), (**Supplementary Table S10**). The high content in hypothetical proteins and transposable elements is associated with the presence of new genes acquired by HGT during microbial adaptation (Dobrindt et al., 2004; Hsiao et al., 2005; Roberts and Kreth, 2014) and recent duplications in *S. ruber* (González-Torres et al., 2015). Enriched COG M genes, mainly coding for sulfotransferases and glycosyltransferases, and located within the O-antigen cluster, are involved in processes of recognition and immunospecificity virus–host recognition (Peña et al., 2010; Wang and Quinn, 2010; López-Pérez and Rodríguez-Valera, 2016). The high diversity and intraspecific genome dynamics of surface genes

TABLE 1 | Characteristics of the genomic islands found in the eight *Salinibacter ruber* strains compared in this study.

Genome	GI Classification	GI ID	Number ORFs	Size (kb)	Initial position (bp)	Final position (bp)	%GC	%Genome covered
M8 M(1999)	fGI	fGI1 (HRV1)	127	151.71	249,507	401,287	61.3	3.96
	fGI	fGI2 (HRV2)	134	129.92	829,018	958,941	58.2	3.39
	fGI	fGI1 (HRV1)	111	134.12	223,961	358,080	62.1	3.74
M31	fGI	fGI2 (HRV2)	82	88.14	774,703	862,838	59.1	2.46
M(1999)	HGT-GI	GI3-M31	44	42.75	1,360,489	1,404,241	55.8	1.19
M1 M(1999)	fGI	fGI1 (HRV1)	151	177.85	235,218	413,070	60.5	4.72
	fGI	fGI2 (HRV2)	44	51.25	841,588	892,833	62.2	1.36
P13 SP(1999)	fGI	fGI1 (HRV1)	139	140.56	246,865	387,420	61.9	3.90
	fGI	fGI1 (HRV1)	77	74.90	819,096	893,996	60.8	2.08
	HGT-GI	GI3-P13	30	27.30	2,335,400	2,362,700	50.6	0.76
P18	fGI	fGI1 (HRV1)	87	109.18	237,245	346,422	62.7	2.92
SP(1999)	fGI	fGI2 (HRV2)	50	56.79	793,772	850,557	62.4	1.59
	HGT-GI	GI3-P18	18	33.80	1,990,500	2,024,300	58.7	0.90
	HGT-GI	GI4-P18	35	43.50	2,544,500	2,588,000	66.0	1.16
	HGT-GI	GI5-P18	105	114.61	2,608,780	2,723,390	61.4	3.06
SP38	fGI	fGI1 (HRV1)	145	168.38	230,386	398,753	62.5	4.55
SP(2006)	fGI	fGI2 (HRV2)	42	50.10	817,511	867,607	62.5	1.35
	HGT-GI	GI3-SP38	26	23.80	3,584,200	3,560,400	51.0	0.64
SP73 SP(2006)	fGI	fGI1 (HRV1)	205	196.16	256,113	452,273	60.4	5.03
	fGI	fGI2 (HRV2)	103	104.73	872,601	977,331	58.6	2.69
	HGT-GI	GI3-SP73	106	113.67	2,993,000	3,106,670	61.4	2.91
RM158	fGI	fGI1 (HRV1)	127	156.47	249,611	406,078	60.7	4.03
M(2006)	fGI	fGI2 (HRV2)	97	94.72	840,001	934,723	58.7	2.44
	HGT-GI	GI3-RM158	102	113.68	2,943,003	3,056,685	61.5	2.93
	HGT-GI	GI4-RM158	18	27.64	3,700,270	3,727,910	57.6	0.72

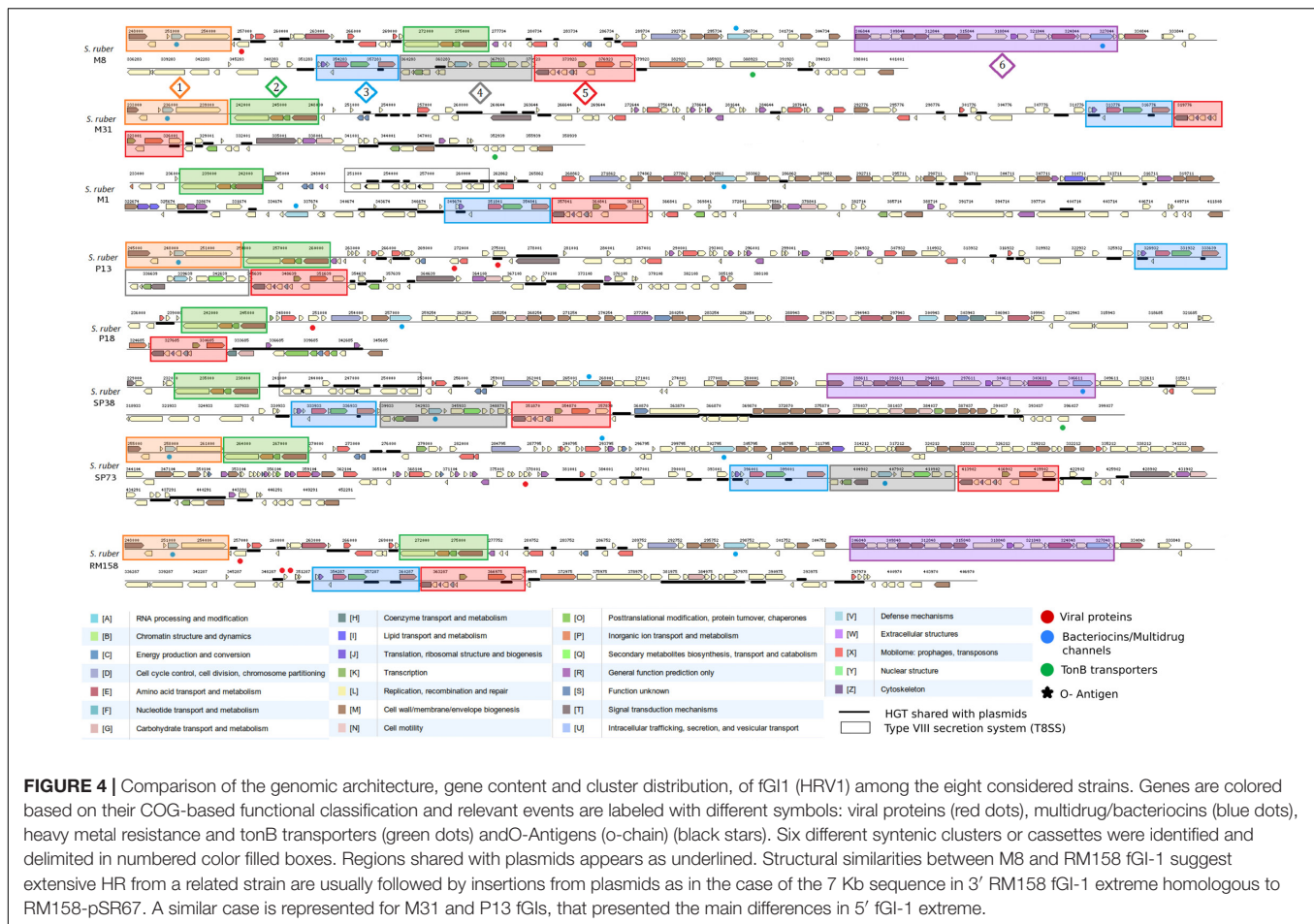
was consistent with recent metabolomic studies in *S. ruber* (Rosselló-Mora et al., 2008; Antón et al., 2013) and the known high adaptive pressure exerted by viruses from the environment (Santos et al., 2010; Martin-Cuadrado et al., 2015; López-Pérez and Rodríguez-Valera, 2016). Differences in cell surfaces affect susceptibility to phages and phage-host range (Peña et al., 2010; Villamor et al., 2017) and thus result in selection either by differential lysis impact or by processes of signal transduction, according to the “Constant Diversity” model, and in favor of a metaclonal scenario (Rodríguez-Valera et al., 2009). Bacteriocin and multidrug resistance proteins (including betalactamases) have been associated to molecules mediating signaling between bacterial populations in nature (Davies, 2006; Bernier and Surette, 2013). Indeed, efflux pumps, drug resistance, and antibiotic related pathways have been associated with adaptive mechanisms of expression response that mediate intra-specific interactions between co-occurring strains of *S. ruber* (González-Torres et al., 2015). Thus the mechanisms that modulate the content of such elements in the accessory genome could play important roles in adaptation to halophylic environments.

