Chapter 12 from:

Acidophiles

Life in Extremely Acidic Environments

Edited by Raquel Quatrini and D. Barrie Johnson

ISBN: 978-1-910190-33-3 (hardback)
ISBN: 978-1-910190-34-0 (ebook)

Caister Academic Press https://www.caister.com/acidophiles

The Flexible Genome of Acidophilic Prokaryotes

12

Raquel Quatrini, Francisco J. Ossandon and Douglas Rawlings

Abstract

Over the last couple of decades there has been considerable progress in the identification and understanding of the mobile genetic elements that are exchanged between microbes in extremely acidic environments, and of the genes piggybacking on them. Numerous plasmid families, unique viruses of bizarre morphologies and life cycles, as well as plasmid-virus chimeras, have been isolated from acidophiles and characterized to varying degrees. Growing evidence provided by omic-studies have shown that the mobile elements repertoire is not restricted to plasmids and viruses, but that a plethora of integrative elements ranging from miniature inverted repeat transposable elements to large integrative conjugative elements populate the genomes of acidophilic bacteria and archaea. This chapter reviews the diversity of elements that have been found to constitute the flexible genome of acidophiles. Special emphasis is put on the knowledge generated for *Sulfolobus* (archaea) and species of the bacterial genera *Acidithiobacillus* and *Leptospirillum*. Also, recent knowledge on the strategies used by acidophiles to contain deleterious exchanges while allowing innovation, and the emerging details of the molecular biology of these systems, are discussed. Major lacunae in our understanding of the mobilome of acidophilic prokaryotes and topics for further investigations are identified.

Introduction

Bacteria and archaea have evolved efficient genome modification mechanisms that allow them to adapt rapidly and effectively to ever-changing environmental conditions and to colonize a plethora of ecological niches. Some of these mechanisms are universal (e.g. transposition), while others are significantly more frequent in the prokaryotic world (e.g. transduction, conjugation). Prokaryotic genomes are modified by intrareplicon, interreplicon and interorganismal gene movement of pre-evolved traits, occasionally circumventing the species barrier. These mechanisms are collectively known as horizontal gene transfer (HGT).

Scientific literature on the adaptive role of HGT in prokaryotes is extremely rich; for instance, the acquisition of resistance to harmful compounds and elements or the acquisition of metabolic capabilities allowing the use of new resources are frequent events in all well-studied ecosystems. Extensive HGT between lineages that occupy similar ecological niches, which may drive radical changes of lifestyle, have also been documented. HGT seems to have contributed to the adaptation of *Thermoplasmatales* archaea to extremely acidic and hot environments. *Thermoplasma acidophilum* and *Picrophilus torridus* share approximately the same fraction of genes (65% of their proteomes) between them, as they do with *Sulfolobus solfataricus*, a very distant relative that often inhabits the same environment, whereas only 35% of their genes are present in *Pyrococcus furiosus*, a closer relative (Fütterer *et al.*, 2004). Beyond its role in prokaryotic genome innovation and evolution, HGT also influences dramatically the function of complex prokaryotic communities.

Gene flow between and within microbes is driven by a vast repertoire of Mobile Genetic Elements (MGEs). MGEs are mostly DNA-based platforms that encode enzymes and other proteins that mediate or facilitate self movement and that support different genetic cargoes. The full set of MGEs of a microorganism and the genes they carry

constitute the mobilome or the flexible genomic compartment. Research in microbial MGEs has progressed rapidly over the last couple of decades. Some MGEs are mainly intrachromosomal (e.g. transposons) and others extrachromosomal or episomal (e.g. plasmids), and yet others exist in a fluctuating equilibrium between an integrated and an episomal state (e.g. temperate viruses). MGEs are classified depending on the molecular mechanism underlying their physical movement. Translocative elements, including transposons and insertion sequences, depend on transposases which catalyse 'cut and paste' or 'copy–paste' movements of DNA. Integrative elements relying on site-specific recombination of DNA mediated by a diverse family of enzymes called integrases (or recombinases), include integrons, temperate viruses and a wide diversity of chimeric elements. Other replicative and non-replicative elements which can move between microorganisms, such as plasmids and genomic islands, are reckoned as dispersive or transmissible elements and are further classified as conjugative (those that code all components needed for transfer) or mobilizable according to their transfer ability.

Thus, the mobilome is an assortment of selfish MGEs (Fig. 12.1) composed of many different types of elements, and accretion of individual MGEs (e.g. transposon) on higher-order mobile elements (e.g. a conjugative plasmid). Several chimeric elements such as transposable proviruses, integrative plasmids and mobile integrons muddle the picture considerably and favour the view of MGEs as mosaics of functional gene modules of diverse evolutionary origins.

Lack of informative classification systems and adequate tools for MGE prediction in sequenced genomes has hampered analysis of the prokaryotic mobilome. Both impediments have began to be addressed in recent years. Improved classification criteria have began to emerge for certain types of MGEs. Such is the case of proteobacterial plasmids, for which a relatively comprehensive picture of the diversity, types and mechanisms has started to emerge (Garcillán-Barcia *et al.*, 2013). Also, several MGE prediction strategies, tools and pipelines have been developed and made publicly available in recent years. Although comprehensive genomics of the prokaryotic 'mobilome' still lags behind most other genomics initiatives, sequencing efforts and comparative genomics and metagenomics studies have revealed that the different types of MGEs are found in all branches of the prokaryotic tree of life, and that they occur at varying frequencies, even in the most extreme environments. From these studies new relevant definitions have emerged, for example (i) core genome, which defines the genes present in all strains of a prokaryotic species; (ii) flexible (or dispensable) genome, which concerns genes that are present in some, but not all, strains of a species; (iii) pan genome, which entails the sum of the first two, and (iv) supergenome (or communal gene pool), which consists of all the genes encoded on MGEs and that are readily available to all permissive prokaryotes within a particular setting. This has signified a paradigm shift in that the individual gene pool is now thought to represent a vast resource for MGEs to exploit, enabling the sharing of beneficial traits among prokaryotic populations restricted only by the host ranges of the resident MGEs and the selection pressures exerted by the surrounding environment. Stable appropriation of foreign genetic material by different hosts is determined by a number of factors and is less likely as the phylogenetic distances increase. In the case of plasmids, the host-range is influenced by factors affecting their capacity of transference and replication within their new hosts. In the case of viruses, it is mainly determined by specific host attachment sites and cellular factors necessary for viral multiplication. In addition, a number of molecular mechanisms encoded by prokaryotes that counteract uptake and stabilization of foreign DNA molecules serve to further limit MGE host ranges: these include CRISPR (clustered regularly interspaced short palindromic repeats)-Cas and RM (restriction–modification) systems.

Given that the mobilome and its genetic cargo seem, in many cases, tightly linked to the environment of the host population, it is thought to vary considerably from one environment to another. In this chapter, current knowledge gathered on MGEs that favour HGT in extreme acidic econiches and the strategies used by acidophiles to contain their dispersal, are explored. Knowledge on the mobilome of acidophiles has been organized on the basis of the type of element and covers both acidophilic bacteria and archaea. Longstanding lacunae in our understanding of the of the mobilome of acidophilic prokaryotes are highlighted.

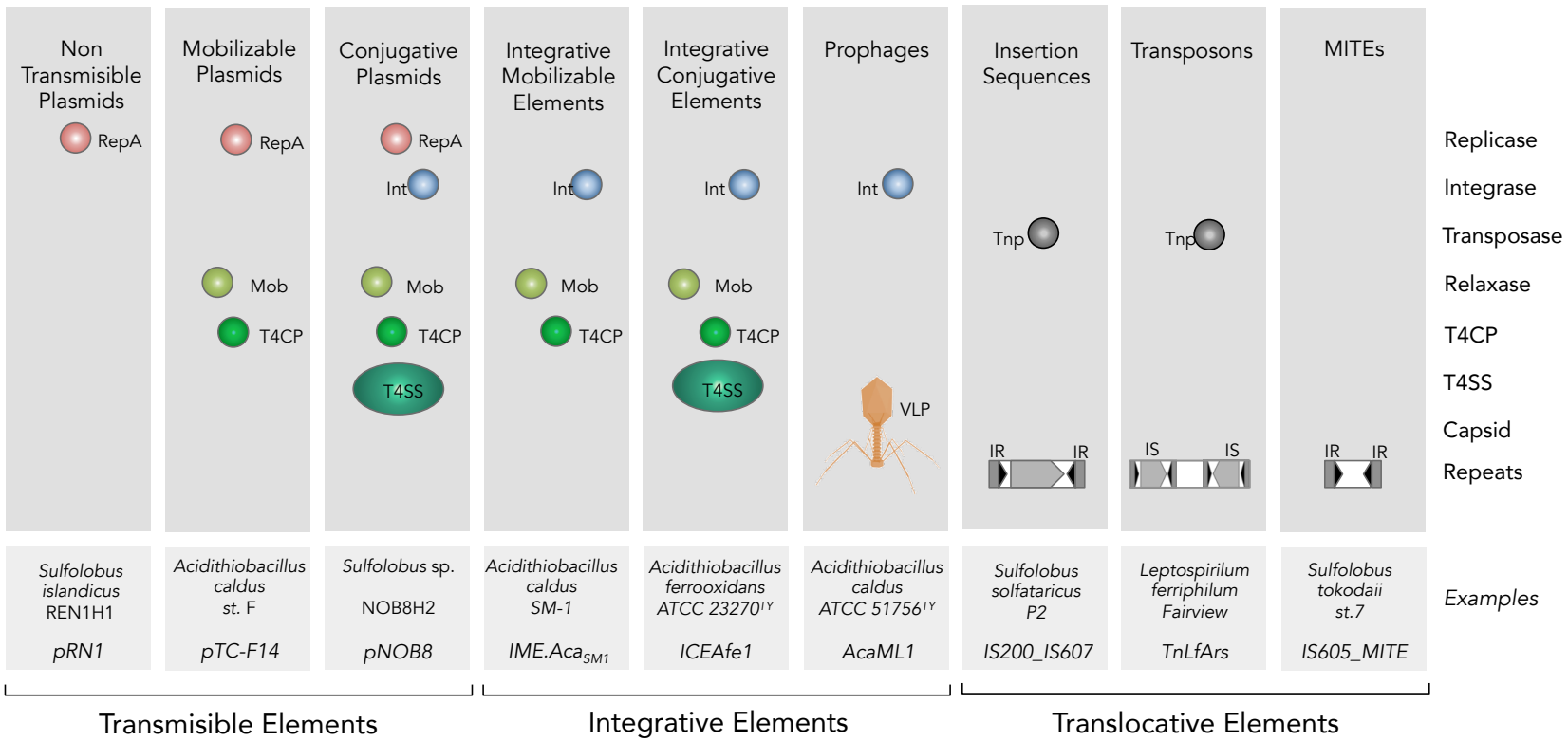

Figure 12.1 Diversity of elements in the mobilome of acidophiles.

Transmissible elements

Plasmids are self-replicating, extrachromosomal, transmissible replicons that act as key agents of change in microbial populations. They range from a few kilobase (kb) to few hundred kb replicons, termed megaplasmids. Although more than a thousand plasmids from all three domains of life and almost every conceivable environmental niche have been described, only 70 plasmids from nearly 160 acidophiles are known by means of their sequence or description (Table 12.1).

