Chapter 5
From Genomics to Microevolution and Ecology: The Case of *Salinibacter ruber*

Arantxa Peña, Hanno Teeling, Jaime Huerta-Cepas, Fernando Santos, Inmaculada Meseguer, Marianna Lucio, Philippe Schmitt-Kopplin, Joaquín Dopazo, Ramon Rosselló-Móra, Margarete Schüler, Frank Oliver Glöckner, Rudolf Amann, Toni Gabaldón, and Josefa Antón

5.1 Introduction

Salinibacter ruber is an extremely halophilic bacterium of the *Bacteroidetes* phylum that inhabits hypersaline environments, such as crystallizer ponds from solar salterns. Haloarchaea usually dominate the microbial communities in such

A. Peña • F. Santos • J. Antón (✉)
Departamento de Fisiología, Genética y Microbiología, Universidad de Alicante, Apartado 99, Alicante, Spain
e-mail: anton@ua.es

H. Teeling • M. Schüler • R. Amann
Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen, Germany

J. Huerta-Cepas • T. Gabaldón
Bioinformatics and Genomics Program, Centre for Genomics Regulation (CRG), UPF, Dr. Aiguader, 83, 08003 Barcelona, Spain

I. Meseguer
Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández. Avda. de la Universidad s/n., Elche (Alicante), Spain

M. Lucio • P. Schmitt-Kopplin
Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Ecological Chemistry, Neuherberg, Germany

J. Dopazo
Functional Genomics Node (INB), Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain

R. Rosselló-Móra
Marine Microbiology Group, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), C/Miquel Marqués 21, 07190 Esporles, Spain

F.O. Glöckner
Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen, Germany
and
Jacobs University Bremen GmbH, Campusring 1, Bremen, Germany

A. Ventosa et al. (eds.), *Halophiles and Hypersaline Environments*,
DOI 10.1007/978-3-642-20198-1_5,

salterns, while *S. ruber* can account from 2 to 30%. *S. ruber* is globally widespread, but nonetheless its 16S rRNA gene and 16S–23S rRNA internal transcribed spacers (ITS) sequences are highly congruent (see below). At the genomic level, however, this species presents a high level of microdiversity (Peña et al. 2005; Antón et al. 2008). Indeed, when new *S. ruber* isolates are obtained from the same salterns where the species was first retrieved and their digested genomic DNAs are analyzed by pulsed gel field electrophoresis (PFGE), a remarkably high diversity of patterns is obtained. For instance, 35 new strains isolated in 2008 provided as many new patterns, although all had identical 16S rRNA gene sequences. In addition, none of these patterns corresponded to previously retrieved strains.

To date, genome comparisons offer the most powerful way to study microdiversity. This approach allows unveiling the diversity hidden within a species' population as well as providing clues on the ecological mechanisms that drive diversification. Until recently, such studies were for obvious reasons focused on pathogenic bacteria. In the last years, however, microdiversity has also been studied in aquatic prokaryotic species, with a focus on the characterization of ecotypes (i.e., sub-populations within the same species with distinct ecological strategies) (Cohan and Koeppel 2008). Such comparative genome analyses at the sub-species level include the studies on *Prochlorococcus* sp. (Coleman et al. 2006), and *Alteromonas macleodi* isolates (Ivars-Martínez et al. 2008), for which ecotypes have been found that can be distinguished via their 16S rRNA gene(s) or ITS. However, there is a scale of diversity that falls beyond the ecotype level, like clusters above 99% 16S rRNA similarity as described by Acinas et al. (2004) for coastal bacterioplankton. In hypersaline environments, such clusters can be found as well (Antón et al. 2008; Oh et al. 2010).

The situation of *S. ruber* represents a really short scale of differentiation, since in most cases no differences in the whole ribosome operon can even be observed although, systematically, different strains are being continuously isolated from the same environment. From an ecological perspective, the microdiversity in *S. ruber* is really intriguing, in particular in consideration of the few opportunities for microniche differentiation that crystallizer ponds seem to offer.

5.2 Choosing the Strains: M8 and M31

Our objective was to envisage the microdiversity within the species *S. ruber* and understand its meaning. For this purpose, we chose the two most closely related strains in order to elucidate (a) the extent of the genomic differences among these strains, (b) whether the genomic differences have any ecological benefit, and (c) the mechanisms causing microdiversity in the context of the apparent lack of heterogeneity in the crystallizers that *S. ruber* inhabits.

Among all strains within our collection from different parts of the world, M8 and M31 showed the closest relatedness. Both had identical ribosomal operons (including 5S, 16S, 23S, and ITS), a DNA–DNA hybridization value of around 90%, and

had been sampled at the same time in September 1999 on the Mediterranean island Mallorca from a pond of the solar saltern of Campos. On the other hand, a genomic micro-diversity was detectable, since both strains showed distinct genomic PFGE patterns that, like random amplification of polymorphic DNA (RAPD) patterns, clustered with above 90% similarity (Peña et al. 2005). Since strains M8 and M31 were among the five isolates that had been used for the *S. ruber* species description (Antón et al. 2002), and the genome of strain M31^T had already been sequenced by Mongodin et al. (2005), this provided a unique opportunity for a strain level genome comparison study (Peña et al. 2010).