Most HGT-GIs were short (22–45 kb) and strain specific (Supplementary Figure S7 and Supplementary Table S1), with the exception of three (GI5-P18, GI3-SP73, and GI3-RM158), which were homologous and syntenic, and spanned a large region of 113 Kb. All HGT-GIs were flanked by tRNAs, repetitive regions, XerD or phage-related recombinase, which potentially facilitates integration and excision in HGT

events (Hacker and Kaper, 2000). The five smaller HGT-GIs were enriched for hypothetical proteins. GI-4P18 (43 kb) contained many membrane genes (COG M), some involved in resistance to organic solvents, as well as transcription factors (COG K), including sigma 70. The three larger HGT-GIs (GI5-P18, GI3-SP73, and GI3-RM158) presented an identical content, except for a 3 kb-long indel generated by transposition in the GI3-SP73. This GI, integrated in a previous event next to an indel present in all the strains (in the case of P18 in a reverse direction), contained several genes coding for ion or carbohydrate transporters (COG P and COG G, respectively) and structural membrane elements (COG M). Many of these transporters code for proteins involved in heavy metal resistance (Cu, Mg, Au, Co, Zn, Cd) and cofactor transport (Zn, Fe, Co). Hence, this GI could act as a metal resistance island, by analogy to the halophilic island described in M31 (Mongodin et al., 2005). Incidentally, it could represent a potential bioremediation element. Heavy metal resistance elements are commonly associated to GIs, mobile elements and adaptive processes (López-Pérez et al., 2013; Bellanger et al., 2014; Martin-Cuadrado et al., 2015).

Plasmid Diversity and Plasmid-Chromosome Genetic Exchanges

A total of eighteen plasmids were present in the analyzed strains, in variable numbers (0–4 per strain) and sizes (10–116 Kb). Plasmids showed a lower GC content as compared



to the genome average, and displayed the characteristic low copy replicative origin (ParA/RepB) (Pinto et al., 2012; Ietswaart et al., 2014) (**Supplementary Figure S1** and **Supplementary Table S5**). Similarly to GIs, plasmids showed a high proportion of hypothetical proteins (35%) as compared to the whole genome (chromosome and plasmids) (24%). Smaller plasmids contained the greatest proportion of short ORFs, and of viral recombinase, XerD or terminase (as in case of RM158-pSR31 with 45 ORFs in 31 Kb) (**Supplementary Figure S9**). Larger plasmids contained barrier elements involved in DNA exchange such as CRISPR-Cas, RM systems, secretion systems, as well as membrane elements (**Supplementary Figures S8, S9**). M8-pSR61 plasmid, pSR66-M1, M8-pSR84, and RM158-pSR67 contained clusters of genes coding for type VIII secretion system (T8SS, like curlin) present also in M1 and SP38 fGI-1 (**Supplementary Figures S8, S9**). The presence of secretion systems in virulence or conjugative GIs and plasmids is common among different ecological strategies (Fernández-Gómez et al., 2012; Melville and Craig, 2013). Finally the M8-pSR84 plasmids, RM158-pSR60 and M1-pSR116 contained many genes associated with membrane elements (COG M), some linked to the O-antigen cluster described in the fGIs-II, as well as ionic transporters (COG P) related to resistance to heavy metals and iron transport. We also detected the presence of genes clearly connected with halophylic adaptation such as

bacteriorhodopsin (Oren, 2013) in plasmid M1-pSR116. The presence of these common elements (O-antigen and T8SS) between fGIs and plasmids suggests that the former could serve as the transmitter vehicle, eventually incorporating sequences to the GI. The comparison of this same plasmid, M8-pSR84, with fGIs-1, allowed us to identify in fGIs-1 of SP38 a 30-kb region with high identity (93%, e -value = 0) and synteny. In particular the two last mentioned plasmids, M8-pSR84 and RM158-pSR67, shared a 10 Kb-long syntenic region with a 99.6% sequence identity (block 3, **Supplementary Figures S8, S9**) that included T8SS. The above results suggest that plasmids act as a dynamic vehicle for access to the environmental pool, while the GI would be the gateway to the chromosome.

To further investigate this possible mechanism, we compared the sequences of each plasmid against the genomes of the other strains, and identified plasmidic regions integrated in the chromosomes. Approximately 50% of hits occurred at fGIs-1 and IGF-2, which represent a small fraction of the genome (6.1–10.6%) (**Supplementary Figure S10**). This would explain that strains of different CFs and geographical origins can share accessory genome sequences. Similar findings have been reported for different species (López-Pérez et al., 2013, 2017). Although we failed to find Integrative and Conjugative Elements (ICEs) in the plasmids, we did find some elements coding for proteins

involved in integration and excision, including site-specific tyrosine recombinases (XerD) and integrases. The presence of the genes *comEC* and *comF* in all *S. ruber* strains, suggests that transformation and transduction may provide alternative mechanisms for the transfer of plasmids, as suggested for *A. macleodii* (López-Pérez et al., 2017). Incorporation of new genes in GIs can play a role in modulation of transcription and signal transduction (Coleman et al., 2006). In this respect, the translocation of a gene from a plasmid to a GI could result in increased expression, as some regions in GIs such the O-antigen gene have a much higher expression than plasmids (González-Torres et al., 2015). The rest of plasmids showed peaks of identity between 86 and 94%, which could indicate that these sequences are so ubiquitous in various species that

they may have exchanged sequences from other replicons, as observed previously (López-Pérez et al., 2013). In addition, HR events extensively involved fGIs (fGIs-1 and fGIs-2 to 4), as detailed in **Figure 4**, and included events longer than 100 Kb that involved different CFs, in agreement with the Constant Diversity model and metaclonal scenario (Rodríguez-Valera et al., 2009). In some cases HR events affected short syntenic clusters (5–12 kb), and resulted in phylogenetic incongruencies (**Supplementary Figure S11**) (Fernández-Gómez et al., 2012). Altogether, these results underscore the complex intraspecific dynamic network regulating the access to the accessory gene pool in *S. ruber*. The emerging picture is that of plasmids exchanging genetic material with GIs, and these acting as gateways or hotspots, driving exchanges with the rest of the genome including syntenic