Although the genome sequences of an increasing number of plasmids from microbial acidophiles have been determined, knowledge on their biology, biochemistry and molecular mechanisms lags considerably far behind. Insights into the replication, mobilization/conjugation and regulation of some of these plasmids has been gained, yet in most cases they remain cryptic from a functional perspective, and their impact on genome plasticity remains largely unexplored. Knowledge generated for well-characterized plasmid families is covered herein, including archaeal conjugative plasmids, plasmid–virus hybrids and bacterial mobilizable and cryptic plasmids. Replicons of acidophiles, which based on size criteria fall within the category of megaplasmids (larger than 100 kb) have been omitted from this chapter, since no thorough study of these has been reported to date.

Small cryptic plasmids

Some acidophilic archaea and bacteria possess small plasmids of less than 10 kb, considered to be cryptic in nature. Examples include the high copy number crenarchaeal pRN and 'pRN-like' family plasmids, the medium copy number *Acidiphilum* plasmids pAM/pACRY/pACM and the broad host range, low copy number *Acidithiobacillus* pTF/pAca plasmids, among others recently added from sequencing projects (Table 12.1).

The research on cryptic plasmids of acidophiles has focussed mostly on the pRN plasmid family from *Sulfolobus* spp. and *Acidianus* spp. Members of this family seem to be exclusive to the crenarchaeal phylum and appear unrelated to any other known eubacterial or archaeal plasmids. Several of these plasmids have been used for the construction of *Escherichia coli–Sulfolobus* shuttle vectors (Berkner and Lipps, 2008). The pRN1 plasmid was the first crenarchaeal plasmid isolated and sequenced, and represents the prototype of this rather remarkable family (Lipps, 2009). All family members share a conserved genomic region and three highly conserved gene products which have been characterized biochemically. Recent studies have revealed that a single multifunctional replication protein (ORF904) with helicase, primase and DNA polymerase activity, reminiscent of the type of replication performed by some linear bacteriophages, is required to initiate plasmid replication (Berkner *et al.*, 2014). In marked contrast to all DNA polymerases, ORF904 is able to polymerize DNA *de novo*, i.e. without a primer. The enzyme is not a processive DNA polymerase, it does not have proofreading activity and its specific activity is rather low. Gene orthologues of *orf56* encode a fairly variable sequence-specific DNA-binding protein, designated CopG, that has a role in plasmid copy control through down-regulation of plasmid replication initiation. The *orf80* orthologues are partly annotated as *plrA* (plasmid regulatory gene A) but their function is currently unclear.

The pRN family also includes pSSVx from *Sulfolobus islandicus* REY 15/4, which is a hybrid between a pRN plasmid and a fusellovirus. The genome of pSSVx contains the three highly conserved ORFs of the plasmid family and two conserved ORFs originated from SSVs fuselloviruses, probably involved in packaging. It is believed that pSSVx replicates in the host cell as a plasmid and that upon superinfection with the helper SSV viruses it spreads as a virus satellite (Stedman *et al.*, 2003). pSSVi resembles members of the pRN plasmid family in genome organization but encodes an SSV-type integrase. Indeed, a few other pRN-type elements (pXQ1, pST1 and pST3) have been discovered in an integrated form in the chromosomes of *Sulfolobus* species, all of which encode an SSV-type integrase, likely responsible for plasmid integration into the crenarchaeal genome.

Additional small high-copy-number cryptic plasmids isolated from other *Sulfolobus* species and strains, grouped in the 'pRN-like' family, have been reported and some of them characterized. The plasmids pIT3, pTAU4, pORA1 and pTIK4 differ from the pRN family plasmids in that they encode a much less conserved or a completely different replication protein. This provides evidence that the replication apparatus of *Sulfolobus* plasmids can be quite diverse.

Table 12.1 Transmissible elements found in sequenced acidophiles

Phylum	*Species*	Strain	Accession no.	Size (kb)	No. of CDs	% G+C	Relaxase (MOB-type)	Type IV coupling protein	Conjugation system	Copy number	Host range	Name	Family
Small cryptic													
Crenarchaea	*S. islandicus*	REN1H1	NC_001771	5.4	6	37.3	–	–	–	H	B	pRN1	pRN
Crenarchaea	*S. islandicus*	REN1H1	NC_002101	7.0	6	37.4	–	–	–	H	B	pRN2	pRN
Crenarchaea	*S. islandicus*	HEN7H2	NC_004853	7.8	11	35.0	–	–	–	H	B	pHEN7	pRN
Crenarchaea	*Ad. ambivalens*	Lei 10	NC_005562	7.6	10	38.0	–	–	–	H	B	pDL10	pRN
Crenarchaea	*S. islandicus*	REY 15/4	NC_010011	5.7	8	38.7	–	–	–	H	B	pSSVx	pRN; VPH
Crenarchaea	*S. islandicus*	ARN3/6	NC_010365	7.0	7	39.5	–	–	–	H	B	pXZ1	pRN-like
Crenarchaea	*S. solfataricus*	IT3	NC_005907	5.0	6	44.0	–	–	–	H	B	pIT3	pRN-like
Crenarchaea	*S. neozealandicus*	EC	NC_006906	9.7	13	37.8	–	–	–	H	B	pORA1	pRN-like
Crenarchaea	*S. solfataricus*	P2	NC_013777	5.7	8	39.0	–	–	–	H	B	pSSVi	pRN; VPH
Euryarchaea	*P. oshimae*	DSM 9789	NC_016049	7.6	6	30.5	–	–	–	L	B	pPO1	pRN-like
Euryarchaea	*Tp. acidophilum*	H0–122	NC_008318	15.7	18	46.7	–	–	–	L	–	pTA1	
Proteobacteria	*At. ferridurans*	ATCC 33020	X52699	–	–	–	Q	–	–	L	–	pTF1	
Proteobacteria	*A. multivorum*	AIU301	NC_015181	12.1	14	59.6	Q	–	–	M	–	pACMV6	
Proteobacteria	*A. cryptum*	JF-5	NC_009472	8.8	9	61.0	F	–	–	M	–	pACRY06	
Proteobacteria	*A. multivorum*	JCM8867	NC_008691	5.2	5	57.5	Q	–	–	M	–	pAM5	
Proteobacteria	*At. caldus*	SM-1	NC_015852	9.8	10	59.8	P_A	–	–	L	B	pLAtc1	
Proteobacteria	*At. caldus*	ATCC 51756	CP005989	9.8	11	50.3	–	–	–	L	B	$pAca_{TY}2$	
Proteobacteria	*At. ferrooxidans*	MAL4–1	NC_001520	4.1	2	62.2	P_A	–	–	L	–	pTF4.1	
Proteobacteria	*A. cryptum*	JF-5	NC_009473	5.6	5	57.5	Q	–	–	M	–	pACRY07	
Proteobacteria	*A. multivorum*	AIU301	NC_015189	5.2	8	57.6	Q	–	–	M	–	pACMV7	
Proteobacteria	*A. cryptum*	JF-5	NC_009474	4.9	5	58.7	–	–	–	M	–	pACRY08	
Proteobacteria	*A. multivorum*	AIU301	NC_015182	1.7	1	60.9	–	–	–	M	–	pACMV8	
Firmicutes	*D. acidiphilus*	SJ4	NC_018067	3.9	4	43.3	Q	–	–	–	–	pDESACI.02	
Firmicutes	*Alb. acidocaldarius*	DSM 446	NC_013208	7.9	6	54.3	–	–	–	–	–	pAACI03	

Table 12.1 Continued

Phylum	*Species*	Strain	Accession no.	Size (kb)	No. of CDs	% G+C	Relaxase (MOB-type)	Type IV coupling protein	Conjugation system	Copy number	Host range	Name	Family
Conjugative													
Crenarchaea	*S. islandicus*	EC	NC_006422	28.9	41	37.1	–	+	+	M	B	pKEF9	pKEF-like
Crenarchaea	*S. islandicus*	EC	NC_004852	24.6	38	37.3	–	+	+	M	B	pING1	pKEF-like
Crenarchaea	*S. islandicus*	EC	NC_006425	35.4	54	37.3	–	+	+	M	B	pHVE14	pKEF-like
Crenarchaea	*Sulfolobus* sp.	NOB8H2	NC_006493	41.2	52	37.3	–	+	+	M	B	pNOB8	pKEF-like
Crenarchaea	*S. islandicus*	SOG2/4	NC_010598	26.0	41	36.7	–	+	+	M	B	pSOG2	pKEF-like
Crenarchaea	*S. islandicus*	EC	NC_006423	26.2	40	36.3	–	+	+	M	B	pARN3	pARN-like
Crenarchaea	*S. islandicus*	EC	NC_006424	26.5	1	37.1	–	–	+	M	B	pARN4 (DV)	pARN-like
Crenarchaea	*S. islandicus*	SOG2/4	NC_010597	29.0	46	35.8	–	+	+	M	B	pSOG1	pARN-like
Crenarchaea	*S. islandicus*	Y.N.15.51	NC_012624	42.2	50	36.1	–	+	+	M	B	pYN01	pARN-like
Crenarchaea	*S. islandicus*	L.D.8.5	NC_013770	26.6	32	36.1	–	–	+	M	B	pLD8501	pARN-like
Crenarchaea	*S. solfataricus*	P2	NC_021914	28.0	34	39.5	–	+	+	M	B	pMGB1	pARN-like
Crenarchaea	*S. tengchongensis*	EC	NC_005969	20.4	27	41.4	–	–	+	M	B	pTC	pARN-like
Crenarchaea	*Ad. hospitalis*	W1	NC_011299	28.6	39	35.9	–	–	+	M	B	pAH1	pARN-like
Proteobacteria	*A. multivorum*	AIU301	NC_015187	65.6	69	61.9	F/Q	–	+	–	–	pACMV2	
Proteobacteria	*A. multivorum*	AIU301	NC_015179	54.2	61	61.2	MOB_F	–		–	–	pACMV3	
Proteobacteria	*A. cryptum*	JF-5	NC_009469	89.0	77	62.1	FF	+	+	–	–	pACRY03 (DV)	
Firmicutes	*Sb. thermotolerans*	Y0017	NC_016040	59.2	61	58.2	P/C	+	+ (NP)	–	–	pY0017	
Firmicutes	*Sb. thermotolerans*	L15	JN119829	–	–	–	–	–	–	–	–	pL15	

