Analysis of defence systems and a conjugative IncP-1 plasmid in the marine polyaromatic hydrocarbons-degrading bacterium *Cycloclasticus* sp. 78-ME.

Authors:

Michail M. Yakimov,[1]* Francesca Crisafi,[1] Enzo Messina,[1] Francesco Smedile,[1] Anna Lopatina,[2] Renata Denaro,[1] Dietmar H. Pieper,[3] Peter N. Golyshin,[4] and Laura Giuliano[1]

[1] *Institute for Coastal Marine Environment, CNR, Spianata S.Raineri 86, 98122 Messina, Italy.*

[2] *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel*

[3] *Microbial Interactions and Processes Research Group, HZI – Helmholtz Centre for Infection Research, Inhoffenstraße 7, D-38124 Braunschweig, Germany*

[4] *School of Biological Sciences, Bangor University, ECW Bldg Deiniol Rd, Bangor, Gwynedd LL57 2UW, United Kingdom.*

* For correspondence: E-mail michail.yakimov@iamc.cnr.it; Tel. (+39) 090 6015437; Fax (+39) 090 669007.

Keywords: Obligate marine hydrocarbonoclastic bacteria; polyaromatic hydrocarbons; prokaryotic defence systems; degradation; Mediterranean Sea

Running title: Genomic defence systems in *Cycloclasticus*

Summary

Marine prokaryotes have evolved a broad repertoire of defence systems to protect their genomes from lateral gene transfer including innate or acquired immune systems and infection-induced programmed cell suicide and dormancy. Here we report on the analysis of multiple defence systems present in the genome of the strain *Cycloclasticus* sp. 78-ME isolated from petroleum deposits of the tanker *"Amoco Milford Haven"*. *Cycloclasticus* are ubiquitous bacteria globally important in polyaromatic hydrocarbons degradation in marine environments. Two "defence islands" were identified in 78-ME genome: the first harbouring CRISPR-Cas with toxin-antitoxin system, while the second was composed by an array of genes for toxin-antitoxin and restriction-modification proteins. Among all identified spacers of CRISPR-Cas system only seven spacers match sequences of phages and plasmids. Furthermore, a conjugative plasmid p7ME01, which belongs to a new IncP-1θ ancestral archetype without any accessory mobile elements was found in 78-ME. Our results provide the context to the co-occurrence of diverse defence mechanisms in the genome of *Cycloclasticus* sp. 78-ME, which protect the genome of this highly specialized PAH-degrader. This study contributes to the further understanding of complex networks established in petroleum-based microbial communities.

Introduction

Recent landmark studies of environmental DNA using next-generation sequencing showed that bacteriophages are by far the most abundant and genetically diverse biological entities in marine habitats (Edwards and Rohwer, 2005; Suttle, 2007; Kristensen *et al*., 2010). To withstand the constant exposure to marine viruses, microorganisms have evolved a broad repertoire of defence systems, sometimes at the expense of allocating substantial resources and genomic space (Labrie *et al*., 2010; Stern and Sorek, 2011; Makarova *et al*., 2011a). Although these multiple defence systems (MDS) are very diverse across different prokaryotic life forms, they can be attributed to one of the three general functional categories: (i) innate immunity systems that are based on self-nonself discrimination of foreign and host genomes, (ii) acquired immunity systems that are based on adaptive RNA-based immunity against foreign genetic elements, such as viruses and plasmids, and (iii) suicidal systems that cause programmed cell death or dormancy induced by infection, preventing its further spread (Iranzo *et al*., 2015). Innate immunity among other forms involve restriction-modification systems (RMS), which protect prokaryotic cells from heterologous DNA through cleavage of unmodified foreign DNA molecules by the restriction component of RMS, while a host DNA methylated by the modification component of RMS

remains intact. Acquired (adapted) immunity is facilitated by the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas systems. Usually this system relies on a CRISPR array, a segment of DNA containing series of short identical sequences (repeats) separated by unique sequences of about the same length (spacers), and CRISPR-associated *cas* genes that encode a multifunctional protein complex. CRISPR array is transcribed and processed into individual small CRISPR RNAs (crRNAs). Mature crRNAs in complex with Cas proteins are directed to foreign complementary nucleic acids and base pair with them, which results in target DNA or RNA degradation (Deveau *et al*., 2010). The suicidal systems, such as toxin-antitoxin systems (TAS), are widespread in bacteria and archaea and function through stress-induced cell suicide or dormancy (Makarova *et al*., 2009; Blower *et al*., 2011). The TAS are based on the “poison-antidote” principle and in most cases consist of two genes, which encode, respectively, a toxin (either a protein perforating the cell membrane or an mRNA-cleaving endonuclease) and an antitoxin (either a small RNA that prevents toxin gene translation or a protein that forms an inactive complex with the toxin). Under normal conditions, the toxin is maintained in an inactive state via interaction with the antitoxin gene product. Various stresses, including viral infection, inactivate the antitoxin and thus unleash the toxin, which either kills the affected

cells or induces their dormancy, restricting the impact of the infection (Buts *et al*., 2005). Most of these defence mechanisms are widely distributed across the prokaryotic world, and genomes of free-living bacteria and archaea typically encode multiple defence systems of different classes, which form so-called genomic “defence islands” (DIs) (Makarova *et al*., 2011b).

Here, we analysed multiple defence systems and conjugative IncP-1 plasmid present in the genome of obligate marine bacterium *Cycloclasticus* sp. 78-ME, isolated from petroleum deposits of the sunken tanker *Amoco Milford Haven* (Mediterranean Sea) (Messina *et al*., 2016).