5.3 Differences Between the M8 and M31 Genomes

The first step of the comparative analysis was to sequence the M8 genome (a single chromosome of 3.6 Mbp, and four plasmids ranging from 11.2 to 84.34 kbp) and to analyze features such as GC content, tetranucleotide frequency or codon usage. These parameters turned out not to be homogeneous along the chromosome, since there were two areas (that we called hypervariable regions – HVRs) with deviant GC contents and tetranucleotide frequencies as well as less adapted codon usage (Fig. 5.1). The annotation of the genome indicated a higher density of transposases in these areas as well as genes related to the cell envelope (cell wall, capsules, and outer membrane, including many glycosyltransferases and sulfotransferases).

The genomes of M8 and M31 were compared with different tools. As expected, both genomes were highly similar as it could be observed from the calculations of the average amino acid identity (AAI = 94.2%), average nucleotide identity for orthologous genes (ANIo = 93.5%), and average nucleotide identity (ANIb = 98.5%), but they also displayed some remarkable differences as discussed below. The differences between ANIo and ANIb are basically due to the differences in the thresholds for their calculations. In one side, ANIo indicates the degree of identity among recognized orthologous genes shared between the strains. On the other side, ANIb is a raw indication of the identity of the genome sequences in a similar way to that given by the use of DNA–DNA hybridization assays (Richter and Rosselló-Móra 2009).

With the exception of the HVRs, high levels of gene synteny (gene order and orientation) were observed between both chromosomes (Fig. 5.2). The two HVRs of M8 aligned with two of the three "genomic islands" that were described in M31. In addition, strain-specific genes were identified along the chromosome (represented as dots in the axis of Fig. 5.2). These are genes present only in one of the two strains, and thus part of the accessory genome of the *S. ruber* species, while the remaining genes present in both strains are part of the core genome (Feil 2004). Together, core and accessory genomes form the species' pan-genome (Medini et al. 2005). In total, 325 genes were found in M8 that were absent from M31, which implies that about 10% of the M8 genes were strain-specific. Among these, 164 genes were unique to M8 and constituted new genes absent from any

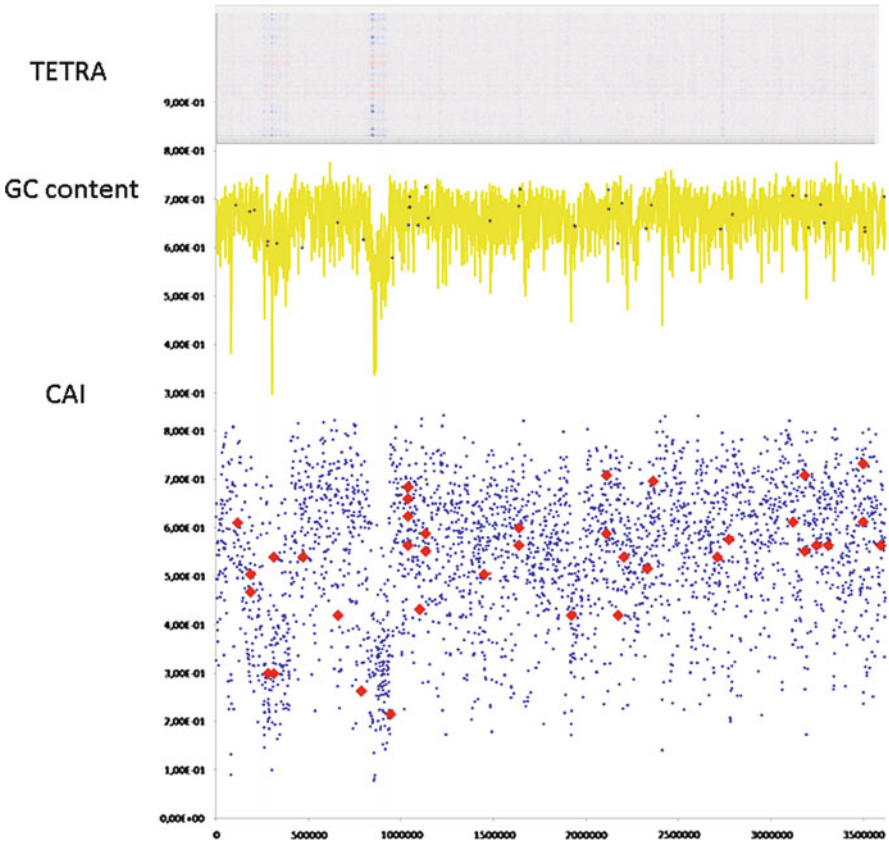

Fig. 5.1 Characteristics of the M8 genome. From *top* to *bottom*: tetranucleotide frequency, GC content and codon adaptation indexes. In *blue* in the GC plot and *red* in the CAI plot, the LGT genes described in Table 5.1. The whole-genome alignment shows the positions of orthologous genes in *S. ruber* M8 and M31. Genes present only in one of the strains appear on the axes. The location of HVR and CR in M8 are marked with *gray bars*. Modified from Peña et al. (2010)

sequence database. It is remarkable that analyzing the genome of a new strain (with a ribosomal operon identical to a known genome) reveals 164 novel genes. This points towards an extensive pan-genome, as discussed at the end of this chapter.