Mobilizable													
Firmicutes	*Alb. acidocaldarius*	DSM 446	NC_013206	91.7	98	54.1	–	+	–	–	–	pAACI01	
Firmicutes	*Alb. acidocaldarius*	DSM 446	NC_013207	87.3	92	55.4	–	+	–	–	–	pAACI02	
Firmicutes	*D. acidiphilus*	SJ4	NC_018066	60.4	70	39.5	–	+	–	–	–	pDESACI.01	
Proteobacteria	*A. cryptum*	JF-5	NC_009470	37.4	32	59.6	P_A	+	–	–	–	pACRY04	
Proteobacteria	*At. caldus*	MNG	NC_010600	65.2	45	57.1	Q/P_A	–	–	L	B	pTcM1	pTcM1
Proteobacteria	*At. caldus*	SM-1	NC_015854	29.7	28	58.8	Q	–	–	L	B	pLAtc3	pTcM1
Proteobacteria	*At. caldus*	ATCC 51756	CP005988	27.5	33	59.4	Q	–	–	L	B	$pAca_{TY}.1$	pTcM1
Proteobacteria	*At. caldus*	SM-1	NC_015853	14.1	13	57.7	Q	–	–	L	B	pLAtc2	
Proteobacteria	*A. cryptum*	JF-5	NC_009471	37.2	28	60.9	F	–	–	–	–	pACRY05	
Proteobacteria	*A. multivorum*	AIU301	NC_015188	40.6	44	60.0	F	–	–	–	–	pACMV4	
Proteobacteria	*A. multivorum*	AIU301	NC_015180	14.3	19	58.9	P_A	–	–	–	–	pACMV5	
Proteobacteria	*At. caldus*	F	NC_004734	14.2	16	55.7	P_A	–	–	M	B	pTC-F14.	
Proteobacteria	*At. ferrooxidans*	EC	M57717	12.2	–	–	P_A	–	–	M	B	pTF-FC2	
Nitrospira	*L. ferrooxidans*	ATCC 49879	NC_006909	28.0	33	55.2	Q	–	–	–	–	p49879.2	
Nitrospira	*L. ferrooxidans*	ATCC 49879	NC_006907	28.9	32	57.8	Q	–	–	–	–	p49879.1	
Others													
Proteobacteria	*At. ferridurans*	ATCC 33020	NC_005023	19.8	13	56.1	–	–	–	–	–	pTF5	

B, broad host range; DV, deletion variant; EC, enrichment culture; H, high plasmid copy number (>50); L, low plasmid copy number (<15); M, medium plasmid copy number (>15, <50); NP, non-pheromone conjugation system; VPH, virus–plasmid hybrid.

Unlike the situation in the crenarchaea, where a large number of genetic elements have been characterized, little is known of extrachromosomal elements in the *Thermoplasmatales*. To date, the only sequenced and characterized plasmids from this phylogenetic order are the low copy number pTA1 isolated from *Tp. acidophilum* and pPO1 isolated from *Picrophilus oshimae*. Small cryptic plasmids are missing in closely related *Thermoplasma volcanicum* and *P. torridus*.

The search for native plasmids in the acidithiobacilli has yielded a large number of cryptic plasmids ranging in size from 4 to nearly 10 kb. Comparative studies of these plasmids using restriction endonuclease digestion and southern hybridization have profiled plasmid occurrence in diverse strains from different geographical origins but have yielded little insight into the nature of the sequence similarity of this group of small independent replicons.

In 1995 Chakravarty and colleagues presented evidence for a common plasmid replicon present in several strains of *At. ferrooxidans* (Chakravarty *et al.*, 1995). They found that four strains (TFI91, TFI92, TFI85 and TFI29) contained a related 9.8 kb plasmid sharing similar restriction endonuclease sites to the prototype plasmid pTFI91. Similarity of pTFI91-family plasmids was also extended to plasmid pTfA-4 and related plasmids present in Chilean isolates of *At. ferrooxidans* obtained from copper mines several hundred kilometres apart, as well as to the larger pTF5 plasmid isolated from the type strain of *Acidithiobacillus ferridurans* (ATCC 33020) (Dominy *et al.*, 1998). This indicated that these plasmids have a very wide geographic distribution and are either positively selected for or difficult to cure. Later findings confirmed that the pTFl9I-family replicon is a unique entity not related at the nucleotide level to other known broad host range plasmids. *Acidithiobacillus caldus* sequenced strains ATCC 51756 and SM-1 also contained 9.8 kbp small cryptic plasmids. Plasmids $pAca_{TY}2$ and pLAtc1 have been characterized bioinformatically and shown to be different from all currently sequenced acidithiobacilli plasmids and also from each other (Acuña *et al.*, 2013; You *et al.*, 2011a). Plasmid pLAtc1 encodes a mobilization system similar to that described for pRSB105, but is otherwise very different from it. In turn, plasmid $pAca_{TY}2$ from the type strain encodes a disrupted orthologue of the replication protein RepA of other acidithiobacilli and a gene whose product is similar to the plasmid exclusion protein Exc1, known to prevent the entry of IncI-type plasmids. Restriction profile analysis has recently revealed that plasmid $pAca_{TY}2$ corresponds to plasmid pTK1 reported in the original description of the type strain (Shelly Dean, Stellenbosch University, personal communication). Representatives of this set of broad host range plasmids or parts of them (e.g. the replication origin) have been recently used for the development of expression vectors (Zhang *et al.*, 2014) and shuttle vectors between *E. coli* and the *Acidithiobacillus* spp. (US Patent 8.163.558 B2; 2012).

Conjugative plasmids

The thermoacidophilic crenarchaeote *Sulfolobus* hosts a number of medium sized, moderate to high copy number, conjugative plasmids (Table 12.1). These appear to be present in about 3% of all isolated *Sulfolobus* strains (Prangishvili *et al.*, 1998). *Sulfolobus* conjugative plasmids have been grouped on the basis of their genomic characteristics as pKEF-like plasmids and pARN-like plasmids. Full genomic sequence comparison and phylogenetic analysis of representative genes has revealed that pKEF9, pING1, pHVE14, pSOG2 and pNOB8 are very similar to each other in gene content and organization, whereas pARN3, pARN4, pSOG1 and pTC form a related, but separate, group (Wang *et al.*, 2015). Natural deletion variants fitting this classification have also been described (e.g. pING2 del. var, of pING4; pARN4 del. var. of pARN3). They all encode a cluster of five or six core proteins that have been implicated in the conjugative process and, with few exceptions, also a non-partitioning type integrase. This integrase is known to be responsible for reversible chromosomal integration of pNOB8 plasmid in *S. tokodaii*, *S. acidocaldarius* and *Acidianus hospitalis* chromosomes. The presence of a defective conjugative plasmid-like element in the chromosome of several *Sulfolobus* species and *Metallosphaera sedula* DSM 5348, and the discovery of pAH1 in *Ad. hospitalis* W1, extends the host range of this group of conjugative plasmids to genera other than *Sulfolobus*. Being also integrative elements, these plasmids have been implicated in facilitating chromosomal DNA conjugation for some *Sulfolobus* species.

The best-characterized plasmids are the pKEF family plasmids. They range in size between 20 and

40 kb and are all medium copy number plasmids in *S. solfataricus* test strain PH1 (20–40 copies/chromosome); although slightly lower copy numbers occur in their natural hosts (Prangishvili *et al.*, 1998). Conserved genes within this family are organized in three functionally distinct genomic sections grouping the conjugation genes, the origin of replication and the proposed replication genes which resemble pRN family plasmids *repA*, *copG* and *plrA* genes. The genomic regions between these sections are more variable in size, gene content and sequence, and harbour putative recombination motifs which contribute to plasmid variability. Indeed, mixtures of variant plasmids have been detected upon propagation of these plasmids in laboratory *S. solfataricus* strains. Even so, pKEF plasmids are stably maintained in the original plasmid-carrying strains, which is consistent with them carrying orthologues of bacterial proteins involved in partitioning (ParA, ParB).

Despite the fact that conjugational exchange of plasmid DNA has been observed between *Sulfolobus* spp. (Prangishvili *et al.*, 1998), it is still unclear what the exact mechanism is, what proteins are involved, and what the origin of transfer looks like. However, one of the well-conserved regions of pKEF plasmids has provided hints in this respect. This region contains six ORFs predicted to encode the core conjugative apparatus in *Sulfolobus*, and two of these genes products (pKEF9: p12 and p01) exhibit sequence motifs and domain structures characteristic of the bacterial Type IV secretion system coupling protein (VirD4/T4CP) and relaxase (VirB4) proteins required for DNA transfer. The *Sulfolobus* T4CPs are considerably larger than bacterial VirD4, while the VirB4-like proteins are slightly smaller than their bacterial counterparts, and both proteins have C-termini and/or N-termini that are unrelated to other family members. The other four ORFs encode predicted membrane proteins with up to 10 transmembrane helix motifs, and are believed to be involved in transmembrane pore formation required for conjugational exchange of plasmid DNA, mimicking the single membrane spanning protein that facilitates self-transfer in some small plasmids of Gram-positive bacteria. Although VirB2 orthologues are absent from pKEF family plasmids, genes encoding small proteins with pilin-like hydropathy profiles (pKEF9: p05) are collinear with genes encoding VirB4-like subunits and might assemble to form the pili detected on *Sulfolobus* pKEF9 donor cells (Schleper *et al.*, 1995).

To date, limited progress has been made in characterizing the mechanisms of replication and copy number control of the conjugative plasmids of archaeal acidophiles, and very little is known about the interplay between the different elements within a host. Superconjugation immunity studies, in which a plasmid containing host is challenged with a second plasmid after successive conjugative transfers, have been performed for some of these plasmids (Prangishvili *et al.*, 1998). These experiments have revealed incompatibility of closely related plasmids and different quantitative responses on the copy number of certain conjugative plasmids. While the incompatibility of bacterial plasmids is mainly a consequence of competition between two plasmids for the same replication or partition apparatus, the mechanisms behind incompatibilities between archaeal plasmids still needs to be determined.

In the case of acidophilic bacteria, only a few of the known plasmids encode the T4SS required for conjugative self-transfer. These include the 59 kb pY0007 (and its related pL15) from *Sulfobacillus thermotolerans* (Dean and Rawlings, 2011) and two MOBF-type large plasmids, pACVM2 and pACRY03, from *Acidiphilium multivorum* and *Acidiphilium cryptum*. A deletion variant of pACRY03 lacking the coupling protein and the mating bridge genes is also found in *A. multivorum* under the pACMV3 designation. Both *Sb. thermotolerans* large, high G+C content plasmids pY0017 and pL15 encode a potential non-pheromone conjugation 14-gene operon typical of Gram-positive bacteria. Although relatively little is known about this conjugation system, it has been suggested that it might be involved in establishing physical contact between donor and recipient cells in order to facilitate conjugal transfer of the plasmids (Dean and Rawlings, 2011). None of the *Acidiphilium* conjugative plasmids have been analysed to any extent to date and conjugative transfer capacity still awaits experimental evaluation for most acidophilic bacterial conjugative plasmids.

Mobilizable plasmids

A considerably larger number of mobilizable plasmids have been identified in acidophilic bacteria,

mostly in the *Proteobacteria*. Most of the reported or predicted mobilizable plasmids are fairly large elements (ranging from ~ 5 to 90 kb), encoding at least a relaxase and/or a coupling protein. Species-specific plasmid backbones accompanied by different accessory genes appear to be a recurring theme in acidophiles such as *At. ferrooxidans* (e.g. Dominy *et al.*, 1998) and *At. caldus* (e.g. van Zyl *et al.*, 2008a).

