


# SCIENTIFIC REPORTS



OPEN

## Genomics of high molecular weight plasmids isolated from an on-farm biopurification system

Received: 12 February 2016

Accepted: 31 May 2016

Published: 20 June 2016

María C. Martini<sup>1</sup>, Daniel Wibberg<sup>2</sup>, Mauricio Lozano<sup>1</sup>, Gonzalo Torres Tejerizo<sup>1</sup>, Francisco J. Albicoro<sup>1</sup>, Sebastian Jaenicke<sup>2</sup>, Jan Dirk van Elsas<sup>3</sup>, Alejandro Petroni<sup>4</sup>, M. Pilar Garcillán-Barcia<sup>5</sup>, Fernando de la Cruz<sup>5</sup>, Andreas Schlüter<sup>2</sup>, Alfred Pühler<sup>2</sup>, Mariano Pistorio<sup>1</sup>, Antonio Lagares<sup>1</sup> & María F. Del Papa<sup>1</sup>

The use of biopurification systems (BPS) constitutes an efficient strategy to eliminate pesticides from polluted wastewaters from farm activities. BPS environments contain a high microbial density and diversity facilitating the exchange of information among bacteria, mediated by mobile genetic elements (MGEs), which play a key role in bacterial adaptation and evolution in such environments. Here we sequenced and characterized high-molecular-weight plasmids from a bacterial collection of an on-farm BPS. The high-throughput-sequencing of the plasmid pool yielded a total of several Mb sequence information. Assembly of the sequence data resulted in six complete replicons. Using *in silico* analyses we identified plasmid replication genes whose encoding proteins represent 13 different Pfam families, as well as proteins involved in plasmid conjugation, indicating a large diversity of plasmid replicons and suggesting the occurrence of horizontal gene transfer (HGT) events within the habitat analyzed. In addition, genes conferring resistance to 10 classes of antimicrobial compounds and those encoding enzymes potentially involved in pesticide and aromatic hydrocarbon degradation were found. Global analysis of the plasmid pool suggest that the analyzed BPS represents a key environment for further studies addressing the dissemination of MGEs carrying catabolic genes and pathway assembly regarding degradation capabilities.

Biopurification systems (BPS) were developed to mitigate the direct contamination of surface water bodies with pesticides. BPS operate as biofilter systems in which pesticides are removed from the wastewater by sorption and biodegradation in the filter matrix<sup>1</sup>. BPS receive high loads of pesticides at relatively high concentrations for long periods of time, thus creating a strong and long-term selective pressure for the development and growth of pesticide-tolerant or -degrading bacteria<sup>2</sup>. Despite the increasing application of on-farm BPS worldwide, information on the involved microbiology is still scarce. Exposure of the indigenous bacteria to mixtures of pollutants might have fostered adaptational responses *via* horizontally acquired mobile genetic elements (MGEs). Microbial activities that promote the occupancy of a particular ecological niche may be encoded on MGEs that can move across a microbial community. Accordingly, approaches targeting the mobilome<sup>3</sup> provide access to this yet-unknown genetic resource.

It is well known that horizontal genetic transfer (HGT) plays an important role in bacterial adaptation and evolution. Plasmids are significant contributors to HGT to a considerably extent<sup>4</sup>. However, knowledge on the incidence and diversity of plasmids in bacteria from different environments is still limited to date.

<sup>1</sup>Instituto de Biotecnología y Biología Molecular (IBBM), CONICET- Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calles 47 y 115 (1900) La Plata, Argentina.

<sup>2</sup>Center for Biotechnology (CeBiTec), Bielefeld University, Institute for Genome Research and Systems Biology, D-33615 Bielefeld, Germany. <sup>3</sup>Microbial Ecology, GELIFES, University of Groningen, The Netherland. <sup>4</sup>Servicio Antimicrobianos, Departamento de Bacteriología, Instituto Nacional de Enfermedades Infecciosas-ANLIS Dr. Carlos G. Malbrán, Buenos Aires, Argentina. <sup>5</sup>Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-Consejo Superior de Investigaciones Científicas (CSIC), 39011 Santander, Spain. Correspondence and requests for materials should be addressed to A.L. (email: lagares@biol.unlp.edu.ar) or M.F.D.P. (email: floppy@biol.unlp.edu.ar)

To broaden our view of the entire plasmid pools present in bacteria from environmental habitats, modern approaches utilize different methods. Among them, 1) Cesium Chloride Ethidium Bromide -gradient ultracentrifugation<sup>5</sup>, 2) transposon aided capture (TRACA<sup>6</sup>), 3) bioinformatical derivation of plasmid associated contigs/genes<sup>7,8</sup>, and 4) degradation of linear DNA with exonuclease followed by multiple displacement amplification<sup>9–11</sup> are the most employed (see revision of Jorgensen *et al.*<sup>12</sup> and Smalla *et al.*<sup>13</sup>). Although all methods used for plasmid isolation have several limitations they have been successfully used as strategies for plasmid targeting.