Results and discussion

Biodegradation potential of Cycloclasticus sp. *78-ME*

Cycloclasticus sp. 78-ME is a Gram-negative, obligate aerobic, marine gammaproteobacterium isolated from the tar residues disposed on the seabed at the wreck site of the supertanker *“Amoco Milford Haven”* (08°42.086’E, 44°22.242’N, 78 m depth). This accident, which happened close to the Genoa-Voltri coastline (Mediterranean Sea) in April 1991, released to the sea approximately 30,000 metric tons of heavy crude oil

and was considered as one of the top-ten oil spills in the human history. A mixture of phenanthrene : pyrene was used in the enrichment experiments as the only carbon sources (Messina *et al*., 2016). In addition to phenanthrene and pyrene, strain 78-ME was able to use naphthalene, methylnaphthalene, 2,6-dimethylnaphthalene, biphenyl, fluorene, acenaphthene, dibenzofuran, dibenzothiophene and anthracene as single carbon sources for growth. Currently, strain 78-ME is the second known *Cycloclasticus* strain capable to uptake pyrene, as the only carbon source (Lai *et al*., 2012a). Noteworthy, in the presence of various PAHs with two-four condensed rings, *Cycloclasticus* sp. 78-ME is capable to transform benz[α]pyrene (Fig. S1). Such striking PAH-degradation capabilities imply the presence of a very sophisticated enzymatic machinery. Indeed, genome analysis of strain 78-ME revealed the presence of 72 different enzymes belonging to four classes of ring-cleavage dioxygenases (Table S1). Among them, 20 genes encoding α-subunits of Rieske non-heme iron oxygenases with 2 of the 8 subunits belong to the biphenyl and are most closely related to naphthalene dioxygenases of *Proteobacteria* (cluster XXIV, Duarte *et al*. 2014).

Analysis of two "defence islands" in Cycloclasticus *sp. 78-ME genome*

Genome innovation and evolution in prokaryotes is *inter alia* dependent on the acquisition of the DNA from external sources, the process generally termed 'lateral gene transfer' (Ochman *et al*., 2000; Frost *et al*., 2005). Nevertheless, an acquisition of foreign genetic information does frequently lead to the disruption and inactivation of genetic determinants of the host organism at the insertion location. With a high probability this error-generation process can be lethal for organisms with minimized and streamlined genomes, such as that of *Cycloclasticus* sp. 78-ME. To protect the integrity of their genomes and withstand permanent extensive exposure to exogenous DNA, the prokaryotic organisms have evolved defence systems, which are typically clustered in the "defence islands" (DIs) (Makarova *et al*., 2011b; 2013). Several examples of DIs were found in the genome of *Cycloclasticus* sp. 78-ME.

One of them contained CRISPR-Cas and TAS systems. The pair of genes (CYCME_2159 and CYCME_2160) was predicted as a toxin-antitoxin pair. This DI is located upstream of the *cas* operon of the type I-E CRISPR-Cas system and consists of two genes, the *hipA* (hip is for "high persistence") encoding a toxin (the closest homologue is HipA protein from *Syntrophus aciditrophicus*, e-value 1e-140) and *hipB* encoding an antitoxin (the closest homologue is XRE family transcriptional regulator of

zetaproteobacterium TAG-1, e-value 2e-24). The products of *hipA* and *hipB* genes are recognized as a major factor involved in persistence to a wide variety of stresses (Wen *et al*., 2014), biofilm formation (Zhao *et al*., 2013) and survival during long-term stationary phase (Kawano and Mori, 2009).

Another 28 kbp-long DI was predicted with IslandViewer3 software (Dhillon *et al*., 2015). It includes 26 ORFs, most of which are clearly mobilome genes such as integrases, plasmid-like integrated elements, and genes of defence systems, such as predicted TAS, RMS and genes encoding virulence proteins (Fig. 1). One of the ORFs of the 28 kbp-long genomic island (CYCME_2453) contains a nucletidyl-transferase domain DUF1814, which was shown to be a signature domain of a widespread superfamily of toxins of type II TAS with unknown mechanism of toxicity (Sberro *et al*., 2013). Interestingly, the DUF1814 domain was also documented in AbiG – a two-gene system involved in abortive infection, suggesting that it might be involved in anti-phage defence (Makarova *et al*., 2011b; Sberro *et al*., 2013). CYCME_2454 located upstream of the toxin gene CYCME_2453 contains DUF2893 domain with unknown function, and the HHpred analysis (http://toolkit.tuebingen.mpg.de/hhpred#) revealed a helix-turn-helix domain. Based on its two-gene nature, CYCME_2453 and CYCME_2454 can be suggested as a novel pair of

type II TAS (shown in yellow in Fig. 1). Among other defence-related ORFs were CYCME_2461, CYCME_2464 and CYCME_2465 (shown in red in Fig. 1) encoding HsdR, HsdS and HsdM, correspondently. All of them are well known components of RMS of type I. Some of the proteins encoded in this DI (shown in white in Fig. 1) have no homology to proteins with known functions or predicted conserved domains and thus could be suggested as a “dark matter” encoding yet unexplored defence (or defence-irrelevant) proteins (Makarova *et al.*, 2014).

Analysis of CRISPR-Cas system in Cycloclasticus *sp. 78-ME genome*

CRISPR-Cas system is a prokaryotic adaptive immune system against foreign genetic elements such as viruses and plasmids (Makarova *et al.*, 2006, Makarova *et al.*, 2011a, Barrangou *et al.*, 2007). Consisting of stretches of interspaced repetitive DNA fragments and associated cas genes, CRISPR-Cas has been revealed as a unique defence system in prokaryotes, which recognizes fragments of nucleic acid of foreign origin, like phages and plasmids, and degrades it using a CRISPR-associated protein complex (Deveau *et al.*, 2010). A 15,500-bp long DNA fragment containing CRISPR-Cas system was found in *Cycloclasticus* sp. 78-ME using Pilercr v1.02 (Edgar, 2007). Seven *cas* genes were