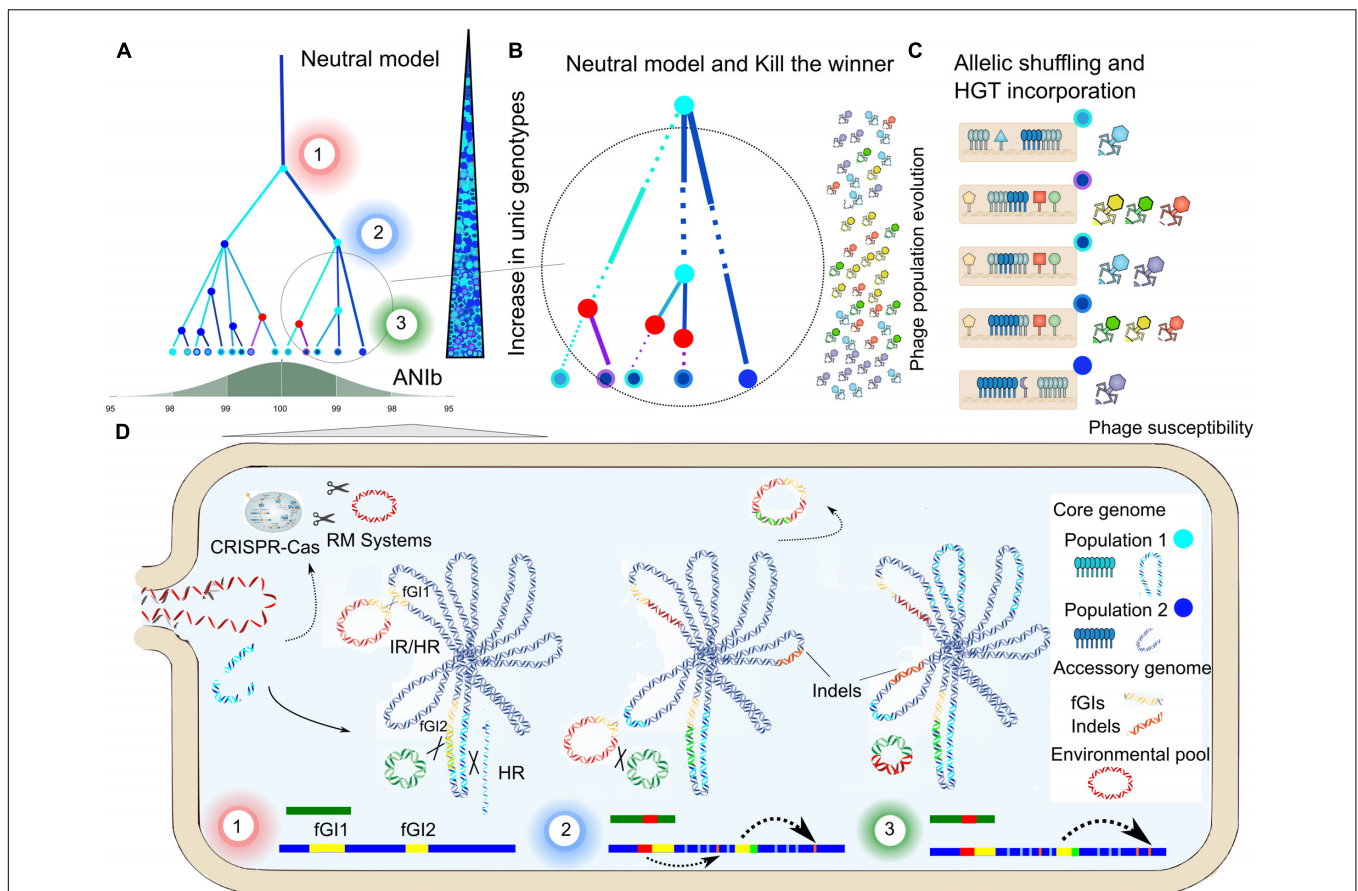


FIGURE 5 | Genome dynamics and ecological models. Schematic representation of: **(A)** the cohesive effect of HR, as predicted by the Neutral Model, limits sequence variation in two *S. ruber* populations (dark blue and clear blue dots). The number of unique genotypes (represented as dots outlined and filled with different colors) increases over time due to the core genome shuffling, although ANIb is maintained over 98% due to the cohesive effect. Significant HR events are highlighted as colored nodes, in which the color indicates the donor (red nodes refers to HGT gene incorporation). Numbered from 1 to 3, different snapshots over the population tree represented in **(D)**. **(B)** Detail of the population tree represented in **(A)**, incorporating the simultaneous effect of phage predation. Branch thickness is proportional to clonal frame abundance, which is modulated periodically by phage populations, as predicted by the Kill the Winner model. **(C)** Allelic shuffling in core genome genes (blue) and phage susceptibility, resulting from glycosyltransferases and surface receptors content (colored with the same code of the corresponding phage). **(D)** Detail of the population tree represented in **(A)**, representing the genome evolution (main chromosome in dark blue and plasmid in green) of a clonal frame from the population 2. Exchanges between clonal frames, by means of illegitimate recombination (IR) or homologous recombination (HR), would be modulated according to the contents in RM and CRISPR-Cas systems. Genomics Islands (GIs) act as gateways of elements of the accessible pool (colored in red) through the integration of plasmidic elements and HR in case of syntenic clusters, contributing to glycosyltransferases. The bars at the bottom represent the linearised genome regions (chromosome in dark blue and plasmid in green) and the recombined regions in a similar way that in the **Supplementary Figure S4**.

regions. Hence, plasmids would serve as an important vehicle contributing to genome dynamics, in which the barrier elements encoded by them would modulate inter and intraspecific genetic flows in a strain, thereby affecting its capabilities and abundance (Figure 5D).

The Influence of RM and CRISPR-Cas Systems Evolution on DNA Exchanges

The presence of RM and CRISPR-Cas systems in the accessory genome reflects the complex interactions that exist between gene exchange processes (Corvaglia et al., 2010; Fernández-Gómez et al., 2012; Bellanger et al., 2014). A clear example is provided by the presence of the IC-prr type MR system in the chromosomes of M1, P13, and SP38 strains, in equivalent positions, flanked by XerD (Supplementary Figure S12). This unique indel flanked by a recombinase located in an otherwise conserved region, suggests a provenance from accessory regions (plasmids or GIs) (Figure 5). In the case of M1-pSR61 plasmid, we also identified an additional RM system, type 1A/B. The presence of different sets of RM systems among strains in a population determines access to the environmental pool and intraspecific exchanges favoring incipient speciation (Cadillo-Quiroz et al., 2012). In this sense, *S. ruber* strains in the same CF presented the same RM system content, and similar HR levels, which were higher in those strains that contained a RM system. This finding supports the idea that there are limits to intraspecific exchanges between strains coding for different RM type systems (Corvaglia et al., 2010). Additionally, we detected the presence of complete CRISPR-Cas systems in the plasmids M1-pSR66 (subtype I-b of halophilic) and SP73-pSR118 (subtype IE *E. coli*) (Makarova et al., 2011) with 38 and 105 spacers, respectively (Supplementary Figure S13 and Supplementary Table S11), as well as additional minor CRISPR arrays in plasmids from these same strains that could act in trans. The CRISPR-Cas system detected in SP73 was syntenic to the *Rhodothermus marinus* R-10 (DSM4254) (Nolan et al., 2009), a species phylogenetically related to *S. ruber*. The presence of these elements in the accessory genome could indicate an adaptive role against phage predation in extreme environments (Mojica et al., 2000; Sorek et al., 2008). Comparisons of spacers against halophylic metagenomes/metaviromes and *S. ruber* (Villamor et al., 2017) and *Haloquadratum walsbyi* (Martin-Cuadrado et al., 2015) halophage sequences confirmed this fact (Supplementary Tables S4, S11), associating a viral or prokaryotic protospacer in almost 70% of the cases in M1 (Supplementary Figure S13). Two M1 spacers and another from SP73 hit contig 170 from Santa Pola crystalizer CR30 metavirome (Santos et al., 2010). The independent acquisition of spacers for the same mobile element, and the lack of shared spacer sequences between both strains, suggest that, although Mallorca and Santa Pola saltern ponds share a pool of viruses, both viral populations presented biogeographical differences promoting specific adaptations. The presence of M1 spacers whose precursor exogen sequence, protospacer, corresponds to isolated viruses has been related to lower infection susceptibilities (Villamor et al., 2017). Acquisition of independent resistance among CFs would prevent the clonal sweep in periodic episodes, limiting the loss of diversity (Lythgoe