We recently reported the characterization of a collection of 35 high molecular weight (HMW) plasmids harboured by bacteria obtained from a BPS contaminated with several pesticides<sup>14</sup>. The sampled BPS was suggested to constitute a suitable source for the isolation of genes encoding novel catabolic enzymes and resistances to toxic elements/compounds based on the high abundance and diversity of MGEs detected in this habitat<sup>15,16</sup>. In particular, the isolate collection characterized by Martini *et al.*<sup>14</sup> harbors more than 50 HMW plasmids amounting to several megabases of sequence information.

In this study, we address the isolation and genetic characterization of large circular plasmids from the bacterial collection originated from a BPS. First, extrachromosomal DNA was isolated and purified, followed by high-throughput sequencing (Illumina sequencing technology, MiSeq system). The sequence data were annotated using the GenDB platform<sup>17</sup> facilitating different BLAST searches against defined databases to get an overview of plasmid-specific modules and plasmid accessory genetic elements. The analysis was focused on the detection of plasmid-specific sequences, antibiotic resistance genes and pesticide degradation genes within the plasmid DNA pool. The results presented herein provide insights into the genetic diversity and structure of these circular replicons and their accessory genetic elements. Moreover, the role of such plasmids for adaptation of the community members to the specific conditions of the BPS habitat (representing a model environment of agricultural interest) is discussed.

## Results and Discussion

**Plasmid DNA sequencing and global analysis of plasmid content.** A collection of plasmid-containing bacteria isolated from an on-farm BPS used for pesticide removal was previously characterized at the taxonomic and phenotypic level<sup>14</sup>. To gain deeper insight into the genetic content of large plasmids harbored by these bacteria, 35 plasmid-containing isolates were chosen for plasmid sequencing. The purified plasmidic DNA was pooled and sequenced using Illumina MiSeq shotgun sequencing technology. The output of the sequencing approach amounted to a total of 2,065,817 reads with an average sequence length of 250 bp. Assembly of the 516 Mb of sequence data using the Newbler assembler (version 2.6) generated 19,962 contigs larger than 500 bp accounting for 20.1 Mb of non-redundant sequence information. The average contig size was 1,513 bp with the largest contig being 76.6 Kb in length. Automatic gene prediction using Prodigal resulted in a total of 24,069 coding sequences (CDS) in contigs larger than 500 bp.

Usually, plasmid-DNA preparations are contaminated with chromosomal DNA. To get rid of contigs representing chromosomal sequences, the assembled dataset was filtered *in silico* for contigs originating from chromosomal DNA. For this purpose, the contigs were compared to host genome sequences available in publicly-accessible databases. Taxonomic assignments of plasmid-containing host bacteria have previously been determined by sequencing of their 16S rRNA gene regions<sup>14</sup>. Contigs that were more than 95% identical to the chromosomes of reference strains for more than 90% of their lengths were assumed to represent chromosomal contamination and therefore were discarded. As a result, 9,386,079 bp of non-redundant sequence information remained which in first instance was supposed to originate from plasmids (hereafter designated the plasmid dataset). Thus, plasmid contigs comprised a total of 11,839 CDSs with an average GC content of 56.24%. Interestingly, a large proportion of sequences (~36%) did not match (*e* value:  $1 \times 10^{-10}$ ) to any known sequences deposited in the public nucleotide and genome databases. A similar observation was reported in other plasmid metagenomes studies<sup>5</sup>. Therefore, plasmids may not only supply their prokaryotic hosts with known auxiliary functions of ecological and adaptative value, but also have to be considered as a resource of so far unknown genetic information.

**Diversity of genes involved in plasmid replication, mobilization and stabilization.** Plasmid-related functions were analyzed by comparison of the amino acid sequences deduced for all predicted CDSs to the domains and reference proteins deposited in the protein family (Pfam) database. The analysis was carried out using the advanced metagenomics analysis platform MGX that allows processing of large datasets such as those generated on Illumina sequencing platforms. Of the 2,065,817 reads generated by Illumina sequencing, 2,055,622 yielded hits against the Pfam database of which 368,055 were related to plasmid functions, indicating that at least 18% of the sequences represented plasmid-related genes. The protein families related to plasmid functions, i.e., replication (*rep*), stabilization/partition (*sta-par*) and mobilization (*mob*)/conjugative transfer genes found in the dataset are listed in Supplementary Tables S1, S2 and S3, respectively.