detected: *cas3*, *cse1*, *cse2*, *cse4/cas7*, *cas5*, *cse3*, *cas1*, *cas2*, and the CRISPR-cassette made up by 116 spacers and 117 repeat sequences (5'-GTGTTCCCCACAAGCGTGGGGATGAACCG-3') (Fig. 2 and Table S2a). Noteworthy, this palindromic sequence was 97% identical to the repeat sequence detected in a CRISPR cassette of hydrogenotrophic methanogenic euryarchaeon Candidatus "*Methanosphaerula palustris*" E1-9c^{T} (Cadillo-Quiroz *et al.*, 2015). Following the current classification, CRISPR-Cas system of 78-ME was affiliated to I-E (*E. coli*) or CASS2 subtype, widespread in *Proteobacteria* (Makarova *et al.*, 2011a). Cas proteins showed 63-80% amino acid similarities to corresponding proteins detected in gammaproteobacteria *Cronobacter sakazakii* 701 (Cas2), *Escherichia coli* str. K-12 substr. MG1655 (Cas1), *Methylobacter tundripaludum* SV96 (Cse3) and *Methylomicrobium album* BG8 (from Cas5 to Cas3). As postulated elsewhere (Barrangou *et al.*, 2007; Garneau *et al.*, 2010; Snyder *et al.*, 2010; Anderson *et al.*, 2011; Makarova *et al.*, 2011a), during viral or plasmid infection new spacers can be acquired by active CRISPR-Cas systems in the CRISPR array, thus providing a genetic record of coevolution of host and its predators. To obtain information on the origin of foreign genetic elements used by *Cycloclasticus* sp. 78-ME for

acquired immunity, sets of spacer sequences of 78-ME were manually analysed as described in Supporting Information.

Twenty-nine spacers matched environmental sequences from the blast-env db (maximum 2 mismatches between a spacer and a protospacer allowed). Obtained environmental sequences were further analysed using BlastX algorithm (Table S2a and Table S2b), whereas 6 out of 29 sequences had no homologs in NR database. The remaining 12 spacers have homologies in bacterial (eight spacers) or viral (four spacers) genomes. The fact that only four out of 116 spacers analysed have viral origin is in coincidence with the fact that the vast majority of marine phages remains unknown and underrepresented in nt, env_nt, and wgs databases (Kristensen *et al.*, 2010). Two of these spacers (n° 8 and n° 108 that differ only in one nucleotide) matched a sequence encoding endolysin of uncultured Mediterranean phage uvMED, detected in metagenomic fosmid library of the 0.2-5 µm-size plankton, collected from the deep chlorophyll maximum (DCM) (Mizuno *et al.*, 2013). Noteworthy, many obligate marine hydrocarbonoclastic microorganisms, including PAH-degraders, were recently isolated as associated to microphytoplakton (Gutierrez *et al*., 2012; 2013). The spacer n° 45 matched a sequence encoding a portal protein of uncultured Mediterranean phage uvMED from the same library

(Mizuno *et al.*, 2013). The spacer n° 47 (Table S2b) matched a bacteriophage head-to-tail joining protein (pfam12236), previously detected in the genome of petroleum-degrading *Thalassospira* sp. HJ (Kiseleva *et al.*, 2015). Additionally, four protospacers targeting by spacers of the 78-ME CRISPR-Cas system were detected in viral fraction of British Colombia Bay metagenomes (Angly *et al.*, 2006), four were recovered from marine metagenome assembly TARA project_(111_DCM_0.22-3) (Karsenti *et al.*, 2011).

CRISPR interference in type I-E systems requires a functional protospacer adjacent motif (PAM) AWG, located upstream of the protospacer sequence. As shown above, 12 spacers homologous to environmental database entries contained an AAG trinucleotide on their 5' flanking region (Table S2b, Pam 5' column). PAM seem to have a fundamental role in recognition of invading elements, triggering the operation of the CRISPR system, and new spacer uptake process (Mojica *et al.*, 2009). So far many of these spacers should be "real" or active spacers utilizable for foreign elements recognition, probably appertaining to yet uncharacterized phages or plasmids.

The consequences of a rapid evolution and niche-specificity of defence systems in prokaryotes are their frequently observed patchy phyletic distributions. In particular, bacterial or archaeal strains that are otherwise closely related often differ in the content of

defence systems (Makarova *et al.*, 2011a; 2011b; Iranzo *et al.*, 2015). Corroborating with this statement, *Cycloclasticus pugetii* PS-1, isolated from coastal waters in Puget Sound (North Pacific Ocean; Dyksterhouse *et al.*, 1995), possesses CRISPR-Cas system totally different from that the strain 78-ME while sharing more than 98% average nucleotide identity with it (Supporting Information). A 14 kbp-long genome fragment is composed of 6 *cas* genes (*cas1*, *cas3*, *csy1*, *csy2*, *csy3* and *cas6f*) and a CRISPR array, and belongs to I-F (Ypest) or CASS3 subtype. Compared to 78-ME, the PS-1 CRISPR region is much shorter and is formed by 29 spacer sequences and 30 short palindromic repeats (5'-GTTCACTGCCGCACAGGCAGCTTAGAAA-3'). This sequence is identical to repeat sequences detected in *Photorhabdus temperata* subsp. *thracensis* strain DSM 15199 (Kwak and Shin, 2015). All these elements have nothing in common with corresponding structures of 78-ME CRISPR-Cas system, suggesting that acquired immunity of two *Cycloclasticus* strains could have been adapted to yet unexplored pool of geographically-specific viruses and other mobile DNA elements. Noteworthy, using updated (at January 2016) env_nt database we found that the 78-ME CRISPR matched with environmental DNA sampled in the Bizerte lagoon located in Northern Tunisia (37°16'08.9"N, 9°53'20.1"E; Mediterranean Sea) (Bargiela *et al.*, 2015). This site is a subject to petroleum

pollution, determined by activity of adjacent oil refinery, and likely represents an optimal ecological niche for many obligate marine hydrocarbonoclastic microorganisms, including PAH-degrading *Cycloclasticus*. Interestingly, the obtained Bizerte environmental DNA (scaffolds AZII01000540-AZII01000544, see Fig. 2) contained portions of CRISPR arrays with repeat sequence identical to that of 78-ME, moreover, from 92 spacers found in these scaffolds 44 were identical to those present in 78-ME CRISPR array. Previously, we have reported on ubiquitous distribution in the Mediterranean Sea of *Alteromonas macleodii* AltDE (stands for “Deep Ecotype”) strains harbouring identical CRISPR-Cas systems (Smedile *et al.*, 2013). In present case, the *Cycloclasticus* CRISPR-Cas systems of Mediterranean Sea seem also remarkably static. One of the possible explanations is that in spite of the postulated huge diversity of marine phages, petroleum-contaminated water masses and sediments of Mediterranean Sea are characterized by a relative stability and uniformity of environmental parameters and are likely to sustain rather uniform viral communities. A further confirmation of these results is given by the analysis of the spacers detected in Bizerte environmental scaffolds different from *Cycloclasticus* 78-ME. Although they target organisms that are different from that seen for *Cycloclasticus* 78-ME CRISPR-Cas systems, the corresponding protospacer were detected in the same environments that

hold 78-ME protospacers (Table S2a). Thus, one could imagine that Mediterranean *Cycloclasticus* strains are only resistant to such niche-specific phages.