and Chao, 2003; Held et al., 2010). Moreover, we found a remarkable number of hits (75) in Lake Tyrrell metavirome of 2010, surpassing the number detected in 2007 (5) in these same ponds (Supplementary Figure S13). These findings support the persistence of spacers from viruses that may have had periodic abundance fluctuations within a single pond, fitting the virus-host dynamics predicted by the “Constant Diversity” model under a metaclonal scenario (Rodríguez-Valera et al., 2009).

CONCLUSION

Our comparative analysis of eight whole *S. ruber* genomes revealed a diverse and open pangenome, in which each strain presented a unique plasmid and GI profile. In contrast, the core genome is extremely conserved in terms of sequence identity and gene order. We show that HR constitutes the main evolutionary force acting on core genomes. Indeed HR affects over 30% of the genome, breaks clonal structures, and limits sequence divergence within population clusters. Moreover the presence of transposable elements, XerD, recombinases and phage terminases, also abundant in fGIs, HGT-GIs and plasmids suggested a complete and dynamic network of genetic exchange between the accessory and the core genome by means of homologous or illegitimate recombination (Figure 5). In this model, GIs act as gateway of elements of the accessible pool through the integration of plasmidic elements and HR in case of syntenic clusters, contributing to glycosyltransferases and metabolic capacities diversity through transfection or transformation. Exchanges between clonal frames would be modulated according to the contents in RM and CRISPR-Cas systems.

The observed HR evolutionary parameters, together with genomic traits (ANIb and dN/dS) and data from metagenomics recruitments, fitted with Fraser’s Neutral model (Fraser et al., 2007), reflecting the cohesive effect of HR over *S. ruber* populations maintaining the species identity. This model highlights the role of HR as a passive, neutral mechanism driven solely by genetic divergence. However, we detected functions that are over-represented in genes that formed part of recombined regions. In many cases these over-represented functions are potentially related to adaptations to the halophilic niche inhabited by *S. ruber*, suggesting the influence of natural selection. These include, among others, accessory genes involved in intraspecific cellular communication, multidrug resistance, ionic transport, as well as genes involved in glycosylation and thereby in resistance to phage predation. Resistance to phage predation is the key trait considered by the “Kill the Winner” model, which contemplates frequency-dependent selection on bacterial phage-susceptibility types as mediated by phages. It has been suggested that periodic phage predation dynamics may be considered simultaneously with other evolutionary models (Rodríguez-Valera et al., 2009) and also that different ecological and evolutionarily processes may play an important role on parallel trajectories under a metaclonal scenario (Good et al., 2017).

We consider that our results provide support for a synergistic action of processes postulated by both the Neutral and the Constant Diversity models. Our proposed framework (**Figure 5**) considers a metaclonal scenario, in which core and accessory genomes would be differentially affected by HR and HGT, resulting in the observed high microdiversity levels in *S. ruber*, while preventing clonal sweep and limiting sequence variation within clonal frames. Both, biotic (viral) and abiotic conditions (temperature, UV radiation, oligotrophy and salinity) justify the presence of recombination machinery and the high levels of HR, which shuffles the content of both the core and flexible genomes. The coexistence of several clones with different characteristics in a population would dilute the selective pressure by phages and would result in a more efficient exploitation of the scarce resources in a stressful environment, as suggested by the Black Queen hypothesis (Morris et al., 2012). This equilibrium could be maintained by a phage population that would prevent any clonal sweep or periodic selection event. Altogether, the high diversity and genome dynamics uncovered by the comparison of the first eight genomes from *S. ruber* highlights the suitability of this species as a model for the study of genome micro-evolution and speciation in hypersaline environments.

AUTHOR CONTRIBUTIONS

TG supervised the project. PG-T and TG planned the sequencing and bioinformatics analysis and wrote the manuscript. PG-T performed the bioinformatics analyses.

FUNDING

PG-T was an FPI-MINECO fellow. TG group acknowledges support of the Spanish Ministry of Economy and Competitiveness grants, ‘Centro de Excelencia Severo Ochoa 2013–2017’ SEV-2012-0208, and BFU2015-67107 cofunded by European Regional Development Fund (ERDF); from the CERCA Program/Generalitat de Catalunya; from the Catalan Research Agency (AGAUR) SGR857, and grant from the European Union’s Horizon 2020 Research and Innovation Program under the grant agreement ERC-2016-724173 the Marie Skłodowska-Curie grant agreement No. H2020-MSCA-ITN-2014-642095. We acknowledge the support of the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership.

ACKNOWLEDGMENTS

We thank Josefa Antón for her scientific insights during the project and for critically reviewing this manuscript. We thank Jesse R. Willis for critically revising the manuscript. We thank the genomics core facility at CRG for their help with the sequencing procedures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01499/full#supplementary-material>

FIGURE S1 | Representation of 24 replicons (8 chromosomes and 18 plasmids) present in the eight strains in this study. The main traits (nomenclature, genome size, gene and GC content) are detailed for each genome. From inner to outside ring: Ring 1: represents GC skew index (GCSI) defined as (G-C/C+C) (green and purple); Ring 2: GC content (%) (colored in black). Genes appears represented in external rings and colored based on clusters of orthologous groups (COGs) functional classification. Colors corresponds to those associated to each strain in **Figure 1**.

FIGURE S2 | Chromosome alignment representation for the eight *S. ruber* strains considered in this study. Extensive syntenic core genome regions appears represented as collinear blocks interrupted by accessory elements; genomic islands and indels. Colored fraction inside each block is proportional to the sequence identity level with respect to M8 strain.

FIGURE S3 | Metagenomic recruitment and discrete populations. Coverage plot representing fragment recruitment of the **(A)** Santa Pola CR30 crystallizer (Spain) (Fernandez et al., 2013); **(B)** Lake Tyrrell metagenome (Australia) and **(C)** two different San Diego High Salt ponds metagenomes (United States) against *S. ruber* main chromosomes and **(D)** Santa Pola CR30 crystallizer (Spain) (Ghai et al., 2011) against plasmids sequences. For each graph, the coverage (*y*-axis) is normalized to the length of the reference and represents the relative *in situ* abundance of the corresponding population.