A family of mobilizable plasmids (pTcM1 family) ranging in size from approximately 30 to 70 kb has been found in a variety of *At. caldus* strains from different continents (Table 12.1). A number of these plasmids have been characterized to different degrees, including pTcM1, pTcF1 and pCSH12, pLAtc3 and $pAca_{TY}1$. These plasmids share a common region that spans 11 to 26 kb and includes the plasmid replicon (RepA-type) and other functions associated with plasmid stability and maintenance (ParA-ParB), as well as an IncQ-like mobilization relaxase of the MobA/RepB-type and a PasAB plasmid addiction system (van Zyl *et al.*, 2003, 2008a). No *oriT* site was identified within the common 26-kb region, yet pTcM1 is mobilizable. The plasmid backbone appears to be conserved with respect to replicons from microorganisms thriving in different environments. Also, a gene encoding a predicted invertase-like protein with motifs in common with the PinR family of site-specific recombinases is present in pTcM1-family plasmids and other plasmids isolated from acidophilic bacteria (pTCF14 and pTF5), the role of which still remains elusive.

Two closely related broad-host-range, mobilizable, medium copy number plasmids (10–15 plasmids per chromosome), were discovered in the acidithiobacilli (Rawlings, 2005). Plasmids pTF-FC2 (IncQ-2a) and pTC-F14 (IncQ-2b) share with other IncQ-family replicons the presence of *repB*, *repA* and *repC* genes encoding the primase, helicase and iteron-binding protein, respectively, as well as a 22-bp iteron based *oriV*. Despite their similarity, their origins of replication are compatible (Gardner *et al.*, 2004). Both plasmids lack an active partitioning system and their stability depends on the toxin–antitoxin system PasAB, which also seem to play a role in copy number control (Matcher and Rawlings, 2009; Deane and Rawlings, 2004; Gardner *et al.*, 2001), resembling CopG-mediated control of crenarchaeal pRN1-family plasmids. Both plasmids have a five-gene mobilization module encoding a MobA relaxase of the MOBP-type. While MobA and MobC appear to be essential for plasmid mobilization, the non-essential MobD and MobE only seem to affect the frequency of mobilization (van Zyl *et al.*, 2003). Even though no conjugative partner plasmids were identified at the time of isolation for pTF-FC2 and pTC-F14, both can be efficiently mobilized by RP4, an IncP conjugative plasmid (van Zyl *et al.*, 2003). The pTC-F14 IncQ replicon has been cloned into plasmid R388, which could be transferred from *E. coli* HB101 to *At. caldus*, and used to allow the replication of this shuttle vector in *At. caldus* for the construction of allelic exchange mutants (van Zyl *et al.*, 2008b).

Additional mobilizable plasmids have been isolated from *At. ferrooxidans* (pTF4.1) and *At. ferridurans* ATCC 33020^T (previously referred to as *At. ferrooxidans* pTF1) encoding either MobA/MobL or MobL/MobS relaxases that enable the plasmids to be mobilized at high frequency by IncP group plasmids (Drolet and Lau, 1992).

Significantly less is known about the plasmids of the leptospirilli. No plasmid from this group of acidophiles was known until 2005 when a survey in 10 strains of *L. ferrooxidans* and 6 strains of *L. ferriphilum*, identified the first two plasmids for a member of this unique group of bacteria (Table 12.1). Plasmids p49879.1 and p49879.2 were isolated from *L. ferrooxidans* ATCC 49879 (previously referred to as R3), and thereafter sequenced and experimentally characterized (Coram *et al.*, 2005). Plasmids p49879.1 and p49879.2 were found to be of similar sizes, i.e. ~30 kb, but with slightly different overall G + C mol% ratios. Approximately half of the ORFs present on the two plasmids had products related to known proteins in databases. Although no clear candidates for replication proteins could be identified and neither plasmid was capable of replication in *E. coli* or *Pseudomonas putida*, both plasmids contained four 20-bp iterons that may serve as replication protein binding sites. Other plasmid-associated functions such as a mobilization and stability systems have also been characterized: two divergently transcribed ORFs encoding MobS-like and MobL-like proteins, a MOBQ relaxase and a region of DNA with features typical of the *oriT* regions of pTF1 and the IncQ plasmids. In spite of this, when cloned in a test plasmid, neither of the candidate *oriT* could be mobilized by IncP

plasmids such as RP4 in *E. coli* S17.1 (van Zyl *et al.*, 2003). Both plasmids encode proteins with high amino acid sequence identity to ParA-like ATPases responsible for active partitioning of plasmids into progeny cells upon cell division, albeit from different ParA families. This divergence may partly account for the observed plasmid compatibility of p49879.1 and p49879.2. Stability experiments showed the *par* system of p49879.2, but not that of p49879.1, to be functional in *E. coli*.

A number of accessory genes carried by the mobilizable plasmids found in acidophiles encode an eclectic collection of functions. These include energy metabolism related functions (e.g. a terminal quinol oxidase complex), cofactor biogenesis functions (e.g. haem maturation proteins), electron carriers (e.g. 4Fe–4S ferredoxin), transporters, detoxification functions and regulators (Coram *et al.*, 2005; Gardner *et al.*, 2001). However, their functionality and whether they contribute to host cell fitness remains unclear.

Integrative genetic elements

Many mobile elements that physically integrate in the chromosome (or accessory replicons) populate most prokaryotes. Although different bacteria and archaea seem to be differently prone to acquire Integrative Elements (IEs), few of them have been characterized in any detail, particularly in acidophiles. This section focuses on genetic elements of acidophiles that integrate in a site-specific fashion in the host chromosome. This group of elements includes proviruses, conjugative self-transmissible elements (known as ICEs; Integrative Conjugative Elements), mobilizable elements (known as IMEs, integrative mobilizable elements) and other elements that are no longer or were never self-transmissible but can be mobilized under certain circumstances with the help of other MGEs, generally referred to as 'genomic islands' (GIs).

These general definitions group together an overarching diversity of elements ranging in size from less than 10 kb to nearly 500 kb, sharing one or more diagnostic features. Hallmarks of IEs are functional or cryptic genes encoding integrases and small (16–50 bp) perfect or almost perfect direct repeats flanking the element. The integrases are necessary and sufficient to mediate integration and are also required for excision, although this process requires additional factors in most cases. Most of the bacterial integrases that mediate site-specific recombination between extra-chromosomal DNA elements and chromosomes are members of the tyrosine recombinase family, and most integration target sites are tRNAs or other structural RNAs (e.g. tmRNAs). Chromosomal integration in archaea differs from bacteria in that the integrase genes become split into two segments which border the integrated element, and that integration can be reversed only if the intact integrase is produced (Wang *et al.*, 2015 and references therein). These mechanistic differences may have implications for productive HGT across the boundaries of the three domains of life.

In recent years, many IEs have been identified on the basis of sequence data and comparative studies of closely related strains or species representing most of the prokaryotic taxonomic spectra. Although several of these elements may actually be degenerate, experimental evidence has began to build up that supports different aspects of their functionality.

Viruses and proviruses

Evidence collected over the past 30 years points to a rich diversity of viruses infecting acidophiles from both high and low temperature environments (Table 12.2). Most of these viruses have been associated with archaeal acidophiles including *Sulfolobus* spp., *Acidianus* spp., the ARMAN nanoarchaea and *Ferroplasma* spp. (Wang et., 2015; Dellas *et al.*, 2014; Andersson and Banfield, 2008; Allen *et al.*, 2007). These viruses have proved to be unique in terms of their morphology, which includes bottle, lemon and droplet shapes, and also in terms of their life cycles, which span integrated and silent proviruses that parasitize their host, infective and lytic forms that kill the cells they infect, and chronic infecting virions that keep their hosts viable while releasing viral progeny (Wang *et al.*, 2015; Dellas *et al.*, 2014). Tailed viruses infecting bacterial extreme acidophiles of the genus *Acidiphilium* and *Acidithiobacillus* have also been reported (Tapia *et al.*, 2012).

Viruses from acidophiles seem to be capable of integrating their genome into that of their host, establishing lysogeny and moving genes around in acidic econiches. Support for the occurrence of proviruses in acidophiles comes from genomic, metagenomic and experimental evidence. Blocks

Table 12.2 Known viruses of acidophilic prokaryotes

Host	Virus			No. of species	Genome size	Release strategies	Staus	Integrase
	Type	Genus	Name					
Sulfolobus	Circular dsDNA	*Fuselloviridae*	SSV	9	14.7–17.6	Non-lytic	Provirus	+
			SMV	1	48.7	Non-lytic		–
		Unclassified	STSV	2	75.2–76	Non-lytic	Provirus	+
		Guttaviridae	SNDV	1	20	Non-lytic		–
		Turriviridae	STIV	2	16.6–17.6	Non-lytic		–
	Linear dsDNA	*Rudiviridae*	SIRV	2	32.3–35.4	Non-lytic		–
Acidianus	Circular dsDNA	*Fuselloviridae*	ASV1	1	24.1	Non-lytic		–
		Bicaudaviridae	ATV	1	62.7	Non-lytic	Provirus	+
	Linear dsDNA	*Ampullaviridae*	ABV	1	23.8	Non-lytic		–
		Rudiviridae	SIRV	2	32.3–35.4	Non-lytic		–
		Lipothrixiviridae	AFV	6	21–41.1	Non-lytic	Provirus	–
	Linear ssDNA	*Lipothrixiviridae*	AFV3	1	40.4	Non-lytic		–
A-/E-/G-plasma	Circular dsDNA	ND	AMDV	4	>10 kb	Non-lytic	Provirus	ND
ARMAN	ND	Lemon-shaped	ND	ND	ND	Non-lytic	ND	ND
		Rod-shaped	ND	ND	ND	Non-lytic	ND	ND
Acidiphilium	Circular dsDNA	*Siphoviridae*	ΦAc1	1	102	Lytic	Provirus	ND
Acidithiobacillus	Circular dsDNA	*Myoviridae*	AcaML1	1	59.3	ND	Provirus	+
Leptospirillum	dsDNA	ND	AMDV1	1	ND	ND	Provirus	ND

ND, not determined.

of unique genes with anomalous G + C contents, of putative viral origin, have been found in the '*F. acidarmanus*' fer1 genome, and in the 'alphabet-plasma' *Thermoplasmatales* and the leptospirilli datasets from the Iron Mountain acid mine drainage (AMD) system (Andersson and Banfield, 2008; Allen *et al.*, 2007). Also fuselloviruses, bicaudaviruses and integrative plasmid–virus hybrids inserted at tRNA genes in the genomes of *Sulfolobus* spp. and *Acidianus* spp. (e.g. Guo *et al.*, 2011) have been described in the literature. In the case of bacteria, a 59-kb inducible temperate Myoviridae-like prophage located in the *srrA* tmRNA of the *At. caldus* type strain genome has been identified (Tapia *et al.*, 2012). Further research on pro/viruses of acidophiles is certain to provide greater insights into the diversity and the unique biology of viruses from acidophiles.

Genomic islands and conjugative/mobilizable elements

A number of non-viral integrated elements have been formally described in the acidithiobacilli in recent years. These include several genomic islands in *At. ferrooxidans* (Bustamante *et al.*, 2012; Orellana *et al.*, 2011) and *At. caldus* (Acuña *et al.*, 2013). Bioinformatic evidence of additional IEs has also been generated for *At. ferrivorans* (Gonzalez *et al.*, 2014) and *At. thiooxidans* (Travisany *et al.*, 2014).