**Classification of genes encoding proteins involved in replication.** Plasmid incompatibility has been the most traditional way to classify replicons (for a review see Taylor *et al.*<sup>18</sup>). Replication proteins identical to those of the IncP-1, IncP-7, IncP-9 and IncW plasmids (BF\_Rep48, BF\_Rep13, BF\_Rep52, and BF\_Rep24, respectively (Supplementary Table S1)), as well as rolling-circle replication initiators (proteins of the Pfam groups Rep\_1 and Rep\_trans in Supplementary Table S1) are present in our plasmid dataset. But most of the replication initiation proteins identified in this work cannot be assigned to the historical incompatibility groups. Since each Inc group includes plasmids with high nucleotide identity, the Pfam database was used in this work as a more inclusive approach for the replication protein classification.

Thirteen of the fourteen families related to plasmid replication deposited in the Pfam database were identified in the plasmid dataset, illustrating the wide diversity of plasmids present in the BPS isolates. The identified Pfams were assigned to 48 different *rep* genes (Table 1). In addition, six *Rep* sequences which did not contain a Pfam

Pfam number	Pfam ID	Description	Different proteins	Organism	% Identity to the closest protein
pfam00436	SSB	Single-strand binding protein family	10	<i>Legionella pneumophila</i>	50%
				<i>Legionella shakespearei</i>	49%
				<i>Ottowia thiooxydans</i>	96%
				<i>Pseudomonas putida</i>	99%
				<i>Ochrobactrum intermedium</i>	93%
				<i>Microbacterium yannicii</i>	52%
				<i>Pseudomonas syringae</i>	86%
				<i>Acinetobacter baumannii</i>	99%
				<i>Salmonella enterica</i>	100%
pfam01051	Rep_3	Initiator Replication protein	6	<i>Providencia alcalifaciens</i>	64%
				<i>Salmonella enterica</i>	99%
				<i>Pseudomonas putida</i>	100%
				<i>Pseudomonas taiwanensis</i>	71%
				<i>Salmonella enterica</i>	99%
pfam02486	Rep_trans	Replication initiation factor	5	<i>Acidovorax citrulli</i>	56%
				<i>Streptococcus parasanguinis</i>	97%
				<i>Acidovorax delafieldii</i>	48%
				<i>Streptococcus parasanguinis</i>	97%
				<i>Acidovorax citrulli</i>	56%
pfam11740	KfrA_N	Plasmid replication region DNA-binding N-term	5	<i>Pseudomonas aeruginosa</i>	100%
				<i>Pseudomonas putida</i>	46%
				<i>Serratia marcescens</i>	89%
				<i>Salmonella enterica</i>	100%
				<i>Pseudomonas sp.</i>	100%
pfam11800	RP-C_C	Replication protein C C-terminal region	5	<i>Ochrobactrum anthropi</i>	83%
				<i>Ochrobactrum anthropi</i>	89%
				<i>Ochrobactrum rhizosphaerae</i>	99%
				<i>Sinorhizobium meliloti</i>	94%
				<i>Ochrobactrum anthropi</i>	88%
pfam03428	RP-C	Replication protein C N-terminal domain	4	<i>Ochrobactrum rhizosphaerae</i>	88%
				<i>Sinorhizobium meliloti</i>	94%
				<i>Ochrobactrum rhizosphaerae</i>	95%
				<i>Ochrobactrum rhizosphaerae</i>	94%
pfam04796	RepA_C	Plasmid encoded RepA protein	3	uncultured bacterium	62%
				uncultured bacterium	53%
				Enterobacteriaceae	100%
pfam07042	trfA	TrfA protein	2	<i>Thaueria terpenica</i>	68%
				Pseudomonadales	100%
pfam05472	Ter	DNA replication terminus site-binding protein (Ter protein)	2	<i>Escherichia coli</i>	99%
				<i>Pseudomonas sp.</i>	100%
pfam01446	Rep_1	Replication protein	2	<i>Bacillus pumilus</i>	94%
				<i>Bacillus pumilus</i>	98%
pfam03090	Replicase	Replicase family	2	<i>Pseudomonas sp.</i>	90%
				<i>Aeromonas hydrophila</i>	99%
pfam06504	RepC	Replication protein C (RepC)	1	<i>Pseudomonas aeruginosa</i>	96%
pfam02387	IncFII_repA	IncFII RepA protein family	1	<i>Klebsiella pneumoniae</i>	96%