FIGURE S4 | Homologous recombination map. Genomic map representing the inferred recombination events among the considered strains. Each strain appears associated to a color code (top bar) consistent with **Figure 1** and **Supplementary Figure S1**. The bottom bars represent the recombination events, which are colored according to the predicted donor strain. FGI1, fGI2 and the conserved region (CR) described in previous studies (Peña et al., 2010) are highlighted in the figure.

FIGURE S5 | Metagenomic recruitment and GI identification. Alignment of the eight *S. ruber* chromosome sequences with metagenomic sequences from Santa Pola saltern pond, CR30 crystallizer (Spain) (Ghai et al., 2011). Values on the *y*-axis indicate nucleotide percent identity, usually over 95%. Areas with unusually low representation in the metagenome corresponded to under-recruiting genomic islands: fGIs1 and fGIs2 (indicated in blue and green, respectively) and HGT-GIs in green.

FIGURE S6 | Overall comparison of the genomic architecture, gene content and clusters distribution, of fGI2 (HRV2) among the eight analyzed strains. Regions shared with plasmids are underlined. Six different syntenic clusters or cassettes were identified and delimited in numbered colored boxes. Gene colors and symbols were consistent with **Figure 4**.

FIGURE S7 | Gene content and functional traits of HGT-GI islands among the eight analyzed strains. RM systems are delimited with red boxes and tRNA genes are labeled with triangles. Gene colors and symbols were consistent with **Figure 4**.

FIGURE S8 | Gene content and functional traits of plasmids detected amongst three of the eight strains analyzed in this study (*S. ruber* M8, *S. ruber* M31, *S. ruber* M1). RM systems, type 8 secretion systems (T8SS) and CRISPR-Cas systems are delimited in boxes and genes are labeled with a triangle and appear flanking HGT-GI islands. Three different events were identified and delimited in numbered color filled boxes representing three cases of extensive homologous regions involving plasmids of four different strains: M8 (pSR84), SP73 (pSR118), SP38 (pSR76), and RM158 (pSR60, pSR67). Gene colors and symbols were consistent with **Figure 4**.

FIGURE S9 | Gene content and functional traits of plasmids detected amongst three of the eight strains analyzed in this study (*S. ruber* SP73, *S. ruber* M38, and *S. ruber* RM158). RM systems, type 8 secretion systems (T8SS) and CRISPR-Cas

systems are delimited in boxes and genes are labeled with a triangle and appear flanking HGT-GI islands. Three different events were identified and delimited in numbered color filled boxes representing three cases of extensive homologous regions involving plasmids of four different strains: M8 (pSR84), SP73 (pSR118), SP38 (pSR76), and RM158 (pSR60, pSR67). Gene colors and symbols were consistent with **Figure 4**.

FIGURE S10 | The main bar represents the cumulative percentage (*y*-axis) of 18 plasmid sequences, included in the genome of six *S. ruber* strains (*x*-axis), that presented homology with chromosome sequences among the eight different *S. ruber* strains (colored in the legend). Among these sequences, an additional unique secondary bar per strain represents the percentage of total sequences that presented homology with fGI1 (in red) and fGI2 (in black). For P13 and P18 strains (values not represented), the last distributions were 37.11 and 4.6% for fGI1 and 42.26 and 14.81% for fGI2, respectively.

FIGURE S11 | ML trees of 6 of the 10 syntenic clusters or cassettes shared by at least four strains and described in fGI1 and fGI2 (**Figure 4** and **Supplementary Figure S3**). Strains are indicated in all the branches and clonal frames (CFs) correspondence indicated by colored dots consistent with **Figure 4**. The line segment at the bottom part of each tree shows the branch length scale (substitutions/site units).

FIGURE S12 | RM systems diversity and microevolutionary mechanisms. **(A)** Genome alignment showing the syntenic genomic environment (underlined in red) interrupted in case of *S. ruber* M1, SP38 and P13 strains by an indel (underlined in black) that contained a type 1C-prr RM system, flanked by tyrosin recombinases XerD, likely involved in the chromosome integration. **(B)** Most common RM

systems present in bacteria (upper part) and those present amongst *S. ruber* strains (bottom).

FIGURE S13 | CRISPR-Cas systems microdiversity. **(A)** CRISPR-Cas systems found in *S. ruber* SP73 and *S. ruber* M1 plasmids, indicating the type and cas genes content and spacer sequence. CRISPR arrays were represented with dotted red lines. **(B)** Comparison and syntenic similarities between CRISPR-Cas system found in *S. ruber* SP73 strain and its best hit in *Rhodothermus marinus* DSM 4252 strain. **(C)** Distribution of the main protospacer hits amongst environmental metagenomes (yellow), metaviromes (red), and unknown (blue). **(D)** Number of hits found mean protospacer recruitment versus lake Tyrrell Site A and B metaviromes from 2007 and 2010 years.

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A., and Pevzner, P. A. (2016). plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* 32:btw493. doi: 10.1093/bioinformatics/btw493
- Antón, J., Lucio, M., Peña, A., Cifuentes, A., Brito-Echeverría, J., Moritz, F., et al. (2013). High metabolomic microdiversity within Co-occurring isolates of the extremely halophilic bacterium *Salinibacter ruber*. *PLoS One* 8:e64701. doi: 10.1371/journal.pone.0064701
- Antón, J., Rosselló-Mora, R., Rodríguez-Valera, F., and Amann, R. (2000). Extremely halophilic Bacteria in crystallizer ponds from solar salterns. *Appl. Environ. Microbiol.* 66, 3052–3057. doi: 10.1128/AEM.66.7.3052-3057.2000. Updated
- Ashburner, M., Ball, C. A., and Blake, J. A. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Bellanger, X., Payot, S., Leblond-Bourget, N., and Guédon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. *FEMS Microbiol. Rev.* 38, 720–760. doi: 10.1111/1574-6976.12058
- Bernier, S. P., and Surette, M. G. (2013). Concentration-dependent activity of antibiotics in natural environments. *Front. Microbiol.* 4:20. doi: 10.3389/fmicb.2013.00020
- Bolam, D. N., and Koropatkin, N. M. (2012). Glycan recognition by the Bacteroidetes Sus-like systems. *Curr. Opin. Struct. Biol.* 22, 563–569. doi: 10.1016/j.sbi.2012.06.006
- Boni, M. F., Posada, D., and Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176, 1035–1047. doi: 10.1534/genetics.106.068874
- Boujelben, I., Yarla, P., Almansa, C., Villamor, J., Maalej, S., Antón, J., et al. (2012). Virioplankton community structure in Tunisian solar salterns. *Appl. Environ. Microbiol.* 78, 7429–7437. doi: 10.1128/AEM.01793-12
- Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., Reno, M. L., et al. (2012). Patterns of gene flow define species of the thermophilic Archaea. *PLoS Biol.* 10:e1001265. doi: 10.1371/journal.pbio.1001265
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Höfle, M. G., Klappenbach, J., et al. (2011). Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella* bacteria from the Baltic Sea. *ISME J.* 5, 131–140. doi: 10.1038/ismej.2010.93