The core genome of *S. ruber* included genes coding for (conserved) hypothetical proteins, that were present only in M8 and M31 but not in the available databases; these genes can be considered as species-specific. Most of the genes shared by both strains displayed high levels of high amino acid identity, but 18% of the shared genes were rather divergent (AAI below 90%) with a few very divergent genes (AAI below 50%). In general, the divergent genes were distributed along the whole chromosome, albeit with a higher density in the HVRs (Peña et al. 2010).

Thus, the types of observed differentiations between M8 and M31 included: the HRVs, insertions and deletions outside of the HRVs, divergence among homologs,

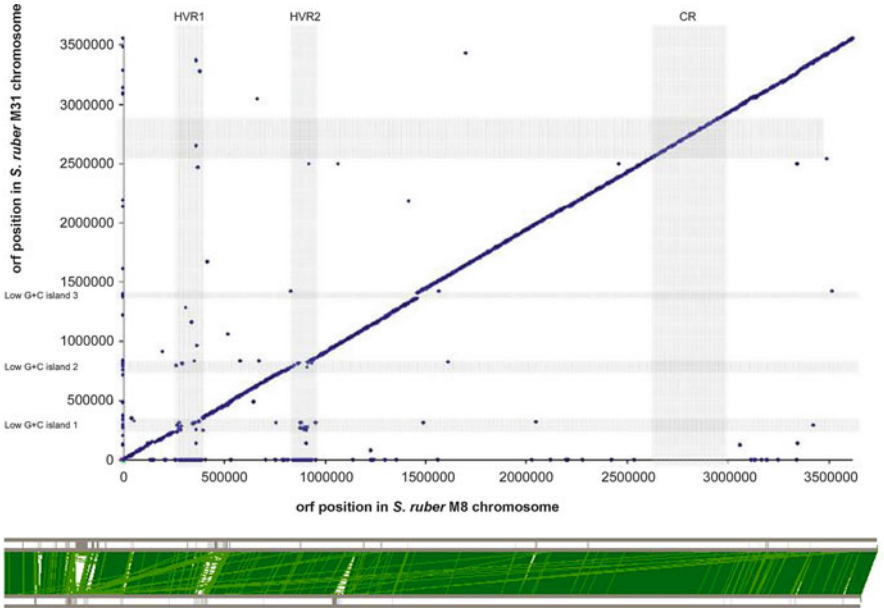

Fig. 5.2 Whole-genome alignment between M8 and M31. The *green bars* linking both genomes represent orthologs matches identified by FASTA analysis. The sequences have been aligned from the predicted replication origin. From Peña et al. (2010), with permission

and specific gene duplications in every genome. Of course, this raised the question whether these differences cause as yet undetected phenotypic strain-level differences. Therefore, we analyzed the strain's metabolomes and we indeed detected differences, mainly in the extracellular fractions, that showed a higher content of sulfonated and glycosylated metabolites in M8 (Fig. 5.3). In addition, we also found that both strains were able to compete to each other: when M8 and M31 were co-cultured in a salt-saturated medium, M31 inhibited the growth of M8 (Peña et al. 2010). Therefore, the differences found between the genomes of both strains were indeed having an effect on their phenotypes and therefore cannot be considered neutral from an ecological point of view.

5.4 The Role of LGT in the Shaping of *S. ruber* Genome

Salinibacter shares its habitat with haloarchaea (family *Halobacteriaceae*). In most, but not all (Oh et al. 2010) of the hypersaline environments analyzed so far, the most abundant prokaryote in the community is the square archaeon *Haloquadratum walsbyi*, while *S. ruber* is the most abundant bacterium. In other cases, *Halorubrum* and *Halobacterium* representatives dominate the community and the gammaproteobacterium *Salicola* spp. dominates the bacterial assemblage.

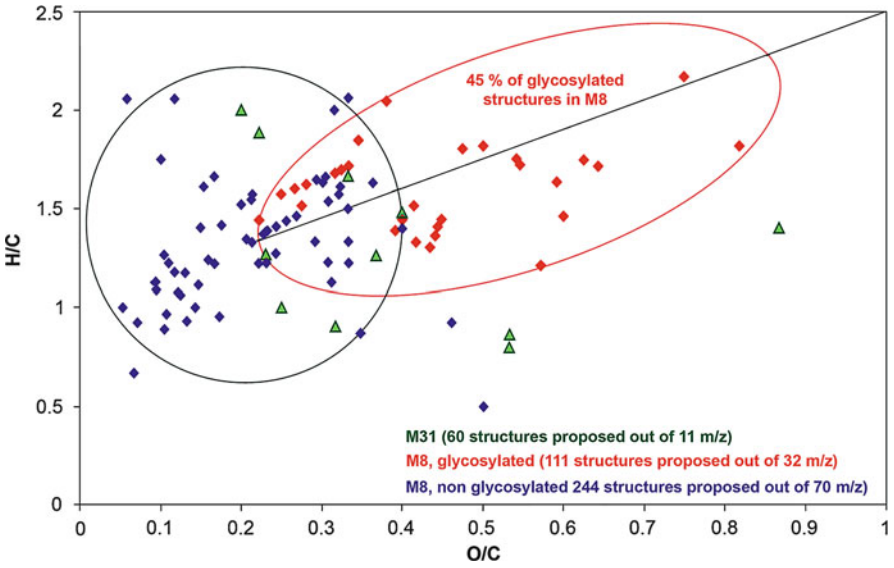

Fig. 5.3 Van Krevelen diagram structural visualization of the discriminative masses only of M8 versus M31 showing the importance of glycosylated structures in M8 as obtained after annotation of the experimental exact masses in the KEGG database (http://www.MassTRIX.org)