**Table 1.** List of genes and classification of encoded proteins according to replication- related Pfam numbers.

domain strictly associated with plasmid replication were identified in the assembled data using the GenDB annotation platform (included in Supplementary Table S1). These putative Rep proteins contained Pfam domains belonging to the helix-turn-helix DNA binding domain families (HTH\_36 and HTH\_38) as well as to other domains. In addition, they were more than 87% identical to reference replication proteins deposited in the NCBI

database. Most of the identified *rep* genes could be assigned to plasmids that were previously found in different species of the Alpha-, Beta- and Gamma-proteobacteria, *Firmicutes* and *Actinobacteria* (Table 1). These results are in concordance with the taxonomic assignments of the plasmid-containing host bacteria from the BPS as determined by 16S rRNA gene sequencing<sup>14</sup>. Some of the reference plasmids encoding the identified Rep proteins, such as pCARI<sup>19</sup>, pND6-1<sup>20</sup>, pSTY<sup>21</sup>, and pDK1<sup>22</sup> contain degradative genes, and others such as pMET-1<sup>23</sup>, RP4<sup>24</sup>, pAHH01<sup>25</sup>, pIE321<sup>26</sup>, and pTBI1<sup>27</sup> contain antibiotic resistance genes. Identity values between the identified Rep proteins and the best BLASTp hit in the NCBI database ranged from 46% to 100% (Table 1). Out of the 54 replication associated proteins, 40 were between 80 and 100% identical to the closest protein deposited in the NCBI database, while the identity values of the remaining 14 ranged from 46 to 79% (Supplementary Table S1). Identification of replication proteins that are not very closely related to known ones extends our understanding of the diversity of plasmid replication modules and highlights such diversity.

**Classification of genes involved in plasmid stability and partitioning.** Stabilization and partitioning functions concern genes involved in plasmid maintenance and inheritance during the process of cell division. The two main mechanisms responsible for plasmid stabilization/partition, i.e. post-segregational killing and active partitioning systems, were highly represented in the dataset. A final amount of 128 genes associated with these functions were found. In total, 21 of 23 known *par-sta* related Pfam families were identified, including ParBc (PF02195) and CbiA (PF01656) as being the most abundant ones (see Supplementary Table S2).

**Analysis of the mobilization gene set in the BPS plasmid dataset.** The spread of antimicrobial-resistance, virulence and biodegradation traits are outstanding examples of the impact of HGT on the adaptation of bacterial communities under stress, with relevant consequences for the anthropogenic activities<sup>28</sup>. Plasmids have been pinpointed as the main DNA vehicles that transfer genes between bacterial chromosomes, with conjugation as the preponderant transfer mechanism.

Bacterial plasmids that are transmissible by conjugation can be either conjugative (i.e., self-transmissible) or mobilizable. The former plasmids encode all genes required for transfer (namely genes for a relaxase, a coupling protein (T4CP) and a mating-pair formation system (MPF), while the latter generally encode only the relaxase gene but lack VirB4 and MPF and then need a helper plasmid for transfer<sup>29–34</sup>. Analysis of the phylogenetic relationships among the relaxases allows the classification of transmissible plasmids into eight (MOB) families<sup>29,30,32–34</sup>. Self-transmissible plasmids were in turn also grouped into eight MPF families<sup>34,35</sup>.