- Caro-Quintero, A., and Konstantinidis, K. T. (2012). Bacterial species may exist, metagenomics reveal. *Environ. Microbiol.* 14, 347–355. doi: 10.1111/j.1462-2920.2011.02668.x
- Castillo-Ramírez, S., Harris, S. R., Holden, M. T. G., He, M., Parkhill, J., Bentley, S. D., et al. (2011). The impact of recombination on dN/dS within recently emerged bacterial clones. *PLoS Pathog.* 7:e1002129. doi: 10.1371/journal.ppat.1002129
- Chu, T.-C., Lu, C.-H., Liu, T., Lee, G. C., Li, W.-H., and Shih, A. C.-C. (2013). Assembler for de novo assembly of large genomes. *Proc. Natl. Acad. Sci. U.S.A.* 110, E3417–E3424. doi: 10.1073/pnas.1314090110
- Cohan, F. M. (2006). Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1985–1996. doi: 10.1098/rstb.2006.1918
- Coleman, M. L., Sullivan, M. B., Martiny, A. C., Steglich, C., Barry, K., Delong, E. F., et al. (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311, 1768–1770. doi: 10.1126/science.1122050
- Conesa, A., and Göttsch, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832
- Corvaglia, A. R., François, P., Hernandez, D., Perron, K., Linder, P., and Schrenzel, J. (2010). A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical *Staphylococcus aureus* strains. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11954–11958. doi: 10.1073/pnas.1000489107
- Cuadros-Orellana, S., Martín-Cuadrado, A.-B., Legault, B., D'Auria, G., Zhaxybayeva, O., Papke, R. T., et al. (2007). Genomic plasticity in prokaryotes: the case of the square haloarchaeon. *ISME J.* 1, 235–245. doi: 10.1038/ismej.2007.35
- Cui, Y., Yang, X., Didelot, X., Guo, C., Li, D., Yan, Y., et al. (2015). Epidemic clones, oceanic gene pools, and Eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. *Mol. Biol. Evol.* 32, 1396–1410. doi: 10.1093/molbev/msv009
- Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. doi: 10.1371/journal.pone.0011147
- Davies, J. (2006). Are antibiotics naturally antibiotics? *J. Ind. Microbiol. Biotechnol.* 33, 496–499. doi: 10.1007/s10295-006-0112-5
- Didelot, X., and Falush, D. (2007). *Clonal Frame User Guide*. Available at: <http://www.stats.ox.ac.uk/~didelot/files/clonalframe-userguide.pdf> [accessed February 25, 2014]
- Didelot, X., and Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends Microbiol.* 18, 315–322. doi: 10.1016/j.tim.2010.04.002
- Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414–424. doi: 10.1038/nrmicro884
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Emerson, J. B., Andrade, K., Thomas, B. C., Norman, A., Allen, E. E., Heidelberg, K. B., et al. (2013). Virus-host and CRISPR dynamics in Archaea-dominated hypersaline Lake Tyrrell, Victoria, Australia. *Archaea* 2013:370871. doi: 10.1155/2013/370871
- Fernandez, A. B., Ghai, R., Belen Martín-Cuadrado, A., Sanchez-Porro, C., Rodriguez-Valera, F., Ventosa, A., et al. (2013). Metagenome sequencing of prokaryotic microbiota from two hypersaline ponds of a Marine Saltern in Santa Pola, Spain. *Genome Announc.* 1:e00933-13. doi: 10.1128/genomeA.00933-13
- Fernández-Gómez, B., Fernández-Guerra, A., Casamayor, E. O., González, J. M., Pedrós-Alió, C., and Acinas, S. G. (2012). Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* 13:347. doi: 10.1186/1471-2164-13-347
- Fraser, C., Hanage, W. P., and Spratt, B. G. (2007). Recombination and the nature of bacterial speciation. *Science* 315, 476–480. doi: 10.1126/science.1127573
- García-González, A., Vicens, L., Alicea, M., and Massey, S. E. (2013). The distribution of recombination repair genes is linked to information content in bacteria. *Gene* 528, 295–303. doi: 10.1016/j.gene.2013.05.082
- Ghai, R., Hernandez, C. M., Picazo, A., Mizuno, C. M., Ininbergs, K., Díez, B., et al. (2012). Metagenomes of Mediterranean coastal lagoons. *Sci. Rep.* 2:490. doi: 10.1038/srep00490
- Ghai, R., Pašić, L., Fernández, A. B., Martín-Cuadrado, A.-B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135
- Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–582. doi: 10.1093/bioinformatics/16.7.573
- Gonzaga, A., Martín-Cuadrado, A. B., López-Pérez, M., Mizuno, C. M., García-Heredia, I., Kimes, N. E., et al. (2012). Polyclonality of concurrent natural populations of *Alteromonas macleodii*. *Genome Biol. Evol.* 4, 1360–1374. doi: 10.1093/gbe/evs112
- González-Torres, P., Prysycz, L. P., Santos, F., Martínez-García, M., Gabaldón, T., and Antón, J. (2015). Interactions between closely related bacterial strains are revealed by deep transcriptome sequencing. *Appl. Environ. Microbiol.* 81, 8445–8456. doi: 10.1128/AEM.02690-15
- Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., and Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature* 551, 45–50. doi: 10.1038/nature24287
- Göttsch, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176
- Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 8:172. doi: 10.1186/1471-2105-8-172
- Hacker, J., and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep.* 2, 376–381. doi: 10.1093/embo-reports/kve097
- Hacker, J., and Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu. Rev. Microbiol.* 54, 641–679. doi: 10.1146/annurev.micro.54.1.641
- Halpern, D., Chiappello, H., Schbath, S., Robin, S., Hennequet-Antier, C., Gruss, A., et al. (2007). Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genet.* 3, 1614–1621. doi: 10.1371/journal.pgen.0030153
- Handa, N., Ichige, A., and Kobayashi, I. (2009). Contribution of RecFOR machinery of homologous recombination to cell survival after loss of a restriction-modification gene complex. *Microbiology* 155, 2320–2332. doi: 10.1099/mic.0.026401-0
- Held, N. L., Herrera, A., Cadillo-Quiroz, H., and Whitaker, R. J. (2010). CRISPR associated diversity within a population of *Sulfolobus islandicus*. *PLoS One* 5:e12988. doi: 10.1371/journal.pone.0012988
- Hiller, N. L., Ahmed, A., Powell, E., Martin, D. P., Eutsey, R., Earl, J., et al. (2010). Generation of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog.* 6:e1001108. doi: 10.1371/journal.ppat.1001108
- Hogg, J. S., Hu, F. Z., Janto, B., Boissy, R., Hayes, J., Keefe, R., et al. (2007). Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol.* 8:R103. doi: 10.1186/gb-2007-8-6-r103
- Hsiao, W. W. L., Ung, K., Aeschliman, D., Bryan, J., Finlay, B. B., and Brinkman, F. S. L. (2005). Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet.* 1:e62. doi: 10.1371/journal.pgen.0010062
- Ietswaart, R., Szardenings, F., Gerdes, K., and Howard, M. (2014). Competing ParA structures space bacterial plasmids equally over the nucleoid. *PLoS Comput. Biol.* 10:e1004009. doi: 10.1371/journal.pcbi.1004009
- Jeltsch, A. (2003). Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? *Gene* 317, 13–16. doi: 10.1016/S0378-1119(03)00652-8
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Kimes, N. E., López-Pérez, M., Ausó, E., Ghai, R., and Rodriguez-Valera, F. (2014). RNA sequencing provides evidence for functional variability between naturally co-existing *Alteromonas macleodii* lineages. *BMC Genomics* 15:938. doi: 10.1186/1471-2164-15-938