Whatever the case, *Salinibacter* always shares its habitat with Archaea, which provides the possibility of lateral gene transfer between these two organisms or, in other words, inter-domain (Archaea-Bacteria) gene transfer. In this context, it is noteworthy that *Salinibacter* and haloarchaea share many phenotypic traits that might be resulting either from convergent evolution or LGT: both are extremely halophilic, aerobic, heterotrophic, pigmented, maintain high intracellular potassium concentrations, have a high genomic GC content (with the exception of *Hqr. walsbyi*), and have retinal proton pumps in their membranes. The sequencing of the *S. ruber* type strain (Mongodin et al. 2005) revealed that the extent of archaeal genes in the *Salinibacter* genome was significant, albeit lower than anticipated. In our work (Peña et al. 2010), we used different approaches to identifying genes involved in inter-domain LGT in *Salinibacter* genomes and reached the same conclusion. In addition, we investigated whether these genes (hereafter referred to as LGT or archaeal genes) had been recently incorporated into the *S. ruber* genome, assuming that they were of pristine archaeal origin and not bacterial genes once transferred to archaeal genomes and now mimicking an archaeal origin, and whether these LGT genes were constitutively expressed and thus had become part of the *S. ruber* core genome.

Most of the putative LGT genes found in the M8 genome (Table 5.1) were located outside the hypervariable regions and did not display anomalous GC contents or unusual codon usage (Fig. 5.1), which would have been expected for recent acquisitions. In addition, only six of the 40 LGT genes in M8 were part of the accessory genome (i.e., not present in M31). Thus, assuming that the identified LGT

Table 5.1 Analysis of the expression of genes candidate of inter-domain LGT along the growth curve

ORF	Annotation	Closest relative[a]	Expression[b] M8/M31
94	HTR-like protein	*Halobacterium salinarum* NRC-1	++++/++++
150	Putative light-and oxygen-sensing transcription regulator	*Halobacterium salinarum* NRC-1	−+++/−++−
169	Adaptive-response sensory-kinase	*Haloarcula marismortui* DSM 3752	+++−/++−−
226	**Conserved hypothetical protein containing PIN domain (putative nucleic acid-binding protein)**	***Haloarcula marismortui* DSM 3752**	**++++**/NA
227	**Conserved hypothetical protein containing DUF0175 domain**	***Haloarcula marismortui* DSM 3752**	**++++/++++**
274	**Conserved hypothetical protein**	***Haloarcula marismortui* DSM 3752**	**−−−−/NA**
393	Cytochrome c oxidase subunit II	*Haloarcula marismortui* DSM 3752	++++/++++
538	Deoxycytidine triphosphate deaminase	*Halobacterium salinarum* NRC-1	−++−/−−+−
655	**Glycosyl transferase, group 1**	***Haloarcula marismortui* DSM 3752**	**−−−−/NA**
831	**Conserved hypothetical protein containing PIN domain (putative nucleic acid-binding protein)**	***Pyrococcus kodakaraensis* OT3**	**−−−−/NA**
911	**Probable exodeoxyribonuclease VII small subunit**	***Haloarcula marismortui* DSM 3752**	**++++/NA**
917	Ferredoxin	*Natronomonas pharaonis* DSM 2160	++++/−+++
918	Conserved hypothetical protein, membrane	*Natronomonas pharaonis* DSM 2160	+++−/++−−
919	Bacteriorhodopsin related protein	*Natronomonas pharaonis* DSM 2160	++−+/−−−−
920	Phytoene dehydrogenase	*Haloarcula marismortui* DSM 3752	++++/++++
964	Probable cell division protein	*Methanobacterium thermoautotrophicum* strain Delta H	++++/++++
995	Translation initiation factor eIF-2B alpha subunit	*Haloarcula marismortui* DSM 3752	++++/++++
1011	Conserved hypothetical protein, secreted	*Haloarcula marismortui* DSM 3752	++++/+−−−
1308	**Uncharacterized ACR**	***Archaeoglobus fulgidus* DSM 4304**	**+++−/NA**
1444	Trk system potassium uptake protein trkA	*Haloarcula marismortui* DSM 3752	++++/−++−
1448	D-lactate dehydrogenase, putative	*Natronomonas pharaonis* DSM 2160	−−−−/−−−−
1707	Conserved hypothetical protein	*Haloarcula marismortui* DSM 3752	++++/−−++
1709	Putative zinc-binding dehydrogenase	*Halobacterium salinarum* NRC-1	+−++/−++−
1859	Anion permease	*Halobacterium salinarum* NRC-1	−+++/++++
1860	Phosphate transporter	*Halobacterium salinarum* NRC-1	++++/−−+−
1903	Conserved hypothetical protein	*Haloarcula marismortui* DSM 3752	+−+−/−−+−
1924	Multi antimicrobial extrusion protein MatE	*Haloarcula marismortui* DSM 3752	−−−−/−−−−
2026	Sodium/calcium exchanger protein, membrane	*Natronomonas pharaonis* DSM 2160	+++−/+−+−
2049	Conserved hypothetical protein containing UPF0047	*Methanospirillum hungatei* DSM 864	−−−−/−−−−
2365	Putative bacterio-opsin	*Haloarcula marismortui* DSM 3752	++++/++++
2415	Heme exporter protein C	*Pyrobaculum aerophilum* DSM 7523	++−−/−−−−
2493	Trk potassium uptake system protein	*Haloarcula marismortui* DSM 3752	++++/++++
2657	Glycerol-3-phosphate dehydrogenase subunit A	*Haloarcula marismortui* DSM 3752	+++−/++++
2720	Conserved hypothetical protein, membrane	*Methanothermobacter thermautotrophicus* ΔH	++++/++++
2733	Sensory rhodopsin I (SR-I)	*Haloarcula marismortui* DSM 3752	+++−/+++−