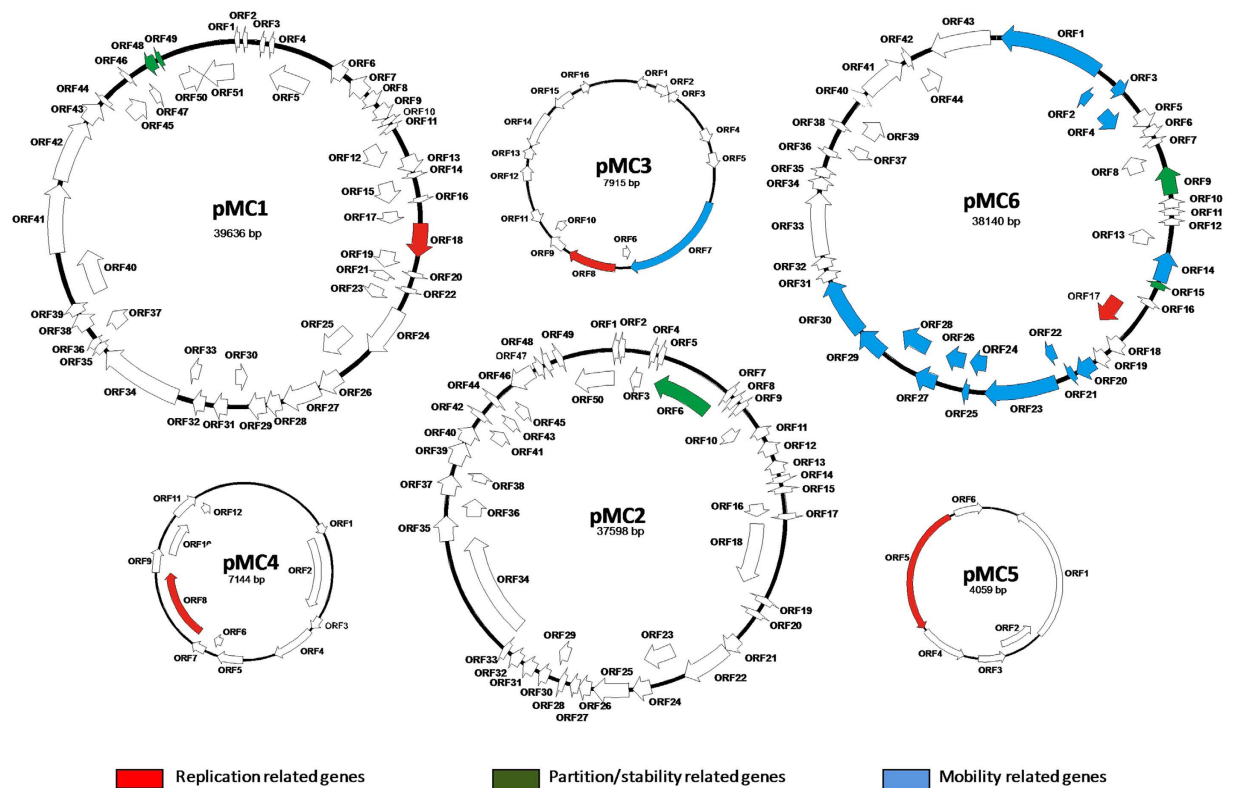
Through the classification of plasmid mobility types the diversity of plasmids and their propagation routes in complex ecosystems can be addressed<sup>36</sup>. In our dataset 12% of the predicted gene products matched a hit involved in bacterial conjugation. Such hits correspond to plasmid sequences whose taxonomic distribution mainly represents the phyla Proteobacteria (Alpha, Beta and Gamma classes), Firmicutes and Actinobacteria. Regarding the completely sequenced plasmids, only two of the six fulfilled the mentioned requirements that allowed the prediction that are transmissible via conjugation. Plasmid pMC3 encodes only a relaxase, and would therefore be only mobilizable, while pMC6 encodes a relaxase next to the T4CP and the MPF system, thus being conjugative (Fig. 1).

Here we analyzed the phylogenetic relationships of the predicted VirB4-like proteins and relaxases of the plasmid dataset to study the diversity of the transmissible plasmids in a BPS environment. VirB4-like proteins are descriptors of the MPF systems and thus serve as an indication of the presence of conjugative plasmids, while relaxases are indicators of either self-transmissible or mobilizable plasmids. Thirteen VirB4 proteins and nineteen relaxases were identified in the plasmid dataset (Supplementary Table S3). On the basis of the enumerations, it appears that the proportion of conjugative versus mobilizable plasmids in our plasmid dataset was higher than that in other metagenome studies of environmental plasmid communities<sup>37</sup>. As can be seen in Fig. 2 all predicted VirB4 proteins identified in this work belong to the MPF<sub>T</sub> type. This MPF type is characterized for including up to eleven proteins (VirB1–VirB11) which are involved in the secretion channel assembly and pilus biogenesis<sup>38,39</sup>. Plasmids encoding MPF<sub>T</sub> systems are broadly distributed in bacterial hosts of all classes of the phylum Proteobacteria<sup>34</sup>.

A total of 19 relaxases were identified in this plasmid dataset: two MOB<sub>P</sub>, four MOB<sub>Q</sub>, one MOB<sub>V</sub> and twelve MOB<sub>B</sub>. The MOB<sub>B</sub> relaxases were placed in the MOB<sub>F11</sub> subfamily (Supplementary Fig. S2). Relaxase BJP\_1037 was identical to those of IncW plasmids (e.g. R388 and R7K), while BJP\_380 was highly similar to relaxases of the IncP-9 plasmids, which were found to be prevalent in polluted environments<sup>40,41</sup> and generally encode xenobiotic-degradation genes (e.g. pWW0 and NAH7). MOB<sub>F11</sub> subfamily only encompasses auto-transmissible (conjugative, non-mobilizable) members<sup>32</sup>, which exhibit a high synteny in their conjugation genes<sup>42</sup>.