245

Cycloclasticus *sp. 78-ME harbors a conjugative plasmid of a new IncP-1θ ancestral archetype without accessory mobile elements*

As recently reported (Messina *et al*., 2016), the genome of *Cycloclasticus* sp. 78-ME consists of two circular replicons: the 2,613,078 bp chromosome (G+C content of 41.84%) and the plasmid p7ME01 of 42,347 bp (G+C content of 53.28%). %). The type strain *Cycloclasticus pugetii* PS-1 has only the 2,383,924 bp chromosome. Since this is the first finding of a naturally-occurring plasmid in *Cycloclasticus* strains, we analysed this plasmid in more details. The complete sequence of p7ME01 revealed that its backbone is very similar to conjugative plasmids of the incompatibility group P-1 (IncP-1). This group of plasmids is an example of highly potent, self-transmissible DNA molecules with a complicated regulatory circuit, which utilize very efficient strategies for stable maintenance in almost all Gram-negative bacteria. In addition to their wide replication range they can even mobilize different “shuttle vectors” to Gram-positive bacteria, cyanobacteria, or even eukaryotic organisms such as yeasts (Sen *et al*., 2013). Recent studies have provided

evidence that IncP-1 plasmid maintenance mechanisms have a lot in common with the systems for chromosome segregation in bacteria (Adamczyk and Jagura-Burdzy, 2003). All known plasmids of this group possess IncP-1-specific backbone modules for their replication, stable inheritance and conjugative transfer. Overwhelming majority of the IncP-1 plasmids has at least one accessory gene or mobile element encoding either degradation of xenobiotic compounds or resistance to antibiotics or heavy metals. These accessory elements have been postulated to be acquired during adaptive evolution of this group of plasmids (Heuer *et al*., 2004). Recently, ancestral IncP-1 plasmids harbouring only backbone archetype and lacking any of these typical accessory genes was identified in various alpha- and gammaproteobacteria (Popowska and Krawczyk-Balska, 2013).

Plasmid p7ME01 possesses a classical IncP-1 structure of backbone modules: two regions involved in plasmid conjugation (the *tra* and *trb* operons), a region carrying the genes for plasmid replication, and a region responsible for central control, stable inheritance and partitioning (Fig. 3). The origins of vegetative replication (*oriV*) and plasmid transfer (*oriT*) are also present in p7ME01. Analysis of the *oriV* nucleotide sequences revealed conserved IncP-1 features in this region. Following the classification of Adamczyk and Jagura-Burdzy (2003), the DNA segment providing *oriV* activity in p7ME01 is

approximately 650bp long and consists of nine 16-mers repeats called iterons, A+T-rich and G+C-rich regions. Iterons are organized in four groups: containing two, one, five, and one copy, respectively (Fig. 3). The biggest group of iterons n° 4 - n° 8 together with A+T-rich region form a cluster representing minimal replication origin activated by initiator protein TrfA. Annotation of the sequence data revealed that p7ME01 contains 52 ORFs and their localizations and predicted functions are presented in Supporting Information (Table S3). Forty-five ORFs correspond to well-conserved backbone modules (Fig. 3). Using approach of Norberg *et al*. (2011), we compared the concatenated backbone regions A, B and C of p7ME01 with corresponding regions of 23 IncP-1 plasmids retrieved from GenBank through BLAST and literature searches. As supported by phylogenetic analysis of the IncP-1 backbones, p7ME01 does belong to a novel clade, hereafter called θ (from the Greek word "θαλαθθα or *thalassa"*, meaning "sea") (del Castillo *et al*., 2013). Additionally to p7ME01, this novel IncP-1θ clade currently consists of three other closely related plasmids (Fig. 4). Noteworthy, all of them were found in marine biofilm-forming gammaproteobacteria: *Alcanivorax hongengensis* A-11-3 (Lai and Shao, 2012b), *Marinobacter adhaerens* HP15 (Gärdes *et al*., 2010) and *Methylophaga frappieri* JAM7 (Auclair *et al*., 2010; Villeneuve *et al*., 2013). Besides the type of habitat and capability of

biofilm-formation, these bacteria do not seem to have any significant common physiological features shared among themselves or with *Cycloclasticus* sp. 78-ME. None of the IncP-1θ plasmids carried by these marine bacteria were studied in detail, and their "accessory regions" remain uncharacterized. As shown on Fig. 3, apart from the genes encoding proteins of backbone modules, p7ME01 plasmid harbours nine cryptic ORFs, located in two "accessory regions": ORF1 and ORF2 are situated between *oriV* and *klcA*; ORF3-9 – between *trbV* and *traC* (Fig. 3 and Table S3). With exception of *Marinobacter adhaerens* plasmid pHP-42, neither of the ORFs located in the *trbV-traC* region of p7ME01 were present in any of IncP-1θ conjugative plasmids. An average G + C content of this region (46.0%) is far below than that of the entire p7ME01 plasmid (53.3%), indicating its different origin from the rest of ORFs. Additionally, the *trbV-traC* region does not contain any insertion sequences and thus does not appear to be a mobile element. These findings suggested that restriction site-associated repeat sequences detected within the proximity of *parA* might play a role in the insertion event of ORF3-9 into p7ME01 rather than in acquisition of mobile elements.