(*continued*)

Table 5.1 (continued)

ORF	Annotation	Closest relative[a]	Expression[b] M8/M31
2780	Putative small solutes transporter	*Halobacterium salinarum* NRC-1	+++−/+++−
2798	Sensory rhodopsin I (SR-I)	*Natronomonas pharaonis* DSM 2160	++++/++++
2993	Sodium:proline symporter	*Haloarcula marismortui* DSM 3752	+−−−/−−−+
2994	Halorhodopsin	*Natronomonas pharaonis* DSM 2160	++++/++++
3079	Conserved hypothetical protein	*Natronomonas pharaonis* DSM 2160	++++/++++

Modified from Peña et al. (2010)
In boldface, LTG candidates specific of M8
Shaded in light gray, genes in HVRs; in dark gray and white font, genes in CR
[a]Closest relative according to phylogenetic trees
[b]Expression of ORFs along the growth curve as measured by RT-PCR amplification with specific primers. For every culture, 4 points were analyzed corresponding to the early, mid and late exponential phase and to stationary phase. For every point + or − is indicated. *NA* not analyzed (these ORFs were present only in M8 genome and could thus not be analyzed in M31)

genes were acquired from archaeal genomes, this interchange should have happened before the separation of the M8 and M31 lineages. This is also in agreement with the LGT gene's average to high CAI and standard GC values. Further analyses (Peña et al., manuscript in preparation) indicated that most of the LGT genes identified in M8 are also present in strains from different geographical sources. Thus, although modest, LGT has played a role in shaping the *Salinibacter* genome.

In order to see whether these 40 LGT genes were needed for the normal growth of *S. ruber*, their expression together with 8 of the 13 LGT genes present only in M31 was analyzed along the growth curve. Pure cultures of M8 and M31 were grown under standard conditions and samples were taken in early, mid, and late exponential as well as in the stationary phase. After RNA extraction, the expression of each LGT gene was detected by RT-PCR with primers specifically designed for every gene (Peña et al. 2010).

Both strains showed different LGT gene expression patterns, with a large majority of the core genome LGT genes expressed at some point in the growth curve, and a much lower proportion of expressed strain-specific LGT genes (Table 5.1; Fig. 5.4). These data indicate that most of the core genome inter-domain LGT candidates are expressed during normal growth and most likely represent functional genes needed for growth. This is in agreement with the CAI and GC traits discussed above. On the contrary, the strain-specific inter-domain LGT candidates are seldom expressed which, together with their GC and CAI characteristics, indicate that they have probably been acquired recently and did not have time to amend to their genomic environments. In any case, they do not seem necessary for growth under standard conditions.

Although beyond the goal of these experiments, the results shown in Table 5.1 deserve additional comments. The cultures were grown in the dark; however, the light-dependent retinal binding protein halorhodopsin and bacteriorhodopsin (Oren 2002) were constitutively expressed for both strains along the entire growth curve. Sensory rhodopsin (Oren 2002) was also expressed by both strains during the complete exponential phase. This is intriguing, since the expression of these