Two large unique genome segments have been identified in *At. ferrooxidans* strains ATCC 23270^T and ATCC 53993 which account for about 16% difference in gene content between the two strains (Holmes *et al.*, 2009). The first is a ~150 kb long genomic island found exclusively in *At. ferrooxidans* ATCC 53993 that contains genes encoding mercury, arsenic and copper metal resistance determinants (Valdés *et al.*, 2010). Experimental

data have shown that *At. ferrooxidans* ATCC 53993 has a much higher resistance to $CuSO_4$ (>100 mM) than the type strain (<25 mM), which could be possibly explained by the presence of the additional copper resistance genes in this GI (Orellana *et al.*, 2011). Another genomic island, twice as long, is present in *At. ferrooxidans* ATCC 23270^T (Holmes *et al.*, 2009; Levicán *et al.*, 2009). Its genetic organization resembles that of other well-characterized integrative conjugative elements (ICEs), and includes integration/excision, conjugative DNA processing, transfer, maintenance and regulation proteins organized in a number of discrete gene modules. Excision of the element, and expression of a complete set of genes encoding self-transfer functions under normal and DNA-damaging growth conditions, suggests this element is indeed functional (Bustamante *et al.*, 2012). As in most mobile genetic elements, approximately 60% of its ORFs encode proteins with unknown or hypothetical functions, but additional genes encoding features that may confer selective advantages to the host strain have been identified, including a exopolysaccharide biosynthesis cluster, a CRISPR-Cas system and a cluster of 36 tRNA genes covering all 20 amino acids. All these accessory functions have been found to occur in the mobilome of other sequenced acidithiobacilli (Talla *et al.*, 2014; Travisany *et al.*, 2014; Acuña *et al.*, 2013).

Comparative genomic studies of *At. caldus* ATCC 51756 and SM-1 have revealed the presence and activity of >10 integrative mobile elements in *At. caldus* ranging in size from ~10 to 200 kb (Acuña *et al.*, 2013). Five of them encode components of type IV secretion systems (T4SS) producing P-type pili for conjugative plasmid transfer that define them as ICE, but only two of these (ICE-2 and ICE-3 type elements) have all the essential components of a functional T4SS, and one encodes at least the relaxase and the coupling protein typically present in non conjugative mobilizable plasmids (IME-1). Experimental analyses revealed the co-existence of the integrated and excised forms of all elements in a collection of *At. caldus* strains, indicating that the elements are capable of forming circular intermediates suitable for further transfer to recipient hosts. Most ICEs and the IME are widespread geographically, indicating that conserved or variant versions occur in the *At. caldus* population as active excising elements that may be transferred horizontally when needed.

High levels of genomic divergence due to differences in the mobile gene complement have also been identified in other acidophiles. Two dominant and recombining strains of *Leptospirillum* group II from the Iron Mountain AMD showing a very high similarity at the rRNA level (0.3% difference), have differences in gene content approaching 20% (Simmons *et al.*, 2008). A similar scenario has been pinpointed for *Hydrogenobaculum* sp. strains (Romano *et al.*, 2013). Integrated plasmid-like elements, virus-like elements and genomic islands encoding conjugal transfer proteins are present in representative strains of all these acidophilic genera. Several of these IEs encode copper-, silver-, arsenic-, and mercury- resistance genes, among other diverse functions, suggesting that they are relevant contributors to fitness and niche adaptation aiding in the dispersal of these and other functions.

Nevertheless, little evidence on the activity and impact on fitness of these IEs and their genetic cargo has been obtained to date. Proteomic studies performed on Iron Mountain AMD leptospirilli found that few proteins encoded by genes in the integrated mobile regions were identifiable, while many of the conjugative transfer proteins encoded by genes in a similar extrachromosomal plasmid were indeed expressed (Goltsman *et al.*, 2009). Further studies have shown that proteins associated with MGEs in the leptospirilli were significantly elevated in low developmental stage biofilms from this environment (Mueller *et al.*, 2010). In turn, studies on *Leptospirillum* group IV grown in bioreactors have shown that genes encoded in mobile elements are overrepresented in transcriptomic data sets (Goltsman *et al.*, 2013). Further work is required to understand the role and relevance of these large MGEs in adaptation and evolution of acidophiles.

Translocative elements

Translocative or transposable elements (TPEs) are components of nearly all prokaryotic genomes. They are capable of moving from one chromosome location to another, regardless of the existence of sequence homology between the element and the target site, by the action of a transposase (TPase) in a process called transposition. Transposases

constitute a highly diverse group of enzymes and they are the single largest component in sequenced genomes (encompassing 1–10% of the genome) and environmental metagenomes. By transposing into transmissible and integrative elements, they can transfer between hosts and serve as sites of homologous recombination. This allows plasmids and other elements to become inserted into chromosomes and to loop out again, sometimes carrying segments of adjacent chromosomal DNA. These elements can also serve as sites of chromosomal DNA deletions, inversions and a variety of other rearrangements; therefore, they contribute greatly to genetic flexibility (or plasticity) and the entry of genetic material into the horizontal gene pool. Microbial transposable elements can be classified into two large groups: (i) autonomous elements with open reading frames that encode the products required for transposition (i.e. the TPase) together with insertion sequences (ISs) and transposons; (ii) non-autonomous elements that derive from the former by deletion of the TPase while retaining the sequences necessary for *in trans* transposition by cognate TPases, which include miniature inverted repeat transposable elements (MITEs).

Insertion sequences

Insertion sequences (ISs) are fairly abundant in acidophilic archaea and bacteria (Fig. 12.2). More than 200 copies of intact ISs of at least 25 different types were found in the *S. solfataricus* P2 genome and a comparable number in *S. tokodaii*, being this the largest number of transposable elements found in any sequenced genome to date. In the *Sulfolobales*, the types of ISs vary substantially between species, and even closely related strains can differ greatly with respect to their complement of total and active ISs (Blount and Grogan, 2005). While *S. acidocaldarius* maintains a very stable genome organization, evidence of extensive IS-catalysed rearrangements in *S. solfataricus P2* and *S. tokodaii* genomes has been described. In turn, in *S. islandicus* and *Ad. hospitalis*, the majority of potentially transposable autonomous and non-autonomous mobile elements, as well as many degenerate copies of the former, fall into a large variable region of their genomes, helping preserve overall genome organization (e.g. Guo *et al.*, 2011; You *et al.*, 2011b). In contrast to the crenarchaeota, acidophilic euryarchaeota have significantly fewer IS elements. Allen *et al.* (2007) reported less than 100 transposase insertions in the genome of the '*F. acidarmanus*' fer1 isolate and ~50 in the environmental population fer1(env) from which fer1 was derived, representing at most 10 families of ISs. Even fewer ISs are present in *Tp. volcanium* (27 ISs) and *Tp. acidophilum* (four ISs), while *P. torridus* seems to be completely devoid of them. The small size of *P. torridus* genome (1.55 Mb) and its extreme gene density have been suggested to explain its lack of ISs, yet the dramatic differences observed between these acidophilic archaeal phyla still lack an adequate explanation.

Although none of the sequenced acidophilic bacteria contains as many ISs as the *Sulofolobales*, the number and diversity of IS is also high in this group (López de Saro *et al.*, 2013). The occurrence of many IS elements has been reported in recent genome announcements and comparative genomic analyses, but descriptions on active transposable elements have only been published for a small number of acidophiles (e.g. Holmes *et al.*, 2001). Early southern hybridization experiments discovered two families of actively transposing ISs of about 1.5 kb in size (originally named IST1 and IST2) in a variety of laboratory strains and natural isolates of *At. ferrooxidans* obtained from different parts of the world. Among IS-positive strains, the number of IST2 copies varied from 15 to 25, and the number of IST1 copies, from 1 to 10 (Yates and Holmes, 1987). IST2 occurs in a small region of the genome of *At. ferrooxidans* ATCC 19859, considered to be a hot spot of transposition, and seems to exhibit two different patterns of insertion depending on the growth conditions (Cádiz *et al.*, 1994). ISTI (currently known as ISAfe1) is an ISL3 family insertion sequence from *At. ferrooxidans* ATCC 19859 (Holmes *et al.*, 2001). It is about 1.3 kbp in size, has 26-bp imperfectly paired inverted terminal repeats, 5-bp target duplications, and encodes a transposase (OrfA) and a second ORF of unknown function (OrfB). Specific changes in the position of IST1 have been associated with the formation of a large-colony variant of *At. ferrooxidans* ATCC 19859 that reversibly lost the capacity to oxidize iron and exhibited a swarming phenotype. Phenotypic switching was later shown to be correlated with the high frequency insertion and excision of ISAfe1 into, and out of, the *resB* gene which encodes a cytochrome c-type maturation protein relevant for the iron oxidation pathway (Cabrejos

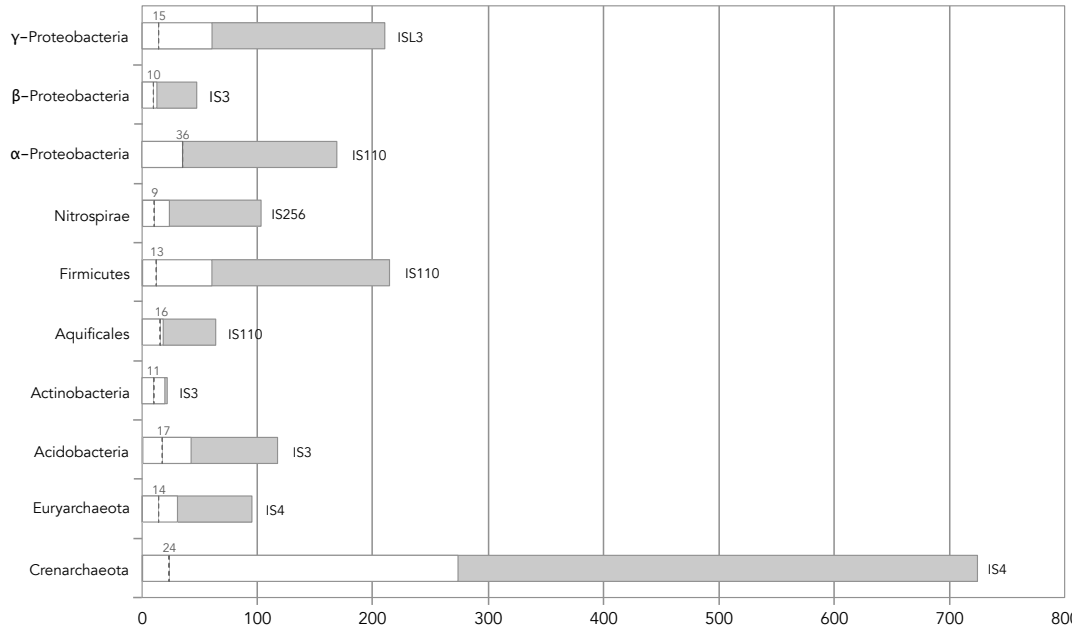

Figure 12.2 Number of insertion sequences in complete genomes of acidophilic prokaryotes. The number of predicted insertion sequences found in sequenced genomes of acidophiles, as organized by phylum, is shown in grey. The number of ISs for the most abundant family in each phylum is shown in white. The number of ISs normalized by the number of genomes in each category is indicated above the dotted lines.

et al., 1999). The cause of the swarming phenotype remains unexplained.