As it is well established, vegetative replication of the IncP-1 plasmids is accompanied by either post-segregational killing (*psk*) or multimer resolution (*mrs*)

systems providing a stable inheritance of plasmids in the host populations (Adamczyk and Jagura-Burdzy, 2003). While the *psk* systems are typically based on toxin-antitoxin mechanisms described above, the absence of the *mrs* systems caused catenation of circular molecules at each replication cycle leading to the so-called “dimer catastrophy” (Summers *et al*., 1993). We inspected p7ME01 and other closely related IncP-1θ plasmids for presence of these systems and could not identify any *psk* systems. In the absence of killing gene system, these plasmids seem to rely only on active partitioning as a stable inheritance function and thus, on the essential *mrs* system. Indeed, in the accessory region located between *trb* and *tra* operons we found two genes, which likely provide the *mrs* function (Fig. 3): the *parA* gene coding for an enzyme resolving plasmid multimers, and the *yacC* gene for an exonuclease, which likely converses concatenated plasmid dimers to the monomeric form.

Proposed defensive role of p7ME01 in Cycloclasticus *sp. 78-ME*

Due to physiological constraints of *Cycloclasticus* sp. 78-ME and its inability to grow on common organic compounds-rich media, conventional conjugative transfer/mating

20

experiments with this marine bacterium as a plasmid donor were hardly possible. Because of some similarities between the conjugation and competence-related DNA transfer machineries (Chen *et al*., 2005), we replaced conjugation experiment with the estimation of transformation rates of purified p7ME01 plasmid using naturally competent marine bacteria *Photobacterium angustum* ATCC 25915 as a recipient. The plasmid mobilisation rates were estimated to be 3.6×10^{8} per μg plasmid, which is roughly equivalent to transformation efficiency of 1.5×10^{-4} . This indicates that within marine microbial communities, p7ME01 can be easily taken up by cells possessing a natural competence.

The biological reason of such mobilization capability and ubiquitous distribution of p7ME01 plasmid is still unclear. In general, the IncP-1 plasmids without any accessory mobile elements are rarely found in microbial communities and to our knowledge there are only four other such IncP-1 plasmids known so far: pA1 (Harada *et al*., 2006), pBP136 (Kamachi *et al*., 2006), and abovementioned pHP-42 and an unnamed plasmid from *Alcanivorax hondengensis* A-11-3. Their existence may be inconsistent with the hypothesis that plasmids are maintained in bacterial communities because they confer one or several advantageous traits to their host, which are intrinsically unnecessary for usual growth and survival (Bergstrom *et al*., 2000). To analyse the relative costs of maintaining

the p7ME01 plasmid, we estimated its copy number per single cell of *Cycloclasticus* sp. 78-ME. Using the qPCR approach with *gyrA*- and plasmid–specific primers (Table S4), we found that p7ME01 is a low-copy number plasmid, whose quantity does not exceed 1.3 copies $cell^{-1}$ (Fig. S2). Thus, the energy costs of p7ME01 maintaining seem to be very small, but they are nevertheless not zero, and it is still unclear how this plasmid persists in *Cycloclasticus* sp. 78-ME and other hosts. As we mentioned above, the most evident explanation is that inheritance capacity of p7ME01 is high enough to overcome their cost and occasional segregational loss, which allows them to persist stable in microbial populations. Thus, the p7ME01 plasmid could be maintained as a parasitic genetic element. Alternatively, p7ME01 may provide some yet unknown advantage to the host. As it was suggested elsewhere (Ghigo, 2001), some of the IncP-1 plasmid backbone genes confer advantage to host organisms in biofilm development. Although further studies are needed to elucidate the role of IncP-1θ plasmids in biofilm formation, it is worth to notice, that all currently known marine bacteria harbouring these plasmids were described as the active biofilm-forming organisms. Under second assumption, the cryptic gene products (ORF1-7) provide as yet unknown benefits to *Cycloclasticus* sp. 78-ME and therefore the p7ME01 plasmid is maintained in its population.

Another possible hypothesis is that the p7ME01 plasmid can be maintained in cells as a part of defence system and is needed to suppress the incorporation of exogenous DNA, which may be beneficial for stability/integrity of the hosts' genome. Studying the capacity of *P. angustum* to take up and mobilise exogenous DNA in form of linearized plasmids, we realized that natural competence of p7ME01-carrying cells, was completely inhibited (Table 1). Noteworthy, the plasmids we used through this study (pGEM and pTA) have replicons that are different from and therefore compatible with that of p7ME01. A similar phenomenon of plasmid-host interference was recently described for naturally competent *Bacillus subtilis* cells after their acquisition of large conjugative plasmid pLS20 (Singh *et al*., 2012). At present, we can only speculate about the biological function of the p7ME01-mediated inhibition of competence. As far as development of competence has been reported to be associated with fitness costs for the host (Haijema *et al*., 2001), the inhibition of this development is energetically favourable, especially in case with oligotrophic lifestyle of *Cycloclasticus*. Alternatively, the self-defence mechanisms, based on suppression of possible recombination between p7ME01 and exogenous DNA, may facilitate the plasmid integrity. This assumption is in the context with the present study, i.e. observed inhibition of competence may be beneficial for genetic stability of backbone

plasmids and as a consequence, for the integrity of the host's genome. Thus, the p7ME01 plasmid can be attributed to a factor of acquired immunity.

380

Conclusion

Members of the genus *Cycloclasticus* are recognized as globally important polycyclic aromatic hydrocarbons (PAH)-degrading bacteria in marine ecosystems including shallow and deep-sea water and oceanic sediments (Yakimov *et al*., 2007; Staley, 2010). One of the 385remarkable features of all known *Cycloclasticus* strains is their highly specialized substrate specificity towards the PAH (Yakimov *et al*., 2007). All of them possess relatively small (about 2.6 Mb) and streamlined genomes, which are highly attenuated to basic physiological properties related to hydrocarbonoclastic lifestyle in oligotrophic marine environments. Four currently known genomes of *Cycloclasticus* strains isolated from different marine 390ecosystems all over the world, share more than 98 % of average nucleotide identity which suggests they all belong to the same species (Goris *et al*., 2007), and which appears to be a consequence of such "genome minimization and streamlining" (Lynch, 2006). Noteworthy, all *Cycloclasticus* genomes are significantly impoverished in either expansive "accessory genes" or "selfish" mobile genetic elements. This indirectly indicates the presence of

efficient multiple defence systems which suppress the acquisition of exogenous DNA and thus may be beneficial for stability/integrity of such minimized and streamlined genome of *Cycloclasticus* sp. 78-ME. Two “defence islands” were identified in its genome: one contained CRISPR-Cas and toxin-antitoxin system, while the second was composed of an array of genes for toxin-antitoxin and restriction-modification proteins. Among 116 spacers of CRISPR-Cas system only seven spacers matched phages and plasmids. Additionally, we showed that *Cycloclasticus* sp. 78-ME harbours a conjugative plasmid p7ME01 of a new IncP-1θ ancestral archetype, which likely suppresses the acquisition of exogenous DNA by this organism. Based on this finding, we suggested that the adaptive immunity of *Cycloclasticus* sp. 78-ME is linked with the acquisition of this plasmid.