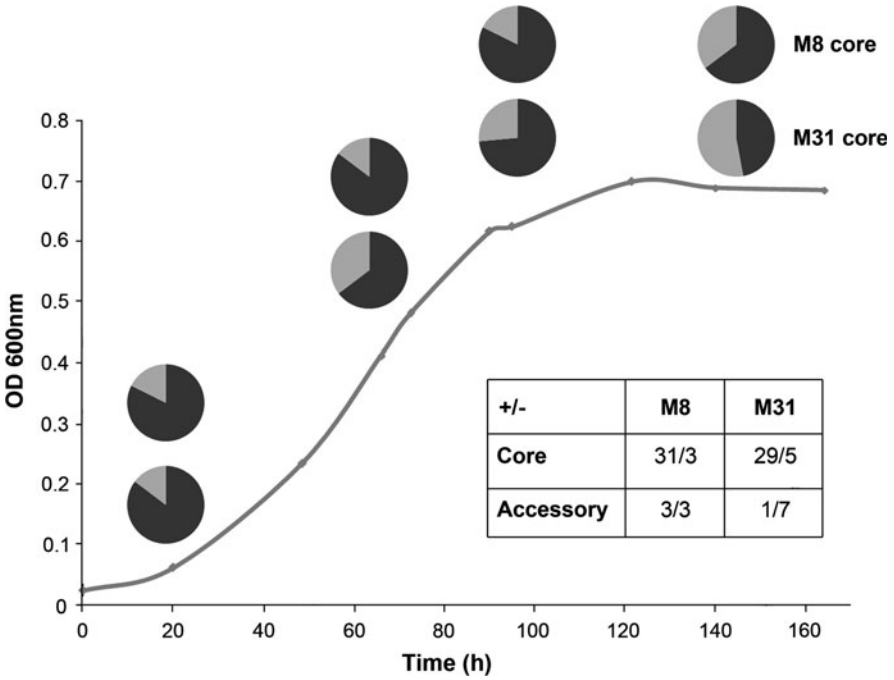

Fig. 5.4 Expression of candidate LGT genes along the growth curve. Pure cultures of M8 and M31 were set up and samples taken at four points along the growth. For every point, the proportion of core-genome LGT expressed genes for M8 and M31 (*dark gray*) are represented in *circles*. The insert shows the number of genes whose expression could (+) or could not be detected (−)

proteins should be light-dependent. Further experiments are needed to address this question and find the physiological role of these proteins, for which the activity has not been demonstrated in *S. ruber*.

5.5 *Salinibacter* and the Phages

As discussed above, many of the M8-specific genes in the HVRs were related to cell-surface properties that could be involved in phage recognition and evasion. This was also the case for the three islands in M31 (Mongodin et al. 2005). In a recent work, Pašić et al. (2009) compared the M31 genome to metagenomic fragments recovered from climax saltern crystallizers with 454 pyrosequencing technology. The three M31 islands were only scarcely represented in the metagenome and thus appear to vary among co-occurring *S. ruber* cells. Furthermore, the islands showed evidence of extensive genomic corruption with atypically low GC contents, low coding densities, and high numbers of pseudogenes and short hypothetical proteins. The annotable proteins in the islands were largely involved in

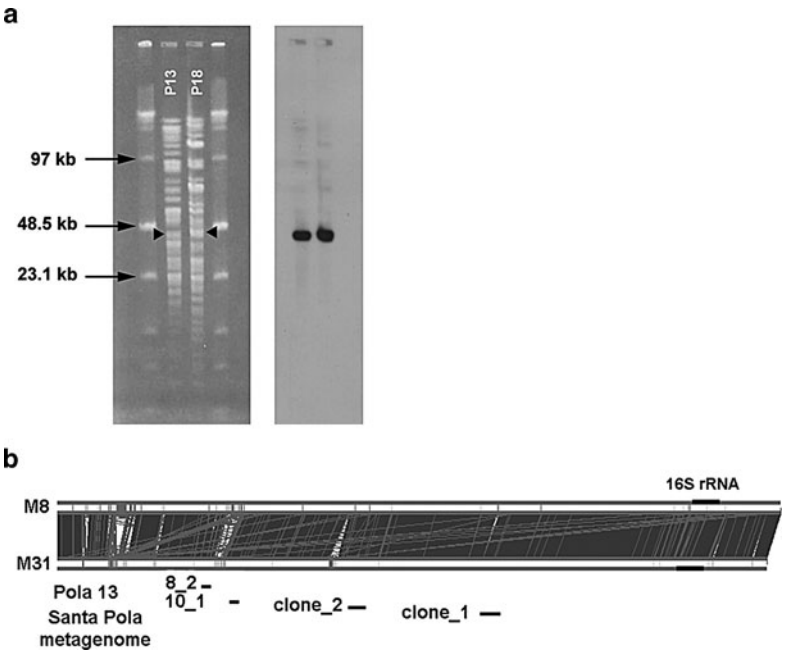

Fig. 5.5 (**a**) *Xba*I digestion of agarose embedded genomic DNA from *S. ruber* strains Pola 13 (P13) and Pola 18 (P18) (*left*). *Red arrows* point to the bands that were excised from the gel and cloned in fosmids. These bands contained the 16S rRNA gene, as hybridization with a digoxigenin-labeled probe (*right*) indicated. (**b**) Alignment of the P13 and Santa Pola metagenomics clones against M8 and M31 whole genomes. The position of the 16S rRNA gene in the genome of both microorganisms is marked in *blue*

variable cell surface characteristics. The authors proposed that "this variation... probably reflects a global strategy of bacteria to escape phage predation."