The MOB<sub>Q</sub> relaxases that were predicted in the BPS plasmid dataset grouped in two monophyletic clusters (Supplementary Fig. S3). Protein AP\_265 grouped with the conjugative relaxases of *Rhizobium/Agrobacterium* plasmids in subfamily MOB<sub>Q2</sub>. The MOB<sub>Q2</sub> relaxases are involved in the plasmid conjugation process being different from the MOB<sub>P2</sub> VirD2-like proteins that perform the T-DNA transfer from bacteria to plant cells<sup>32</sup>. The predicted proteins BJP\_6246, BJP\_4415 and BJP\_1723 clustered in a clade that has not been classified<sup>32</sup>. It includes plasmid relaxases from diverse hosts belonging to the Alpha- and Gamma-proteobacteria. A gene for a MOB<sub>V</sub> relaxase was found in one of the completely sequenced plasmids, pMC3. This predicted protein, pMC3\_7, clustered with MOB<sub>V4</sub> relaxases (Supplementary Fig. S4). Such relaxases are generally encoded on small mobilizable plasmids hosted in *Bacillus* and *Streptococcus*. MOB<sub>V</sub> mobilizable plasmids are transferred by a wide variety of helper systems<sup>43</sup>.

MOB<sub>P</sub> is the most abundant and diverse relaxase family<sup>32,34</sup>. Accordingly, most predicted relaxases found in our metamobilome belonged to this family (proteins BJP\_1498, AP\_669, BJP\_441, BJP\_847, BJP\_634, BJP\_2740, AP\_394, AP\_113, BJP\_237, BJP\_387, BJP\_198 and pMC6\_1) (Supplementary Fig. S4). The first three constitute



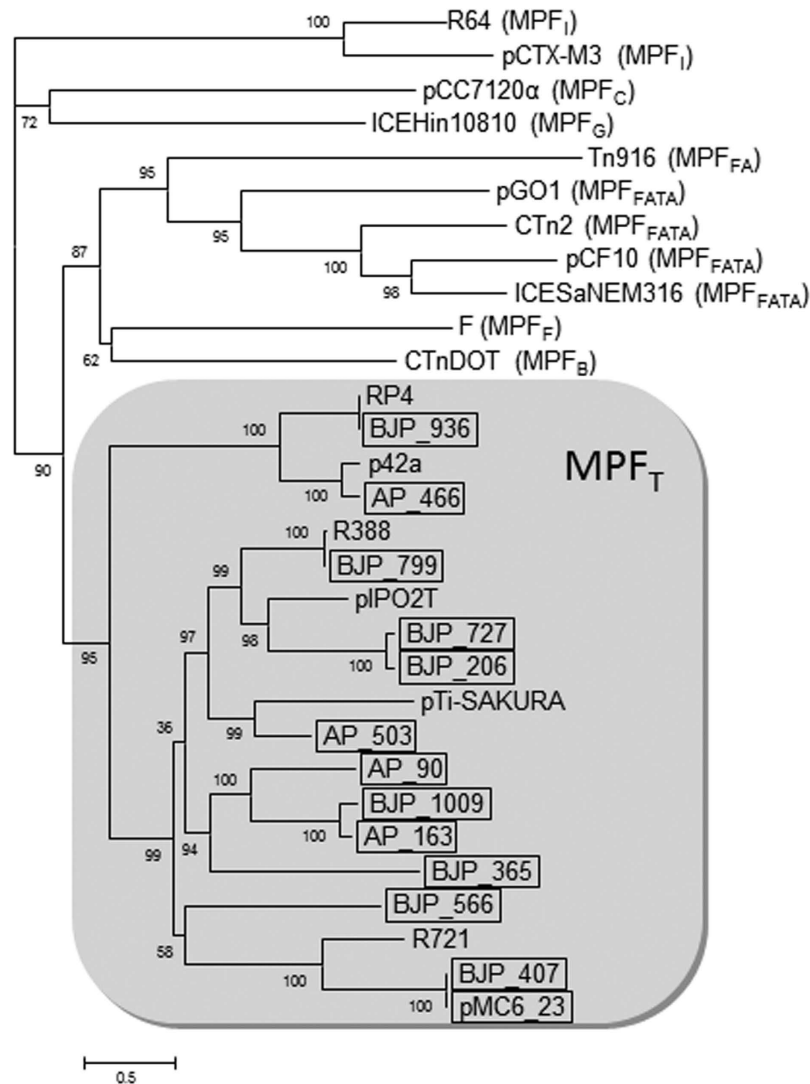
**Figure 1. Physical maps of the six closed replicons obtained from sequence assemblies.** Maps of complete nucleotide sequences of the plasmids pMC1, pMC2, pMC3, pMC4, pMC5 and pMC6 indicating locations of predicted open reading frames (ORFs) are depicted. ORFs are colour coded according to their predicted function as indicated in the associated key. Predicted products and more details regarding putative functions of annotated ORFs are provided in Table S4.