Acknowledgements

We thank Domenico Maiolino and the Diving Center *Ecosfera* for skilful help during the tar sampling, Maurizio Catalfamo for excellent technical assistance in gas chromatography analysis. This work was supported by the European Commission’s program, under the MAGIC-PAH (FP7-KBBE-2009-3-5-01), MAMBA (FP7-KBBE-2009-2B-

25

226977), KILL•SPILL (FP7-OCEAN.2011-2-287589) and INMARE (H2020-BG-2014-2634486) Projects. Authors declare no conflict of interest.

References

Adamczyk, M., and Jagura-Burdzy, G. (2003) Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim Pol* **50**: 425–453.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Anderson, R.E., Brazelton, W.J., and Baross, J.A. (2011) Using CRISPRs as a metagenomic tool to identify microbial hosts of a diffuse flow hydrothermal vent viral assemblage. *FEMS Microbiol Ecol* **77**: 120–133.

Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., *et al.* (2006) The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.

Auclair, J., Lepine, F., Parent, S., and Villemur, R. (2010) Dissimilatory reduction of nitrate in seawater by a Methylophaga strain containing two highly divergent narG sequences. *ISME J* **4**: 1302–1313.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bargiela, R., Mapelli, F., Rojo, D., Chouaia, B., Tornés, J., Borin, S., *et al.* (2015) Bacterial population and biodegradation potential in chronically crude oil-contaminated marine sediments are strongly linked to temperature. *Sci Rep* **5**: 11651.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709–1712.

Bergstrom, C.T., Lipsitch, M., and Levin, B.R. (2000) Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**: 1505–1519.

Blower, T.R., Salmond, G.P., and Luisi, B.F. (2011) Balancing at survival's edge: the structure and adaptive benefits of prokaryotic toxin-antitoxin partners. *Curr Opin Struct Biol* **21**: 109–118.

Buts, L., Lah, J., Dao-Thi, M.H., Wyns, L., and Loris, R. (2005) Toxin-antitoxin modules as bacterial metabolic stress managers. *Trends Biochem Sci* **30**: 672–679.

Cadillo-Quiroz, H., Browne, P., Kyrpides, N., Woyke, T., Goodwin, L., Detter, C., et al. (2015) Complete Genome Sequence of *Methanosphaerula palustris* E1-9C^{T}, a

Hydrogenotrophic Methanogen Isolated from a Minerotrophic Fen Peatland. *Genome Announc* **3**: e01280-15.

Chen, I., Christie, P.J., and Dubnau, D. (2005) The ins and outs of DNA transfer in bacteria. *Science* **310**: 1456–1460.

del Castillo, C.S., Jang, H.B., Hikima, J., Jung, T.S., Morii, H., Hirono, I., *et al.* (2013) Comparative analysis and distribution of pP9014, a novel drug resistance IncP-1 plasmid from *Photobacterium damselae* subsp. *piscicida*. *Int J Antimicrob Agents* **42**: 10–18.

Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* **64**: 475–493.

Dhillon, B.K., Laird, M.R., Shay, J.A., Winsor, G.L., Lo, R., Nizam, F., *et al.* (2015) IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucl Acid Res* **43**: W104-108.

Duarte, M., Jauregui, R., Vilchez-Vargas, R., Junca, H. and Pieper, D.H. (2014) AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database* bau118.

Dyksterhouse, S.E., Gray, J.P., Herwig, R.P., Lara, J.C., and Staley, J.T. (1995) *Cycloclasticus pugetii* gen. nov., sp. nov., an aromatic hydrocarbon-degrading bacterium from marine sediments. *Int J System Bacteriol* **45**: 116–123.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Edgar, R.C. (2007) PILER-CR: fast and accurate identification of CRISPR repeats. *BMC Bioinformatics* **8**: 18.

Edwards, R.A., and Rohwer, F. (2005) Viral metagenomics. *Nat Rev Microbiol* **3**: 504-510.

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3**: 722–732.

Gärdes, A., Kaeppel, E., Shehzad, A., Seebah, S., Teeling, H., Yarza, P., *et al.* (2010) Complete genome sequence of *Marinobacter adhaerens* type strain (HP15), a diatom-interacting marine microorganism. *Stand Genomic Sci* **3**: 97–107.

Garneau, J.E., Dupuis, M.È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., *et al.* (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67–71.

Ghai, R., Martin-Cuadrado, A.B., Molto, A.G., Heredia, I.G., Cabrera, R., Martin, J., *et al.* (2010) Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J* **4**: 1154–1166.

Ghigo, J.M. (2001) Natural conjugative plasmids induce bacterial biofilm development. *Nature* **412**: 442–445.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P, and Tiedje, J.M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81-91.

Gutierrez, T., Nichols, P.D., Whitman, W.B. and Aitken, M.D. (2012) *Porticoccus hydrocarbonoclasticus* sp. nov., an aromatic hydrocarbon-degrading bacterium identified in laboratory cultures of marine phytoplankton. *Appl Environm Microbiol* **78**: 628-637.

Gutierrez, T., Green, D.H., Nichols, P.D., Whitman, W.B., Semple, K.T. and Aitken, M.D. (2013) *Polycyclovorans algicola* gen. nov., sp. nov., an aromatic-hydrocarbon-degrading marine bacterium found associated with laboratory cultures of marine phytoplankton. *Appl Environm Microbiol* **79**: 205-214.