Metabolome analyses also showed differences in the exposed components of M8 and M31. Therefore, it seems reasonable to expect different phage susceptibilities for the two strains. In order to check whether this was indeed the case, we tried to isolate phages infecting M8 and M31 using different salterns waters as a source for viruses. When cultures of either strain were infected with water from the crystallizer of their isolation, we could easily isolate phages infecting M31 but not a single plaque was observed for M8. Then, we used waters from two ponds with salinities of 23.2 and 34.2% from the Santa Pola saltern (near Alicante, Spain) as source of viruses. For the low salinity water we could get a similar amount of phages for M8 and M31 ($4.47 \times 10^2 \pm 61.1$ $6.33 \times 10^2 \pm 65.1$ PFU per ml of natural sample, respectively). The high salinity water yielded a two-order magnitude increase for M8, while no plaques were formed when M31 was infected under the assayed conditions. These data indicate that, as indicated by their genomic differences, M8 and M31 indeed have different phage infection susceptibilities. This can have a profound impact on the ecology of both strains, if we consider that crystallizer ponds from solar salterns show one of the highest numbers of viruses reported for

planktonic systems with values of up to 2×10^9 virus-like particles per ml. In addition, in the salterns were *Salinibacter* thrives, no bacteriovory has been detected at salinities above 25% (Guixa-Boixareu et al. 1996), which leaves viruses as a major factor for mortality. Therefore, compounds mediating interaction with phages (normally associated with cellular envelopes) should be under strong selection. This has been suggested also for the extremely halophilic archaeon *Hqr. walsbyi* (Cuadros-Orellana et al. 2007), which is present at very high abundances in the Santa Pola and Mallorca salterns. Indeed, metagenomic analyses (Santos et al. 2010, 2011) indicate that the most abundant viruses are those infecting this particular prokaryotic group.

Other prokaryotes exposed to high viral predation, such as the marine bacteria *Prochlorococcus* (Coleman et al. 2006) or "Pelagibacter" representatives (Wilhelm et al. 2007), also harbor genomic island(s) or hypervariable regions enriched in surface-related proteins. For these reasons, phages have been proposed as "drivers" of micro-diversification within prokaryote species (Rodríguez-Valera et al. 2009), not only because of their direct selection on lineages carrying specific sets of exposed cellular components, but also because of their role as gene transfer vehicles.

5.6 The Pangenome Grows

In addition to the two strains M8 and M31 from Mallorca salterns, two additional strains, Pola 13 and Pola 18 (isolated by Prof. Aharon Oren from the Santa Pola salterns) were also used for the *S. ruber* species description. To get insights into the genomic differences of Pola and Mallorca strains, we cloned and sequenced PFGE fragments containing the 16S rRNA genes of the strain Pola 13. We identified a genomic fragment containing the marker gene (Fig. 5.4a), cloned it into a fosmid vector, then sub-cloned into plasmids. Subsequent sequencing and assembly (Peña et al., unpublished results) yielded 18 contigs covering 25 kb that were aligned with the M8 and M31 genomes (Fig. 5.4b). The annotation of these contigs indicated that 19 out of the 37 predicted ORFs encoded hypothetical proteins (and thus had no homologs with M8 and M31); in addition only two ORFs (annotated as pyridoxamine phosphate oxidase and acriflavine resistance protein, respectively) shared homology with respective M8 and M31 genes. The rest of the annotated genes (with homologs in sequence databases) did not align with the M8 and M31 genomes. Thus, the analysis of only 25 kb of a new strain provided 19 completely new genes and expanded the pangenome by no less than 35 genes.

Finally, to get further insights into the extent of *S. ruber* genomic diversity, we performed a preliminary characterization of metagenomic clones containing 16S rRNA gene sequences of this bacterium (Peña et al., unpublished results). For this purpose, we constructed a metagenomic fosmid library from DNA directly retrieved from a crystallizer pond of the Santa Pola saltern from which the original Pola 13 and Pola 18 strains were isolated. In this library of 2,000 clones we could

detect three clones containing the 16S marker gene. End-sequencing of the fosmids provided again an extension of *S. ruber* pangenome, since only two of the six fosmid ends were homologous to the M8 and M31 genomes, while the four remaining ones were novel and hence part of the accessory genome. These results indicate that in spite of the homogeneity between strains isolated from Santa Pola and Mallorca with respect to the ribosomal operon, their genome-level diversity is indeed very high. This corroborates the implications of the ecological differences found between M8 and M31, such as the possibility of competition, their different metabolomes and phage susceptibilities. We must keep in mind that the crystallizers in which *Salinibacter* thrives are environments with very low diversity in terms of species, but it seems that the genomic microdiversity is very high. This is the case not only for *Salinibacter*, but also for the square archaeon *Hqr. walsbyi*, which also has been shown to have a broad pangenome (Ivars-Martínez et al. 2008)

5.7 Open Questions and Future Research

The comparative study of *S. ruber* genomes in conjunction with the wet laboratory experiments summarized above have provided novel first insights into the evolution and ecology of this bacterium, and at the same time raised a plethora of further questions. Some of these questions reach far beyond studying a dedicated genus of extremophilic bacteria. We selected the following points as focus of our future work:

- Intra-specific genomic diversity. How many different genomes of *S. ruber* (or any other single species) are present in a given environment? We have isolated 35 new strains from 1 ml of crystallizer water and found 35 different genomic patterns. Sequencing of only 25 kb of a different strain yields an expansion of the pangenome by 35 genes. How many different genomes could we find at a given time? Would the genome pattern change with time for a given environment?
- The pangenome and the environment. When we compare the genomes of two extremely closely related *S. ruber* strains (M8 and M31), we find that 10% of the genes are strain-specific, which is a rather high number for such close relatives. This is also the case with other microorganisms. Which is then the size of the pangenome for these organisms? Is there a pangenome specific of every type of environment? Which is the relationship between the gene pool of an environment and the microdiversity and the "species" richness?
- Interaction between viruses and prokaryotes. Considering that closely related strains of *S. ruber* vary with respect to phage susceptibility, how do the viruses affect *Salinibacter* populations in the environment? Could the phages be mediating lateral gene transfers between *Salinibacter* and halophilic Archaea? Are the phages infecting *Salinibacter* promoting genotype recycling in the community?