- Konstantinidis, K. T., and DeLong, E. F. (2008). Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J.* 2, 1052–1065. doi: 10.1038/ismej.2008.62
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12
- Lapierre, P., and Gogarten, J. P. (2009). Estimating the size of the bacterial pan-genome. *Trends Genet.* 25, 107–110. doi: 10.1016/j.tig.2008.12.004
- López-Pérez, M., Gonzaga, A., Ivanova, E. P., and Rodríguez-Valera, F. (2014a). Genomes of *Alteromonas australica*, a world apart. *BMC Genomics* 15:483. doi: 10.1186/1471-2164-15-483
- López-Pérez, M., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., Chain, F. J. J., Shapiro, J., and Chang Mell, J. (2014b). Homologous recombination is involved in the diversity of replacement flexible genomic islands in aquatic prokaryotes. *Front. Genet.* 5:147. doi: 10.3389/fgene.2014.00147
- López-Pérez, M., Gonzaga, A., and Rodríguez-Valera, F. (2013). Genomic diversity of ‘deep ecotype’ *Alteromonas macleodii* isolates: evidence for pan-mediterranean clonal frames. *Genome Biol. Evol.* 5, 1220–1232. doi: 10.1093/gbe/evt089
- López-Pérez, M., Ramon-Marco, N., and Rodríguez-Valera, F. (2017). Networking in microbes: conjugative elements and plasmids in the genus *Alteromonas*. *BMC Genomics* 18:36. doi: 10.1186/s12864-016-3461-0
- López-Pérez, M., and Rodríguez-Valera, F. (2016). Pangenome evolution in the marine bacterium *Alteromonas*. *Genome Biol. Evol.* 8, 1556–1570. doi: 10.1093/gbe/evw098
- Luo, C., and Konstantinidis, K. T. (2011). Phosphorus-related gene content is similar in *Prochlorococcus* populations from the North Pacific and North Atlantic Oceans. *Proc. Natl. Acad. Sci. U.S.A.* 108, E62–NaN–E66. doi: 10.1073/pnas.1018662108
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1:18. doi: 10.1186/2047-217X-1-18
- Lythgoe, K. A., and Chao, L. (2003). Mechanisms of coexistence of a bacteria and a bacteriophage in a spatially homogeneous environment. *Ecol. Lett.* 6, 326–334. doi: 10.1046/j.1461-0248.2003.00433.x
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577
- Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, 115–122. doi: 10.1093/nar/gkr1044
- Martin, D., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563. doi: 10.1093/bioinformatics/16.6.562
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D., and Lefeuve, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462–2463. doi: 10.1093/bioinformatics/btq467
- Martin-Cuadrado, A.-B., Pašić, L., and Rodríguez-Valera, F. (2015). Diversity of the cell-wall associated genomic island of the archaeon *Haloquadratum walsbyi*. *BMC Genomics* 16:603. doi: 10.1186/s12864-015-1794-8
- Martínez-García, M., Santos, F., Moreno-Paz, M., Parro, V., and Antón, J. (2014). Unveiling viral–host interactions within the ‘microbial dark matter’. *Nat. Commun.* 5:4542. doi: 10.1038/ncomms5542
- Melville, S., and Craig, L. (2013). Type IV Pili in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 77, 323–341. doi: 10.1128/MMBR.00063-12
- Mira, A., Martín-Cuadrado, A. B., D’Auria, G., and Rodríguez-Valera, F. (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int. Microbiol.* 13, 45–57. doi: 10.2436/20.1501.01.110
- Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 60, 174–182. doi: 10.1007/s00239-004-0046-3
- Mojica, F. J. M., Díez-Villaseñor, C., Soria, E., and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 36, 244–246. doi: 10.1046/j.1365-2958.2000.01838.x
- Mongodin, E. F., Nelson, K. E., Daugherty, S., Deboy, R. T., Wister, J., Khouri, H., et al. (2005). The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18147–18152. doi: 10.1073/pnas.0509073102
- Morelli, G., Didelot, X., Kusecek, B., Schwarz, S., Bahlawane, C., Falush, D., et al. (2010). Microevolution of *Helicobacter pylori* during prolonged infection of single hosts and within families. *PLoS Genet.* 6:e1001036. doi: 10.1371/journal.pgen.1001036
- Morris, J. J., Lenski, R. E., and Zinser, E. R. (2012). The Black Queen Hypothesis: evolution of dependencies through adaptive gene loss. *mBio* 3:e00036-12. doi: 10.1128/mBio.00036-12
- Nolan, M., Tindall, B. J., Pomrenke, H., Lapidus, A., Copeland, A., Glavina Del Rio, T., et al. (2009). Complete genome sequence of *Rhodothermus marinus* type strain (R-10). *Stand. Genomic Sci.* 1, 283–290. doi: 10.4056/sigs.46736
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29
- Oliveira, P. H., Touchon, M., and Rocha, E. P. C. (2016). Regulation of genetic flux between bacteria by restriction–modification systems. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5658–5663. doi: 10.1073/pnas.1603257113
- Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Systems* 4:2. doi: 10.1186/1746-1448-4-2
- Oren, A. (2013). *Salinibacter*: an extremely halophilic bacterium with archaeal properties. *FEMS Microbiol. Lett.* 342, 1–9. doi: 10.1111/1574-6968.12094
- Oren, A., and Mana, L. (2003). Sugar metabolism in the extremely halophilic bacterium *Salinibacter ruber*. *FEMS Microbiol. Lett.* 223, 83–87. doi: 10.1016/S0378-1097(03)00345-8
- Osaka, S., Okuzumi, K., Koide, S., Tamai, K., Sato, T., Tanimoto, K., et al. (2018). Genetic shifts in methicillin-resistant *Staphylococcus aureus* epidemic clones and toxin gene profiles in Japan: comparative analysis among pre-epidemic, epidemic and post-epidemic phases. *J. Med. Microbiol.* 67, 392–399. doi: 10.1099/jmm.0.000687
- Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218–225. doi: 10.1006/viro.1999.0056
- Papke, R. T., Koenig, J. E., Rodríguez-Valera, F., and Doolittle, W. F. (2004). Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929. doi: 10.1126/science.1103289
- Pasić, L., Rodríguez-Mueller, B., Martín-Cuadrado, A.-B., Mira, A., Rohwer, F., and Rodríguez-Valera, F. (2009). Metagenomic islands of hyperhalophiles: the case of *Salinibacter ruber*. *BMC Genomics* 10:570. doi: 10.1186/1471-2164-10-570
- Peña, A., Gomariz, M., Lucio, M., González-Torres, P., Huertas-Cepa, J., Martínez-García, M., et al. (2014). *Salinibacter ruber*: The Never Ending Microdiversity? - *Bspus-Horizonpress.com. Genetics and Genomics of Halophiles*. Poole: Caister Academic Press.
- Peña, A., Teeling, H., Huerta-Cepas, J., Santos, F., Yarza, P., Brito-Echeverría, J., et al. (2010). Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains. *ISME J.* 4, 882–895. doi: 10.1038/ismej.2010.6
- Peña, A., Valens, M., Santos, F., Buczolits, S., Antón, J., Kämpfer, P., et al. (2005). Intraspecific comparative analysis of the species *Salinibacter ruber*. *Extremophiles* 9, 151–161. doi: 10.1007/s00792-005-0430-y
- Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Pinto, U. M., Pappas, K. M., and Winans, S. C. (2012). The ABCs of plasmid replication and segregation. *Nat. Rev. Microbiol.* 10, 755–765. doi: 10.1038/nrmicro2882
- Podell, S., Ugalde, J. A., Narasingarao, P., Banfield, J. F., Heidelberg, K. B., and Allen, E. E. (2013). Assembly-driven community genomics of a hypersaline microbial ecosystem. *PLoS One* 8:e61692. doi: 10.1371/journal.pone.0061692
- Polz, M. F. (2013). “Quantitative and theoretical microbial population biology,” in *The Prokaryotes*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin: Springer).