30

Haijema, B.J., Hahn, J., Haynes, J., and Dubnau, D. (2001) A ComGA-dependent checkpoint limits growth during the escape from competence. *Mol Microbiol* **40**: 52–64.

Harada, K.M., Aso, Y., Hashimoto, W., Mikami, B., and Murata, K. (2006) Sequence and analysis of the 46.6-kb plasmid pA1 from *Sphingomonas* sp. A1 that corresponds to the typical IncP-1beta plasmid backbone without any accessory gene. *Plasmid* **56**: 11–23.

Heuer, H., Szczepanowski, R., Schneiker, S., Pühler, A., Top, E.M., and Schlüter, A. (2004) The complete sequences of plasmids pB2 and pB3 provide evidence for a recent ancestor of the IncP-1b group without any accessory genes. *Microbiology* **150**: 3591–3599.

Iranzo, J., Lobkovsky, A.E., Wolf, Y.I., and Koonin, E.V. (2015) Immunity, suicide or both? Ecological determinants for the combined evolution of anti-pathogen defence systems. *BMC Evolut Biol* **15**: 43.

Kamachi, K., Sota, M., Tamai, Y., Nagata, N., Konda, T., Inoue, T., *et al.* (2006) Plasmid pBP136 from *Bordetella pertussis* represents an ancestral form of IncP-1beta plasmids without accessory mobile elements. *Microbiology* **152**: 3477–3484.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J., *et al.* (2011) A holistic approach to marine eco-systems biology. *PLoS Biol* **9**: e1001177.

Kawano, H.H.Y., and Mori, H. (2009) Long-term survival of Escherichia coli lacking the HipBA toxin-antitoxin system during prolonged cultivation. *Biosci Biotechnol Biochem* **73:**117-123.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kiseleva, L., Garushyants, S.K., Briliute, J., Simpson, D.J., Cohen, M.F., and Goryanin, I. (2015) Genome Sequence of the Electrogenic Petroleum-Degrading *Thalassospira* sp. Strain HJ. *Genome Announc* **3**: e00483-15.

Kristensen, D.M., Mushegian, A.R., Dolja, V.V., and Koonin, E.V. (2010) New dimensions of the virus world discovered through metagenomics. *Trends Microbiol* **18**: 11–19.

Kwak, Y., and Shin, J.H. (2015) Complete genome sequence of *Photorhabdus temperata* subsp. *thracensis* 39-8(T), an entomopathogenic bacterium for the improved commercial bioinsecticide. *J Biotechnol* **214**: 115–116.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol* **8**: 317–327.

Lai, Q., Li, W., Wang, B., Yu, Z. and Shao, Z., 2012a. Complete genome sequence of the pyrene-degrading bacterium *Cycloclasticus* sp. strain P1. *J Bacteriol* **194**: 6677.

Lai, Q., and Shao, Z. (2012b) Genome sequence of the alkane-degrading bacterium *Alcanivorax hongdengensis* type strain A-11-3. *J Bacteriol* **194**: 6972.

Leplae, R., Lima-Mendez, G., and Toussaint, A. (2010) ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* **38**: D57–61.

Lynch, M. (2006) Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol* **60**: 327-349.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**: 7–33

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol Direct* **4**: 19.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., *et al.* (2011a) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**: 467–477.

Makarova, K.S., Wolf, Y.I., Snir, S., and Koonin, E.V. (2011b) Defence islands in bacterial and archaeal genomes and prediction of novel defence systems. *J Bacteriol* **193**: 6039–6056.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013) Comparative genomics of defence systems in archaea and bacteria. *Nucl Acid Res* **41**: 4360–4377.

Makarova, K.S., Wolf, Y.I., Forterre, P., Prangishvili, D., Krupovic, M., and Koonin, E.V. **(**2014) Dark matter in archaeal genomes: a rich source of novel mobile elements, defence systems and secretory complexes. *Extremophiles* **18**: 877–893.

Messina, E., Denaro, R., Crisafi, F., Smedile, F., Cappello, S., Genovese, M., *et al.* (2016) Genome sequence of obligate marine polycyclic aromatic hydrocarbons-degrading bacterium *Cycloclasticus* sp. 78-ME, isolated from petroleum deposits of the sunken tanker Amoco Milford Haven, Mediterranean Sea. *Mar Genomics* **25**: 11–13.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. **(**2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733–740.

Mizuno, C.M., Rodriguez-Valera, F., Garcia-Heredia, I., Martin-Cuadrado, A.B., Ghai, R. (2013) Reconstruction of novel cyanobacterial siphovirus genomes from Mediterranean metagenomic fosmids. *Appl Environ Microbiol* **79**: 688–695.

Norberg, P., Bergstrom, M., Jethava, V., Dubhashi, D., and Hermansson, M. (2011) The IncP-1 plasmid backbone adapts to different host bacterial species and evolves through homologous recombination. *Nat Commun* **2**: 268.

Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.

Ogata, H., Ray, J., Toyoda, K., Sandaa, R.A., Nagasaki, K., Bratbak, G., *et al.* (2011) Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *ISME J* **5**: 1143–1151.

Popowska, M., and Krawczyk-Balska, A. (2013) Broad-host-range IncP-1 plasmids and their resistance potential. *Front Microbiol* **4**: 44.

Rabus, R., Kube, M., Heider, J., Beck, A., Heitmann, K., Widdel, F., *et al.* (2005) The genome sequence of an anaerobic aromatic-degrading denitrifying bacterium, strain EbN1. *Arch Microbiol* **183**: 27–36.

Romine, M.F., Carlson, T.S., Norbeck, A.D., McCue, L.A., and Lipton, M.S. (2008) Identification of mobile elements and pseudogenes in the *Shewanella oneidensis* MR-1 genome. *Appl Environ Microbiol* **74**: 3257–3265.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., *et al.* (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.

Sberro, H., Leavitt, A., Kiro, R., Koh, E., Peleg, Y., Qimron, U. et al. (2013) Discovery of functional toxin/antitoxin systems in bacteria by shotgun cloning. *Mol Cell* **50:** 136-148.