As an overarching goal, we would like to unravel the units of selection in the evolution of microbial communities (strains, phage-host systems, assemblages of strains from the same species...). This may be answered for some lab model systems, but there does not seem to be a consensus for microorganisms in natural communities.

Acknowledgment This work was supported by the Spanish Ministry of Science and Innovation grants CGL2006-12714-CO2-01 and 02, and CGL2009-12651-C02-01 and 0–2 (to JA and RRM) that include European Funds (FEDER), and by the grant ACOM09/155 from the Generalitat Valenciana.

References

Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF (2004) Fine-scale phylogenetic architecture of a complex bacterial community. Nature 430:551–554

Antón J, Oren A, Benlloch S, Rodríguez-Valera F, Amann R, Rosselló-Mora R (2002) *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the Bacteria from saltern crystallizer ponds. Int J Syst Evol Microbiol 52:485–491

Antón J, Peña A, Santos F, Martínez-García M, Schmitt-Kopplin P, Rosselló-Mora R (2008) Distribution, abundance and diversity of the extremely halophilic bacterium *Salinibacter ruber*. Saline Systems 4:15

Cohan FM, Koeppel AF (2008) The origins of ecological diversity in prokaryotes. Curr Biol 18: R1024–R1034

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, Chisholm SW (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. Science 311:1768–1770

Cuadros-Orellana S, Martín-Cuadrado AB, Legault B, D'Auria G, Zhaxybayeva O, Papke RT, Rodríguez-Valera F (2007) Genomic plasticity in prokaryotes: the case of the square haloarchaeon. ISME J 1:235–245

Feil EJ (2004) Small change: keeping pace with microevolution. Nat Rev Microbiol 2:483–495

Guixa-Boixareu N, Caledrón-Paz JI, Heldal M, Bratbak G, Pedrós-Alió C (1996) Viral lysis and bacterivory as prokaryotic loss factors along a salinity gradient. Aquat Microb Ecol 11: 215–227

Ivars-Martínez E, Martín-Cuadrado AB, D'Auria G, Mira A, Ferriera S, Johnson J, Friedman R, Rodríguez-Valera F (2008) Comparative genomics of two ecotypes of the marine planktonic copiotroph *Alteromonas macleodii* suggests alternative lifestyles associated with different kinds of particulate organic matter. ISME J 2:1194–1212

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pangenome. Curr Opin Genet Dev 15:589–594

Mongodin EF, Nelson KE, Daugherty S, DeBoy RT, Wister J, Khouri H, Weidman J, Walsh DA, Papke RT, Sánchez-Pérez G, Sharma AK, Nesbø CL, MacLeod D, Baptese E, Doolittle WF, Charlebois RL, Legault B, Rodríguez-Valera F (2005) The genome of *Salinibacter ruber*. Convergence and gene exchange among hyperhalophilic Bacteria and Archaea. Proc Natl Acad Sci USA 102:18147–18152

Oh D, Porter K, Russ B, Burns D, Dyall-Smith M (2010) Diversity of *Haloquadratum* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. Extremophiles 14:161–169

Oren A (2002) Halophilic microorganisms and their environments. Kluwer Academic Publishers, Dordrecht

Pašić L, Rodríguez-Mueller B, Martín-Cuadraro AB, Mira A, Rohwer F, Rodríguez-Valera F (2009) Metagenomic islands of halophiles: the case of *Salinibacter ruber*. BMC Genomics 10:570

Peña A, Valens M, Santos F, Buczolits S, Antón J, Kämpfer P, Busse HJ, Amann R, Rosselló-Mora R (2005) Intraspecific comparative analysis of the species *Salinibacter ruber*. Extremophiles 9:151–161

Peña A, Teeling H, Huerta-Cepas J, Santos F, Yarza P, Brito-Echevarría J, Lucio M, Schmitt-Kopplin P, Meseguer I, Schenowitz C, Dossat C, Barbe V, Dopazo J, Rosselló-Mora R, Schüler M, Glöckner FO, Amann R, Gabaldón T, Antón J (2010) Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains. ISME J 4:882–895

Richter M, Rosselló-Móra R (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA 106:19126–19131

Rodríguez-Valera F, Martín-Cuadrado AB, Rodríguez-Brito B, Pašić L, Thingstad TF, Rohwer F, Mira A (2009) Explaining microbial population genomics through phage predation. Nat Rev Microbiol 7:828–836

Santos F, Yarza P, Briones C, Parro V, Antón J (2010) The metavirome of a hypersaline environment. Environ Microbiol 12:2965–2976

Santos F, Moreno M, Meseguer I, López C, Rosselló-Mora R, Parro V, Antón J (2011) Meta-transcriptomic analysis of extremely halophilic viral communities. ISME J (submitted for publication) doi:10.1038/ismej.2011.34

Wilhelm LJ, Tripp HJ, Givan SC, Smith DP, Giovannoni SJ (2007) Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. Biol Direct 2:19