Recent full and partial sequencing of several strains of *Acidithiobacillus* have provided further insight into the diversity of ISs and their genomic distribution. For instance, natural isolates tend to have more ISs than long laboratory-kept type strains, and these tend to cluster within a few genomic regions, several of which have been identified as different types of genomic islands and transposition hotspots (e.g. Acuña *et al.*, 2013). The genome of an *At. caldus* strain (SM-1) isolated from a gold bioreactor in China carried many more copies of ISL3, IS4, IS5 and IS21 elements than all other known *Acidithiobacillus* genomes (You *et al.*, 2011a), and the genome of *At. thiooxidans* Licanantay strain, isolated from a copper pile in northern Chile, carried eight times more transposases that the type strain (Travisany *et al.*, 2014). Comparative genomic analysis has also provided evidence for recent proliferation of certain families; IS5 and ISL3 in the *At. caldus* (27 and 37 copies, respectively) and for the ISAs1 family in *At. ferrooxidans* ATCC 53993 (24 copies) (López de Saro *et al.*, 2013). Only a fairly small percentage of these elements (<10%) have been found to actually interrupt coding regions or to mediate gene acquisition.

A large number of ISs have also been found in *Leptospirillum* species (Mueller *et al.*, 2010), and comparable differences in the relative abundance of some IS families between closely related species and strains have also been reported for these acidophiles.

Transposons

Tn21 transposons are probably the best examples to illustrate how TPEs can be vehicles of transmission of specific phenotypic traits. Highly arsenic-resistant strains of *At. caldus* and *L. ferriphilum* isolated from the same bio-oxidation process treating a gold-bearing arsenopyrite concentrate at the Fairview mine in South Africa have been found to contain distantly related Tn21-like transposons (Tuffin *et al.*, 2005, 2006). Both the *L. ferriphilum* and *At. caldus ars* transposons (TnLfArs and TnAtcArs) were capable of transposition in *E. coli* and thus, believed to be functional in *L. ferriphilum* and *At. caldus*, respectively (Kotze *et al.*, 2006; Tuffin *et al.*, 2006). Despite their presumed functionality in their cognate acidophilic hosts, TnLfArs and

TnAtcArs were sufficiently different from each other to indicate that they were obtained independently and not to have been transferred from one bacterium to the other during their coexistence in the biooxidation tank. This suggests either that HGT might not be possible between these phylogenetically distant bacteria or that bio-oxidation tanks are not a suitable environment for HGT between them to occur.

Miniature inverted repeat transposable elements

Miniature inverted repeat transposable elements (MITEs), common in eukaryote genomes, are also present in the genomes of acidophilic archaea. Some MITEs are present in multiple copies of identical or near-identical sequence (> 140 copies in *S. solfataricus* P2), consistent with their having recently transposed within the genome (Redder *et al.*, 2001). Although they are more prevalent in *S. solfataricus* P2 than in any other sequenced archaeal genome, MITEs are common in other *Sulfolobus* species, as well as in *Tp. volcanium* and *Ad. hospitalis.* Since MITEs can constitute homologous recombination sites, they have been implicated as being responsible for many of the genomic rearrangements observed in archaea (Brügger *et al.*, 2004). Although MITEs in general show a low level of transpositional activity, they can also transpose to different genomic regions and mobilise genomic regions located between two MITEs. Since they arise from internal deletions in IS elements but retain the terminal sequences, the parent IS element can be frequently identified. Archaeal MITEs have been classified into two main types, type I MITEs, including two representatives in *Tp. volcanium* (which derive from ISE1247) and nine in *S. tokodaii* (which derive from the IS605 family), and type II MITEs present in the genomes of *S. tokodaii* and *S. islandicus* at low abundance and of *S. solfataricus*, where they constitute up to 0.6% of the genome with predicted partner IS elements ISC1048, ISC1217, ISC1058 and ISC1173. One challenge for the future will be to determine whether, and how frequently, the sequence diversity created by MITE insertions has resulted in altered gene expression or gene products in acidophiles.

Defence mechanisms

Protection against potentially deleterious effects of entry and propagation of foreign MGEs has warranted the emergence and selection of alternative defence strategies by both bacteria and archaea. Some of these strategies entail distinction of self DNA from foreign DNA by means of modification (methylation or phosphorothioation) of self DNA and degradation or restriction of non-self DNA; very much like a prokaryotic version of innate immunity. Other systems, in turn, memorize the encounters with MGEs and attack them specifically afterwards. This is the case of the CRISPR-Cas systems, which have been compared to prokaryotic adaptive immunity systems. Little is known about the RM-systems in acidophiles. On the contrary, substantial evidence on the occurrence and relevance of CRISPR-Cas systems in acidophiles has accumulated in recent years. Hints for HGT between distant prokaryotes has been obtained for both RM and CRISPR-Cas systems; interestingly, these systems act to limit HGT of the mobilome elements but are themselves part of it.

CRISPR-Cas systems

CRISPR-Cas systems are used to acquire heritable protection against invading MGEs by both bacteria and archaea. Essential for the function of these systems are the Cas proteins and the CRISPR loci. CRISPRs are made up of a variable number of repetitive DNA motifs, averaging 30 bp, separated by spacer sequences of similar size. The spacers are derived from MGEs and incorporated into active loci after exposure to, and successful destruction of, the foreign elements. Upon entrance of a MGE into the cell, the CRISPR loci precursor transcripts are processed into a number of short RNAs, called crRNAs, and displayed on Cas protein complexes. Recognition of the invading foreign DNA is achieved by an RNA-guided sequence specific mechanism through crRNAs matching the invader DNA or RNA, and subsequent degradation is exerted by Cas nucleases. Because of their inherent mechanism of action, CRISPR-Cas systems are both adaptive and heritable and thus constitute a memory of previous attacks which the cell has survived or which the population as a whole has been able to cope with.

CRISPR loci are well represented in acidophiles

and have been studied quite extensively in a couple of acidophilic archaeal and bacterial models, namely the *Sulfolobales* (reviewed by Garrett, 2015) and the leptospirilli (Tyson and Banfield, 2008). Nearly all sequenced acidophilic archaea and approximately half of acidophilic bacteria encode CRISPR loci and/or Cas proteins in their genomes, fitting general statistics derived from other econiches. By combining evidence from phylogenetic, comparative genomic and structural analyses, basic structural and functional blocks have been recognized and accommodated into a three-types based (and 12-subtypes based) classification (Makarova *et al.*, 2011). Diversity of CRISPR-Cas system types in acidophiles is summarized in Table 12.3. The most frequent types in acidophilic archaea are I-A, III-B and III-A and in bacteria I-B, I-C and III-B, with 60% of the acidophilic archaea and only 30% of acidophilic bacteria having simultaneously a type I and a type III system. Type II interference systems are absent from all currently known acidophiles. The accepted consensus is that most type I systems target dsDNA, while type III systems target RNA. Thus, acidophilic archaea seem to be better prepared to confront DNA and RNA MGEs, (e.g. archaeophages) than bacterial acidophiles. For instance, the *Sulfolobales* host subtypes I-A, I-D, III-B, III-D and III-VIII-I CRISPR-Cas. Such diversity of the CRISPR-Cas system types and spacers (>4000 in complete sequence genomes of the genus) correlate with the considerable diversity and abundance of genetic elements and viruses infecting the *Sulfolobales* and thriving in thermo-acidic econiches (Garrett *et al.*, 2015; Dellas *et al.*, 2015). Indeed, several viral families containing circular or linear dsDNA genomes as well as positive strand RNA viruses have been detected in the econiches dominated by these thermo-acidophiles. Less impressive numbers are found in other acidophiles (Table 12.3), but still more than 70% of the completely sequenced acidophiles carry CRISPRs in their genomes, suggesting that viruses and other MGEs are as common a threat in acidic econiches as in other types of environments. Archaeal acidophiles tend to have more arrays per genome (up to nine) and more repeats per array (up to 457) than bacteria. These observations have been made also in microbial communities inhabiting other types of environments. Such comparisons suggest that extremophiles frequently have more CRISPR-Cas systems and larger numbers of CRISPRs. Most acidophiles carry repeats that are highly conserved in sequence within each predicted locus, fitting well-accepted criteria for functionality. Nevertheless, with few exceptions, functionality and activity against invading MGEs of most CRISPR loci of acidophiles remain to be evaluated.

A number of studies of the CRISPR-Cas systems of acidophiles have contributed significantly to the general understanding of CRISPR biology. The *Sulfolobales* have served as a model of study for several aspects of the basic molecular and structural biology of CRISPR-Cas systems, providing seminal insights into both the adaptation and interference processes, and their regulation in type I and type III systems (for recent review see Garrett *et al.*, 2015). The CRISPR-Cas systems of the acidithiobacilli, which host rather unique loci (type U loci for Unclassified and subtype Aferro for *At. ferrooxidans* ATCC 23270^{T}), have provided insights into the diversity of systems yet to be discovered and the challenges associated with the classification of these systems (Makarova *et al.*, 2011). CRISPR-Cas loci of the acidithiobacilli are located within ICE elements (Acuña *et al.*, 2013; Bustamante *et al.*, 2012), lack the ubiquitous core genes involved in adaptation (*cas1* and *cas2*) and encode proteins that are highly divergent from other known and characterized Cas (Acuña *et al.*, 2013; Holmes *et al.*, 2009). Despite the fact that the number and diversification of spacer sequences suggests these loci to be active, it remains unclear whether these are self-sufficient systems or defective systems that capture and utilize pre-existing CRISPR arrays. Further study of the Aferro-subtype CRISPR-Cas systems is likely to enlarge the list of emerging alternative roles of these systems. Also, population-level genomic studies in the leptospirilli have proved CRISPRs to be amazing tools for matching viruses to their hosts by means of spacer sequence analysis (Anderson and Banfield, 2008), and for grasping the interactions and dynamics of viruses and microbial host populations (Tyson and Banfield, 2008). Further details on CRISPR analyses in the Iron Mountain AMD system are nicely covered in Chapter 13.