- Popa, O., Hazkani-covo, E., Landan, G., Martin, W., and Dagan, T. (2011). Directed networks reveal genomic barriers and DNA repair bypasses to lateral gene transfer among prokaryotes. *Genome Res.* 21, 599–609. doi: 10.1101/gr.115592.110.Freely
- Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13757–13762. doi: 10.1073/pnas.241370698
- Pruesse, E., Peplies, J., and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. doi: 10.1093/bioinformatics/bts252
- Reeves, P. R., Liu, B., Zhou, Z., Li, D., Guo, D., Ren, Y., et al. (2011). Rates of mutation and host transmission for an *Escherichia coli* clone over 3 years. *PLoS One* 6:e26907. doi: 10.1371/journal.pone.0026907
- Roberts, A. P., and Kreth, J. (2014). The impact of horizontal gene transfer on the adaptive ability of the human oral microbiome. *Front. Cell. Infect. Microbiol.* 4:124. doi: 10.3389/fcimb.2014.00124
- Rodriguez-Brito, B., Li, L., Wegley, L., Furlan, M., Angly, F., Breitbart, M., et al. (2010). Viral and microbial community dynamics in four aquatic environments. *ISME J.* 4, 739–751. doi: 10.1038/ismej.2010.1
- Rodriguez-Valera, F., Martin-Cuadrado, A.-B., and López-Pérez, M. (2016). Flexible genomic islands as drivers of genome evolution. *Curr. Opin. Microbiol.* 31, 154–160. doi: 10.1016/j.mib.2016.03.014
- Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pašić, L., Thingstad, T. F., Rohwer, F., et al. (2009). Explaining microbial population genomics through phage predation. *Nat. Rev. Microbiol.* 7, 828–836. doi: 10.1038/nrmicro2235
- Rosselló-Mora, R., Lucio, M., Peña, A., Brito-Echeverría, J., López-López, A., Valens-Vadell, M., et al. (2008). Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber*. *ISME J.* 2, 242–253. doi: 10.1038/ismej.2007.93
- Salminen, M. O., Carr, J. K., and Burke, D. S. (1995). Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retroviruses* 11, 1423–1425. doi: 10.1089/aid.1995.11.1423
- Santos, F., Yarza, P., Parro, V., Briones, C., and Anton, J. (2010). The metavirome of a hypersaline environment. *Environ. Microbiol.* 12, 2965–2976. doi: 10.1111/j.1462-2920.2010.02273.x
- Santos, F., Yarza, P., Parro, V., Meseguer, I., Rosselló-Móra, R., and Antón, J. (2012). Culture-independent approaches for studying viruses from hypersaline environments. *Appl. Environ. Microbiol.* 78, 1635–1643. doi: 10.1128/AEM.07175-11
- Schmidt, M. A., Riley, L. W., and Benz, I. (2003). Sweet new world: Glycoproteins in bacterial pathogens. *Trends Microbiol.* 11, 554–561. doi: 10.1016/j.tim.2003.10.004
- Shapiro, B. J., Friedman, J., Cordero, O. X., Preheim, S. P., Timberlake, S. C., Szabo, G., et al. (2012). Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51. doi: 10.1126/science.1218198
- Shiklomanov, I. A. (1998). *World Water Resources: A New Appraisal and Assessment for the 21st Century: A Summary of the Monograph World Water Resources: IRC*. Paris: UNESCO.
- Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129. doi: 10.1007/BF00182389
- Sorek, R., Kunin, V., and Hugenholtz, P. (2008). CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* 6, 181–186. doi: 10.1038/nrmicro1793
- Takeno, M., Taguchi, H., and Akamatsu, T. (2011). Role of ComFA in controlling the DNA uptake rate during transformation of competent *Bacillus subtilis*. *J. Biosci. Bioeng.* 111, 618–623. doi: 10.1016/j.jbiosc.2011.02.006
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* 45, 1320–1328. doi: 10.4319/lo.2000.45.6.1320
- Thingstad, T. F., and Lignell, R. (1997). Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* 13, 19–27. doi: 10.3354/ame013019
- Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi: 10.1038/nrmicro1234
- Tully, B. J., Emerson, J. B., Andrade, K., Brocks, J. J., Allen, E. E., Banfield, J. F., et al. (2015). De novo sequences of *Haloquadratum walsbyi* from Lake Tyrrell, Australia, reveal a variable genomic landscape. *Archaea* 2015:875784. doi: 10.1155/2015/875784
- van Wolferen, M., Ajon, M., Driessen, A. J. M., and Albers, S. V. (2013). How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles* 17, 545–563. doi: 10.1007/s00792-013-0552-6
- Vasu, K., and Nagaraja, V. (2013). Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* 77, 53–72. doi: 10.1128/MMBR.00044-12
- Ventosa, A., Fernández, A. B., León, M. J., Sánchez-Porro, C., and Rodriguez-Valera, F. (2014). The Santa Pola saltern as a model for studying the microbiota of hypersaline environments. *Extremophiles* 18, 811–824. doi: 10.1007/s00792-014-0681-6
- Vergin, K. L., Tripp, H. J., Wilhelm, L. J., Denver, D. R., Rappé, M. S., and Giovannoni, S. J. (2007). High intraspecific recombination rate in a native population of *Candidatus pelagibacter ubique* (SAR11). *Environ. Microbiol.* 9, 2430–2440. doi: 10.1111/j.1462-2920.2007.01361.x
- Villamor, J., Ramos-Barbero, M. D., González-Torres, P., Gabaldón, T., Rosselló-Móra, R., Meseguer, I., et al. (2017). Characterization of ecologically diverse viruses infecting co-occurring strains of cosmopolitan hyperhalophilic Bacteroidetes. *ISME J.* 12, 424–437. doi: 10.1038/ismej.2017.175
- Vos, M., and Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. *ISME J.* 3, 199–208. doi: 10.1038/ismej.2008.93
- Wang, X., and Quinn, P. J. (2010). Lipopolysaccharide: biosynthetic pathway and structure modification. *Prog. Lipid Res.* 49, 97–107. doi: 10.1016/j.plipres.2009.06.002
- Whitaker, R. J. (2005). Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus islandicus*. *Mol. Biol. Evol.* 22, 2354–2361. doi: 10.1093/molbev/msi233
- Yahara, K., Didelot, X., Jolley, K. A., Kobayashi, I., Maiden, M. C. J., Sheppard, S. K., et al. (2016). The landscape of realized homologous recombination in pathogenic bacteria. *Mol. Biol. Evol.* 33, 456–471. doi: 10.1093/molbev/msv237
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* 12, 635–645. doi: 10.1038/nrmicro3330

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

Copyright © 2018 González-Torres and Gabaldón. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.