MOB<sub>P1</sub> relaxases. BJP\_1498 and AP\_669 were practically identical to the MOB<sub>P11</sub> relaxase TraI of the IncP-1 prototype RP4. IncP-1 plasmids have previously been found to be abundantly present in pesticide-polluted environments<sup>15,40,44</sup>. Relaxase BJP\_441 belonged to another MOB<sub>P1</sub> subfamily, the MOB<sub>P13</sub> cluster. Other previously classified MOB<sub>P</sub> subfamilies were also represented by genes in the BPS plasmid dataset: MOB<sub>P6</sub> includes proteins BJP\_387, BJP\_198 and pMC6\_1, the relaxase of the completely sequenced plasmid pMC6, while MOB<sub>P7</sub> subfamily contained protein BJP\_237 (Supplementary Fig. S4). Placed in different uncharacterized MOB<sub>P</sub> clades were proteins BJP\_847, AP\_113, BJP\_2740 and AP\_394 (these two last in the same clade). As can be observed in Supplementary Fig. S5a, protein BJP\_634 is a singular relaxase with no close homologs. All MOB<sub>P</sub> relaxases of this plasmid dataset were included in subfamilies populated by self-transmissible instead of mobilizable plasmids.

The high diversity of predicted MOB relaxase families found in this collection revealed a significantly larger diversity in the plasmid pool in the on-farm BPS than had been anticipated from the previous analyses by DNA hybridization<sup>14</sup>. The results of this study, based on the DNA sequencing of a BPS mobilome, also revealed that the plasmid types found in this environment differed from those present in other complex environments such as clinical settings (see for instance<sup>45</sup>). This observation suggests the occurrence of plasmid specialization emphasizing the importance of these mobile elements in the dynamics of local bacterial communities.

**Analysis of six completely assembled plasmid replicons.** Assembly of plasmid replicons was particularly challenging due to the short read length of the Illumina sequence reads and the high abundance of insertion sequence elements, transposons, and other repetitive elements in plasmid genomes. Read assembly generated six complete circular replicons which ranged in size from 4 to approximately 40 kb. The main characteristics of the plasmids and their genomic maps are shown in Table 2 and Fig. 1, respectively. The closest matches of the plasmid-encoded gene products as identified by BLASTp analysis are listed in Supplementary Table S4. Sequence analysis of plasmid replicons showed a large number of CDSs coding for hypothetical proteins when compared to the NCBI database, suggesting that these replicons carry many uncharacterized functions. *In silico* analysis of the replicon-encoded accessory modules uncovered (i) replicon-like prophages (pMC1, pMC2), and (ii) toxin-antitoxin-like components (pMC1, pMC3, pMC6).

Accordingly, four of the six assembled replicons could be classified as plasmids (pMC3, pMC4, pMC5 and pMC6) whereas two of them were considered to represent phages (pMC1 and pMC2). Concerning the latter, the majority of their CDSs were predicted to encode phage proteins, although pMC1 contains a hypothetical protein with low identity to a plasmid replication protein. The three smaller plasmids (pMC3, pMC4 and pMC5) contained one *rep* gene each and most of their CDSs code for hypothetical proteins. Plasmid pMC6 harbors *tra*



**Figure 2. Phylogenetic analysis of the VirB4-like proteins of the plasmid dataset.** The analysis includes 13 proteins belonging to the CagE-TrbE-VirB pfam ID in Table S3 and representatives of the eight mating pair formation systems defined in Guglielmini *et al.*<sup>35</sup>. The MPF<sub>T</sub> cluster is shadowed in grey. VirB4-like proteins of the plasmid dataset are boxed.

and *mob* genes including a relaxase gene and MPF class T genes (according to<sup>35</sup>), suggesting that it is conjugative. This plasmid has a chimeric composition, mostly associated with *Salmonella enterica* sequences, with a high percentage of identity, next to a smaller fragment that is more similar to *Pseudomonas* species sequences (see Supplementary Table S4).