Restriction–modification systems

Foreign DNA is frequently cut into pieces by endonucleases encoded in the genome of the recipient cell, which recognize and cut at specific DNA

Table 12.3 Defence systems present in sequenced acidophiles

	Taxonomy	Name	No. of species	No. of strains	Genome size[a] (kb)	CRISPR–Cas System Type[b]									# CRISPR loci	# Spacers	R–M System Type[c]				# R–M loci
						IA	IB	IC	ID	IE	IF	U	IIIA	IIIB			I	II	III	IV	
Archaea	Crenarchaeota	*Acidilobus*	1	1	1.5	+	–	–	–	–	–	–	–	+	6	103	–	+	–	–	2
		Caldisphaera	1	1	1.5	–	–	–	–	–	–	–	–	–	0	0	+	+	–	–	11
		Acidianus	1	1	2.1	+	–	–	–	–	–	–	+	–	6	123	–	+	–	–	2
		Metallosphaera	2	2	1.8	+	–	–	–	–	–	–	+	–	2–5	282–373	–	+	–	–	2
		Sulfolobus	4	17	2.0–2.9	+	–	–	+	–	–	–	–	+	2–6	54–457	+	+	–	–	30
		Vulcanisaeta	2	2	2.3	+	–	–	–	–	–	–	+	+	4–9	62–120	–	+	–	–	2
	Euryarchaeota	*Ferroplasma*	1	1	1.9	–	+	–	–	–	–	–	–	–	2	20	+	+	+	–	10
		Picrophilus	1	1	1.5	–	–	–	+	–	–	–	+	–	4	116	+	+	–	+	3
		Thermoplasma	2	2	1.5–1.6	–	–	–	–	–	–	–	+	–	1–2	34–46	+	+	+	–	19
		Aciduliprofundum	2	2	1.4–1.5	+	+	–	–	–	–	–	+	–	2–2	21–23	+	+	+	–	8
Bacteria	Acidobacteria	*Acidobacterium*	*1*	*1*	4.1	–	–	+	–	–	–	–	–	–	1	23	+	+	+	–	4
		Granulicella	2	2	4.3	–	–	–	–	–	–	–	–	–	0	0	+	+	–	–	4
		Terriglobus	2	2	5.0–5.2	–	–	–	–	–	–	–	–	–	0	0	–	+	–	–	6
		Solibacter	1	1	9.9	–	–	–	–	–	–	–	–	–	0	0	–	+	–	–	1
		Koribacter	1	1	5.5	–	–	–	–	–	–	–	–	–	0	0	+	+	–	–	5
	Actinobacteria	*Acidimicrobium*	1	1	2.1	–	–	–	–	+	–	–	–	–	2	67	–	–	+	–	1
	Aquificae	*Hydrogenobaculum*	1	3	1.5	–	+	–	–	–	–	–	–	–	1–6	29–69	+	+	–	–	5
	Firmicutes	*Alicyclobacillus*	1	2	3.0–3.1	+	–	+	–	–	–	–	–	+	3–4	46–58	–	+	+	–	23
		Kyrpidia	1	1	3.4	–	+	–	–	–	–	–	–	+	4	418	ND	ND	ND	ND	ND
		Desulfosporosinus	1	1	4.9	–	–	+	–	–	–	–	–	–	1	34	+	+	–	+	7
		Sulfobacillus	1	2	3.5	–	+	–	–	–	–	–	–	–	2–2	14–68	–	+	+	–	20
	Nitrospirae	*Leptospirillum*[a]	2	2	2.4 –2.5	–	–	–	–	+	–	–	–	–	1–3	4–157	+	+	+	–	11
	Proteobacteria	*Acidiphilium*	1	1	3.7	–	–	–	–	–	–	–	–	–	0	0	–	+	–	–	8
		Acidithiobacillus	3	5	2.8–3.2	–	–	–	–	–	–	+	–	–	0–2	0–42	+	+	+	–	52
	Verrucomicrobia	*Methylacidiphilum*			2.3	–	–	+	–	–	–	–	–	+	5	19	+	+	–	+	6

Only complete genome have been considered with the exception of the leptospirilli[a].
[a]Data correspond to draft genomes derived from the Iron Mountain AMD.
[b]Data were obtained from CRISPR Database (http://http://crispr.u-psud.fr/crispr/) and CRISPI (http://crispi.genouest.org).
[c]Data were obtained from ReBase Database (http://rebase.neb.com/rebase/rebase.html).

sequences. Cognate methylases, in turn, protect the hosts DNA from the action of the cutting enzyme by chemically modifying the target sequences, thus achieving an accurate distinction between native and exogenous DNA. Overall, the RM systems limit the frequency of HGT and increase the resistance of cells to viruses. Four types of restriction systems (I–IV) are currently recognized, on the basis of subunit composition, NTP requirement and cleavage mechanism. Information stored in the ReBase database reveals that 95% of sequenced bacteria and nearly 100% of archaea encode RM systems, with type II RM systems being by far the most common. This general tendency is conserved in acidophiles (Table 12.3). Despite their widespread occurrence and high diversity, few RM systems have been characterized so far in acidophiles. These include the *Tp. acidophilus* ThaI system and SuaI and SuiI systems from *S. acidocaldarius* and *S. islandicus*, respectively. While all three systems are capable of DNA restriction, their cognate methylases await discovery and evaluation. Multiple RM-systems populate the genomes of acidophilic bacteria (e.g. >50 in sequenced acidithiobacilli) and several of these have been shown to map to MGEs (e.g. Acuña *et al.*, 2013; van Zyl *et al.*, 2008a). Studies of RM-systems of acidophiles in general are required in order to overcome some of the barriers to genetic manipulation that many of these models impose.

Outlook

Although acidophiles can be differentiated from neutrophiles and other extremophiles in many aspects of their biology, their mobilome seems to be as diverse and untapped as that of other prokaryotes. Transmissible, integrative and translocative elements described in this chapter represent evidence of the active exchanges of genetic material that have taken place between acidophiles, and have likely had an impact on their adaptation, differentiation and/or niche expansion. Many of the genes being exchanged along with these MGEs belong to functional categories that have previously been shown to be over-represented in the mobilome of model systems and communities thriving in other econiches. In contrast, several other genes provide a clear ecological advantage under niche-specific (acidic) conditions. Several other genes encode hypothetical proteins whose function awaits further discovery. To date, very little insight into the potential advantages (e.g. gene acquisition, genome plasticity, etc.) and disadvantages (e.g. gene inactivation, genome instability, etc.) for cells carrying MGEs has been obtained for acidophiles. Also, the elucidation of the environmental conditions that enhance the proliferation and propagation or containment of these elements by defence systems, are relevant topics to be addressed to improve the understanding of the physiology and ecology of these environmentally and industrially relevant microorganisms. Many mechanistic details are likely to emerge from further studies. Both bioinformatic approaches and genetic experiments are required to uncover the dynamic interactions between the genome, the transcriptome and the mobilome. For this to be possible, larger sets of closely related genomes need to get sequenced and a more comprehensive toolbox for genetic manipulation of acidophiles needs to be developed. Both these challenging milestones have begun to be addressed in recent years and will shortly provide new opportunities to expand our knowledge and understanding of the acidophilic mobilome.

Acknowledgements

The authors gratefully acknowledge Fundación Ciencia y Vida, Universidad Andres Bello and the University of Stellenbosch for facilitating the undertaking of our studies. The authors also wish to thank FONDECYT (Grant ID 1140048 and 3140005) for financial support and numerous students that have been contributed to our research.

References

Acuña, L.G., Cárdenas, J.P., Covarrubias, P.C., Haristoy, J.J., Flores, R., Nuñez, H., Riadi, G., Shmaryahu, A., Valdés, J., Dopson, M., *et al.* (2013). Architecture and gene repertoire of the flexible genome of the extreme acidophile *Acidithiobacillus caldus*. PLoS One *8*, e78237.

Allen, E.E., Tyson, G.W., Whitaker, R.J., Detter, J.C., Richardson, P.M., and Banfield, J.F. (2007). Genome dynamics in a natural archaeal population. Proc. Natl Acad. Sci. U.S.A. *104*, 1883–1888.

Andersson, A.F., and Banfield, J.F. (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. Science *320*, 1047–1050.

Berkner, S., and Lipps, G. (2008). Genetic tools for *Sulfolobus* spp.: vectors and first applications. Arch. Microbiol. *190*, 217–230.

Berkner, S., Hinojosa, M.P., Prangishvili, D., and Lipps, G. (2014). Identification of the minimal replicon and the

origin of replication of the crenarchaeal plasmid pRN1. Microbiologyopen *3*, 688–701.

Blount, Z.D., and Grogan, D.W. (2005). New insertion sequences of *Sulfolobus*: functional properties and implications for genome evolution in hyperthermophilic archaea. Mol. Microbiol. *55*, 312–325.

Brügger, K., Torarinsson, E., Redder, P., Chen, L., and Garrett, R.A. (2004). Shuffling of *Sulfolobus* genomes by autonomous and non-autonomous mobile elements. Biochem. Soc. Trans. *32*, 179–183.

Bustamante, P., Covarrubias, P., Levicán, G., Katz, A., Tapia, P., Bonnefoy, V., Holmes, D.S., Quatrini, R., and Orellana, O. (2012). ICEAfe1, an actively excising integrative-conjugative element from the biomining bacterium *Acidithiobacillus ferrooxidans*. J. Mol. Microbiol. Biotech. *22*, 399–407.

Cabrejos, M.E., Zhao, H.L., Guacucano, M., Bueno, S., Levican, G., Garcia, E., Jedlicki, E., and Holmes, D.S. (1999). IST1 insertional inactivation of the *resB* gene: implications for phenotypic switching in *Thiobacillus ferrooxidans*. FEMS Microbiol. Lett. *175*, 223–229.

Cádiz, R., Gaete, L., Jedlicki, E., Yates, J., Holmes, D.S., and Orellana, O. (1994). Transposition of IST2 in *Thiobacillus ferrooxidans*. Mol. Microbiol. *12*, 165–170.

Chakravarty, L., Zupancic, T.J., Baker, B., Kittle, J.D., Fry, I.J., and Tuovinen, O.H. (1995). Characterization of the pTFI91-family replicon of *Thiobacillus ferrooxidans* plasmids. Can. J. Microbiol. *41*, 354–365.

Coram, N.J., van Zyl, L.J., and Rawlings, D.E. (2005). Isolation, sequence analysis, and comparison of two plasmids (28 and 29 kilobases) from the biomining bacterium *Leptospirillum ferrooxidans* ATCC 49879. Appl. Environ. Microbiol. *71*, 7515–7522.

Deane, S.M., and Rawlings, D.E. (2004). Plasmid evolution and interaction between the plasmid addiction stability systems of two related broad host-range IncQ-like plasmids. J. Bacteriol. *186*, 2123–2133.

Deane, S.M., and Rawlings, D.E. (2011). Two large, related, cryptic plasmids from geographically distinct isolates of *Sulfobacillus thermotolerans*. Appl. Environ. Microbiol. *77*, 8175–8180.

Dellas, N., Snyder, J.C., Bolduc, B., and Young, M.J. (2014). Archaeal Viruses: Diversity, Replication, and Structure. Annu. Rev. Virol. *1*, 399–426.

Dominy, C.N., Coram, N.J. , and Rawlings, D.E. (1998). Sequence analysis of plasmid pTF5, a 19.8-kb geographically widespread member of the *Thiobacillus ferrooxidans* pTF191-like plasmid family. Plasmid *40*, 50–57.

Dorrington, R.A., and Rawlings, D.E. (1990). Characterization of the minimum replicon of the broad-host-range plasmid pTF-FC2 and similarity between pTF-FC2 and the IncQ plasmids. J. Bacteriol. *172*, 5697–5705.

Drolet, M., and Lau, P.C.K. (1992). Mobilization protein-DNA binding and divergent transcription at the transfer origin of *the Thiobacillus ferrooxidans* pTF1 plasmid. Mol. Microbiol. *6*, 1061–1071.

Fütterer, O., Angelov, A., Liesegang, H., Gottschalk, G., Schleper, C., Schepers, B., Dock, C., Antranikian, G., and Liebl, W. (2004). Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. Proc. Natl. Acad. Sci. U.S.A. *101*, 9091–9096.

Garcillán-Barcia, M.P., and de la Cruz, F. (2013). Ordering the bestiary of genetic elements transmissible by conjugation. Mob. Genet. Elements *3*, e24263.

Gardner, M.N., and Rawlings, D.E. (2004). Evolution of compatible replicons of the related IncQ-like plasmids, pTC-F14 and pTF-FC2. Microbiology *150*, 1797–1808.