Sen, D., Brown, C.J., Top, E.M. and Sullivan, J. (2013) Inferring the evolutionary history of IncP-1 plasmids despite incongruence among backbone gene trees. *Mol Biol Evol* **30**: 154–166.

Sevin, E.W., and Barloy-Hubler F. (2007) RASTA-Bacteria: a web-based tool for identifying toxin-antitoxin loci in prokaryotes. *Genome Biol* **8**: R155.

Singh, P.K., Ramachandran, G., Durán-Alcalde, L., Alonso, C., Wu, L.J., and Meijer, W.J. (2012). Inhibition of *Bacillus subtilis* natural competence by a native, conjugative plasmid-encoded comK repressor protein. *Environ Microbiol* **14**: 2812–2825.

Smedile, F., Messina, E., La Cono, V., Tsoy, O., Monticelli, L.S., Borghini, M., *et al.* (2013) Metagenomic analysis of hadopelagic microbial assemblages thriving at the deepest part of Mediterranean Sea, Matapan-Vavilov Deep. *Environ Microbiol* **15**: 167–182.

Snyder, J.C., Bateson, M.M., Lavin, M., and Young, M.J. (2010) Use of cellular CRISPR (clusters of regularly interspaced short palindromic repeats) spacer-based microarrays for detection of viruses in environmental samples. *Appl Environ Microbiol* **76**: 7251–7258.

Staley, J.T. (2010) Cycloclasticus: A genus of marine polycyclic aromatic hydrocarbon degrading bacteria. In: *Handbook of Hydrocarbon and Lipid Microbiology*, K.N. Timmis (ed.). Berlin, Germany: Springer-Verlag, pp. 1781–1786.

Stern, A., and Sorek, R. (2011) The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**: 43–51.

Summers, D.K., Beton, C.W., and Withers, H.L. (1993) Multicopy plasmid instability: the dimer catastrophe hypothesis. *Mol Microbiol* **8**: 1031–1038.

Suttle, C.A. (2007) Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**: 801–812.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37–43.

Villeneuve, C., Martineau, C., Mauffrey, F., and Villemur, R. (2013) *Methylophaga nitratireducenticrescens* sp. nov. and *Methylophaga frappieri* sp. nov., isolated from the biofilm of the methanol-fed denitrification system treating the seawater at the Montreal Biodome. *Int J Syst Evol Microbiol* **63**: 2216–2222.

Wen, Y.E.B., Felix, J., Elegheert, J., Vergauwen, B., Devreese, B., and Savvides, S.N. (2014) The bacterial antitoxin HipB establishes a ternary complex with operator DNA and phosphorylated toxin HipA to regulate bacterial persistence. *Nucl Acids Res* **42**: 10134-10147.

Zhao, J.Q.W., Li, M., Heijstra, B.D., Wang, S., Liang, Q., and Qi, Q. (2013) *Escherichia coli* toxin gene *hipA* affects biofilm formation and DNA release. Microbiology **159:** 633-640.

Yakimov, M.M., Timmis, K.N., and Golyshin, P.N. (2007) Obligate oil-degrading marine bacteria. *Curr Opin Biotechnol* **18**: 257–266.

Table 1. Transformation efficiency for Amp^R plasmids *pTA* and *pGEM* (pUC origin of replication, incompetence group A) in *P. angustum* wild type and *P. angustum* harboring p7ME01 plasmid. All transformations were done in triplicates.

Vectors	Strains	Tranformants, **CFU** μg^{-1} plasmid
pTA	*P. angustum wild type*	$1.5 \times 10^9 \pm 2.0 \times 10^6$
	P.angustum::p7ME01	0*
pGEM	*P. angustum wild type*	$8.7 \times 10^9 \pm 2.4 \times 10^6$
	P. angustum::p7ME01	0*

* no transformants were observed.

Figure legends

Figure 1. Genome organisation of 28 kb-long defense island in *Cycloclasticus* sp. 78-ME. Coding regions are shown by arrows indicating direction of transcription. Colors of the arrows represent different functional modules: toxin-antitoxin genes are shown in yellow; restriction-modification system genes are shown in red, integrases are shown in violet; ORFs encoding proteins with other functions are shown in grey, ORFs encoding proteins with no predicted function are shown in white.

Figure 2. Structures of CRISPR-Cas systems identified in genomes of *Cycloclasticus* sp. 78-ME and *Cycloclasticus pugetii* PS-1. Partial CRISPR-Cas system found in Bizerte environmental DNA is depicted for comparison. See text for further details on associated protein and repeat regions found.

Figure 3. Genetic map of plasmid p7ME01. Coding regions are shown by arrows indicating the direction of transcription. The positions of the origins of vegetative replication (*ori*V) and plasmid transfer (*ori*T) are marked with black-red circles. The region of *ori*V is shown

40

in more details above the map of the plasmid and does not contain interrupting mobile elements. Positions of G/C- and A/T-rich regions and iterons are shown by yellow ellipses and red boxes, respectively. The height of each base in sequence logo of p7ME01 iterons represents its conservation. The different functional modules of the plasmid are represented in different colours. In addition, the positions of insertion of mobile genetic elements and phenotypic markers in IncP-1 plasmids are indicated by grey sectors placed outside the plasmid map. Three concatenated backbone regions A, B and C were used for phylogenetic and signature analyses (Norberg *et al*., 2011).

Figure 4. Phylogenetic analysis showing the relationship of p7ME01 with other incP-1 plasmids of the IncP-1plasmid backbone. Maximum Likelihood tree was inferred from concatenated backbone regions A, B and C of 23 IncP-1 plasmids belonging to all currently recognized clades (Norberg *et al.,* 2011; del Castillo *et al*., 2013). Novel IncP-1θ clade is shaded in grey. Five IncP-1 plasmids without any accessory mobile elements known so far are highlighted in bold. Sequence of the uncharacterized plasmid MEALZ_p

665(FD082061) from *Methylomicrobum alcaliphilum* 20Z was used as an outgroup. The scale bar represents the probability of amino acid substitutions per site.