Gardner, M.N., Deane, S.M., and Rawlings, D.E. (2001). Isolation of a new broad-host-range IncQ-like plasmid, pTC-F14, from the acidophilic bacterium *Acidithiobacillus caldus* and analysis of the plasmid replicon. J. Bacteriol. *183*, 3303–3309.

Garrett, R.A., Shah, S.A., Erdmann, S., Liu, G., Mousaei, M., León-Sobrino, C., Peng, W., Gudbergsdottir, S., Deng, L., Vestergaard, G., *et al.* (2015). CRISPR-Cas adaptive immune systems of the *Sulfolobales*: unravelling their complexity and diversity. Life *5*, 783–817.

Goltsman, D.S.A., Denef, V.J., Singer, S.W., VerBerkmoes, N.C., Lefsrud, M., Mueller, R.S., Dick, G.J., Sun, C.L., Wheeler, K.E., Zemla, A., *et al.* (2009). Community genomic and proteomic analyses of chemoautotrophic iron oxidizing "*Leptospirillum rubarum*" (group II) and "*Leptospirillum ferrodiazotrophum*" (group III) bacteria in acid mine drainage biofilms. Appl. Environ. Microbiol. *75*, 4599–4615.

Goltsman, D.S., Dasari, M., Thomas, B.C., Shah, M.B., VerBerkmoes, N.C., Hettich, R.L., and Banfield, J.F. (2013). New group in the *Leptospirillum* clade: cultivation-independent community genomics, proteomics, and transcriptomics of the new species "*Leptospirillum* group IV UBA BS". Appl. Environ. Microbiol. *79*, 5384–5393.

González, C., Yanquepe, M., Cárdenas, J.P., Valdés, J., Quatrini, R., Holmes, D.S., and Dopson, M. (2014). Genetic variability of psychrotolerant *Acidithiobacillus ferrivorans* revealed by (meta)genomic analysis. Res. Microbiol. *65*, 726–734.

Greve, B., Jensen, S., Brügger, K., Zillig, W., and Garrett, R.A. (2004). Genomic comparison of archaeal conjugative plasmids from *Sulfolobus*. Archaea *1*, 231–239.

Guo, L., Brügger, K., Liu, C., Shah, S.A., Zheng, H., Zhu, Y., Wang, S., Lillestøl, R.K., Chen, L., Frank, J., *et al.* (2011). Genome analyses of Icelandic strains of *Sulfolobus islandicus*: model organisms for genetic and virus–host interaction studies. J. Bacteriol. *193*,1672–1680.

Holmes, D.S., Zhao, H.L., Levican, G., Ratouchniak, J., Bonnefoy, V., Varela, P., and Jedlicki, E. (2001). ISAfe1, an ISL3 family insertion sequence from *Acidithiobacillus ferrooxidans* ATCC 19859. J. Bacteriol. *183*, 4323–4329.

Holmes, D.S., Cárdenas, J.P., Valdés, J., Quatrini, R., Esparza, M., Osorio, H., Duarte, F., Lefimil, C., and Jedlicki, E. (2009). Comparative genomics begins to unravel the ecophysiology of bioleaching. Adv. Mat. Res. *71–73*, 143–150.

Kotze, A.A., Tuffin, I.M., Deane, S.M., and Rawlings, D.E. (2006). Cloning and characterization of the chromosomal arsenic resistance genes from *Acidithiobacillus caldus* and enhanced arsenic resistance on conjugal transfer of ars genes located on transposon TnAtcArs. Microbiology *152*, 3551–3560.

Levicán, G., Katz, A., Valdés, J., Quatrini, R., Holmes, D.S., and Orellana, O. (2009). A 300 kpb genome segment, including a complete set of tRNA genes, is dispensable

for *Acidithiobacillus ferrooxidans*. Adv. Mat. Res. *71–73*, 187–190.

Lipps, G. (2009). Molecular biology of the pRN1 plasmid from *Sulfolobus islandicus*. Biochem. Soc. Trans. *37*, 42–45.

López de Saro, F.J., Gómez, M.J., González-Tortuero, E., and Parro, V. (2013). The dynamic genomes of acidophiles. In Polyextremophiles: Life Under Multiple Forms of Stress, Seckbach, J., Oren, A., and Stan-Lotter, H. eds. (Springer, Dordrecht, Netherlands). pp.81–97.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., *et al.* (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Matcher, G.F., and Rawlings, D.E. (2009). The effect of the location of the proteic post-segregational stability system within the replicon of plasmid pTF-FC2 on the fine regulation of plasmid replication. Plasmid *62*, 98–107.

Mueller, R.S., Denef, V.J., Kalnejais, L.H., Suttle, K.B., Thomas, B.C., Wilmes, P., Smith, R.L., Nordstrom, D.K., McCleskey, R.B., Shah, M.B., *et al.* (2010). Ecological distribution and population physiology defined by proteomics in a natural microbial community. Mol. Syst. Biol. *6*, 374.

Orellana, L.H., and Jerez, C.A. (2011). A genomic island provides *Acidithiobacillus ferrooxidans* ATCC 53993 additional copper resistance: a possible competitive advantage. Appl. Microbiol. Biotechnol. *92*, 761–767.

Prangishvili, D., Albers, S.V., Holz, I., Arnold, H.P., Stedman, K., Klein, T., Singh, H., Hiort, J., Schweier, A., Kristjansson, J.K., *et al.* (1998). Conjugation in archaea: frequent occurrence of conjugative plasmids in *Sulfolobus*. Plasmid. *40*, 190–202.

Rawlings, D.E. (2005). The evolution of pTF-FC2 and pTC-F14, two related plasmids of the IncQ family. Plasmid *53*, 137–147.

Redder, P., She, Q., and Garrett, R.A. (2001). Non-autonomous mobile elements in the Crenarchaeon *Sulfolobus solfataricus*. J. Mol. Biol. *306*, 1–6.

Romano, C., D'Imperio, S., Woyke, T., Mavromatis, K., Lasken, R., Shock, E.L., and McDermott, T.R. (2013). Comparative genomic analysis of phylogenetically closely related *Hydrogenobaculum* sp. isolates from Yellowstone National Park. Appl. Environ. Microbiol. *79*, 2932–2943.

Schleper, C., Holz, I., Janekovic, D., Murphy, J., and Zillig, W. (1995). A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. J. Bacteriol. *177*, 4417–4426.

Simmons, S.L., Dibartolo, G., Denef, V.J., Goltsman, D.S., Thelen, M.P., and Banfield, J.F. (2008). Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. PLoS Biol. *6*, e177.

Stedman, K.M., She, Q., Phan, H., Arnold, H.P., Holz, I., Garrett, R.A., and Zillig, W. (2003). Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res. Microbiol. *154*, 295–302.

Talla, E., Hedrich, S., Mangenot, S., Ji, B., Johnson, D.B., Barbe, V., and Bonnefoy, V. (2014). Insights into the pathways of iron- and sulfur-oxidation, and biofilm formation from the chemolithotrophic acidophile *Acidithiobacillus ferrivorans* CF27. Res. Microbiol. *165*, 753–760.

Tapia, P., Flores, F.M., Covarrubias, P.C., Acuña, L.G., Holmes, D.S., and Quatrini, R. (2012). Complete genome sequence of temperate bacteriophage AcaML1 from the extreme acidophile *Acidithiobacillus caldus* ATCC 51756. J. Virol. *86*, 12452–12453.

Travisany, D., Cortés, M.P., Latorre, M., Di Genova, A., Budinich, M., Bobadilla-Fazzini, R.A., Parada, P., González, M., and Maass, A. (2014). A new genome of *Acidithiobacillus thiooxidans* provides insights into adaptation to a bioleaching environment. Res. Microbiol. *165*, 743–752.

Tuffin, I.M., de Groot, P., Deane, S.M., and Rawlings, D.E. (2005). An unusual Tn21-like transposon containing an ars operon is present in highly arsenic-resistant strains of the biomining bacterium *Acidithiobacillus caldus*. Microbiology *151*, 3027–3039.

Tuffin, I.M., Hector, S.B., Deane, S.M., and Rawlings, D.E. (2006). Resistance determinants of a highly arsenic-resistant strain of *Leptospirillum ferriphilum* isolated from a commercial biooxidation tank. Appl. Environ. Microbiol. *72*, 2247–2253.

Tyson, G.W., and Banfield, J.F. (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ. Microbiol. *10*, 200–207.

Valdés, J., Cárdenas, J.P., Quatrini, R., Esparza, M., Osorio, H., Duarte, F., Lefimil, C., Sepulveda, R., Jedlicki, E., and Holmes, D.S. (2010). Comparative genomics begins to unravel the ecophysiology of bioleaching. Hydrometallurgy *104*, 471–476.

Wang, H., Peng, N., Shah, S.A., Huang, L., and She, Q. (2015). Archaeal extrachromosomal genetic elements. Microbiol. Mol. Biol. Rev. *79*, 117–152.

Yates, J.R., and Holmes, D.S. (1987). Two families of repeated DNA sequences in *Thiobacillus ferrooxidans*. J. Bacteriol. *69*, 1861–1870.

You, X.Y., Guo, X., Zheng, H.J., Zhang, M.J., Liu, L.J., Zhu, Y.Q., Zhu, B., Wang, S.Y., Zhao, G.P., Poetsch, A., *et al.* (2011a). Unraveling the *Acidithiobacillus caldus* complete genome and its central metabolisms for carbon assimilation. J. Genet. Genomics. *38*, 243–252.

You, X.Y., Liu, C., Wang, S.Y., Jiang, C.Y., Shah, S.A., Prangishvili, D., She, Q., Liu, S.J., and Garrett, R.A. (2011b). Genomic analysis of *Acidianus hospitalis* W1 a host for studying crenarchaeal virus and plasmid life cycles. Extremophiles *15*, 487–497.

Zhang, M.J., Jiang, C.Y., You, X.Y., and Liu, S.J. (2014). Construction and application of an expression vector from the new plasmid pLAtc1 of *Acidithiobacillus caldus*. Appl. Microbiol. Biotechnol. *98*, 4083–4094.

van Zyl, L.J., Deane, S.M., and Rawlings, D.E. (2003). Analysis of the mobilization region of the broad-host-range IncQ-like plasmid pTC-F14 and its ability to interact with a related plasmid, pTF-FC2. J. Bacteriol. *185*, 6104–6111.

van Zyl, L.J., Deane, S.M., Louw, L.A., and Rawlings, D.E. (2008a). Presence of a family of plasmids (29 to 65 kilobases) with a 26-kilobase common region in different

van Zyl, L.J., Deane, S.M., Louw, L.A., and Rawlings, D.E. (2008a). Presence of a family of plasmids (29 to 65 kilobases) with a 26-kilobase common region in different strains of the sulfur-oxidizing bacterium *Acidithiobacillus caldus*. Appl. Environ. Microbiol., *74*, 4300–4308.

van Zyl, L.J., van Munster, J.M., and Rawlings, D.E. (2008b). Construction of arsB and tetH mutants of the sulfur-oxidizing bacterium *Acidithiobacillus caldus* by marker exchange. Appl. Environ. Microbiol. *74*, 5686–5694.