Plasmid pMC1 harboured genes predicted to encode proteins with high similarity to HicA/HicB. In particular, the HicA/B toxin-antitoxin pair belongs to the type II of TA systems. Most predicted gene products of pMC3 showed homology to hypothetical proteins from *Bacillus cereus*. This replicon contains a CDS with high similarity to barstar antitoxin gene. Barstar plays a role as inhibitor of barnase, representing a ribonuclease that cleaves RNA yielding 3'-nucleotides through a 2,3-cyclic intermediate<sup>46</sup>. In plasmid pMC3 is not encoded a barnase homologous protein. However, it is conceivable that presence of a barstar gene may protect a host bacterium from deleterious effects of a corresponding barnase toxin. Replicons pMC1 and pMC2 carry eight and nine phage related genes, respectively. Although pMC1 also contains a CDS that potentially encodes an integrase, no putative replication initiation genes were detected.

Putative replication initiation genes were only identified in pMC3, pMC4, pMC5 and pMC6. The Rep protein from pMC4 and three other proteins encoded in this replicon are related to corresponding *Acidovorax* proteins. In particular the average GC-content of *Acidovorax* sp. approximately is 64% and that of the pMC4 replicon is 57.5%.

The *rep* gene from pMC6 is most closely related to that of *Salmonella enterica* subsp. *enterica* serovar Agona str. SL483 plasmid (GenBank accession number CP001137.1) (99% identity at amino acid level). In addition, further 29 from 44 CDSs are related to *Salmonella* sp. genes. The average GC-content of this genus is approximately

Replicon Name	Replicon type	rep gene	rep gene ID (see Table S5)	rep Pfam	Length	Number of CDSs	% GC content
pMC1	phage*	yes	CDS_18	No detected	39,636 bp	51	51,94
pMC2	phage*	no	–	–	37,598 bp	50	60,9
pMC3	plasmid	yes	CDS_8	No detected	7,915 bp	16	31,78
pMC4	plasmid	yes	CDS_8	pfam02486	7,144 bp	12	57,5
pMC5	plasmid	yes	CDS_6	pfam02486	4,059 bp	6	30,92
pMC6	plasmid	yes	CDS_17	pfam01051	38,140 bp	44	45,69

**Table 2. Main characteristics of the six closed replicons.** \*These replicons were considered to represent phages, since the majority of their CDSs encode phage proteins.

52%, which is higher than GC-content of pMC6 (approx. 45.7%) (see Supplementary Table S4). In this regard, based on genomic data Nishida<sup>47</sup> and Shintani *et al.*<sup>31</sup> revealed that the GC contents of the majority of plasmids is lower than the GC contents of its host chromosome. Plasmid pMC6 also carries genes for three cysteinyl-tRNA synthetase. Although vague, presence of these genes may be beneficial for growth of the host bacterium under cysteine-limiting conditions.

Genes of direct adaptive value such as antibiotic and heavy metal resistance and catabolic genes were not evident in these six closed circular replicons.

**Functional predictions based on assembled contigs from the sequence dataset.** In order to investigate the genomic content of the plasmids, antibiotic and metal resistance genes as well as genes involved in the degradation of pesticides and other toxic compounds were studied. To address the question whether identified accessory genes are plasmid-related, context information represented on assembled contigs was used.

**Presence of antibiotic resistance genes.** Antibiotic resistance genes (ARGs) are ubiquitous in bacterial communities in various ecological niches. In a previous report<sup>14</sup> we showed that the isolates from the BPS collection were able to grow in the presence of different antibiotics and feature a great variability in their resistance levels. It was further demonstrated that different resistance genes are encoded on plasmids. Automatic annotation within the annotation tool GenDB revealed the presence of 112 ARGs predicted to confer resistance to tetracyclines, macrolides,  $\beta$ -lactams, aminoglycosides, bleomycin, fosmidomycin, bacitracin, phenicols and also to acriflavine, (Fig. 3A and Supplementary Table S5), indicating that a large number and different types of resistance genes are present in the plasmid dataset. In order to identify and classify the ARGs present in the sequenced plasmids, a BLAST search was performed against the Antibiotic Resistance Database (ARDB) (<http://ardb.cbcb.umd.edu/>)<sup>48</sup>. In total, 72 CDS were assigned to an antibiotic gene (Fig. 3).

The prevalent ARGs belong to the multidrug resistance/efflux pump systems (41% of all ARGs). Multidrug efflux systems contribute significantly to the increased resistance to multiple antibiotics<sup>49</sup> as well as to other compounds used as antimicrobial agents, such as quaternary ammonium compounds<sup>50</sup> in bacteria. In the assembled sequence data representing the BPS plasmids, three contigs encode multidrug efflux systems. The operons *mar-ABC*, *mdtI* and *mdtON* were identified (shown in Supplementary Table S5).

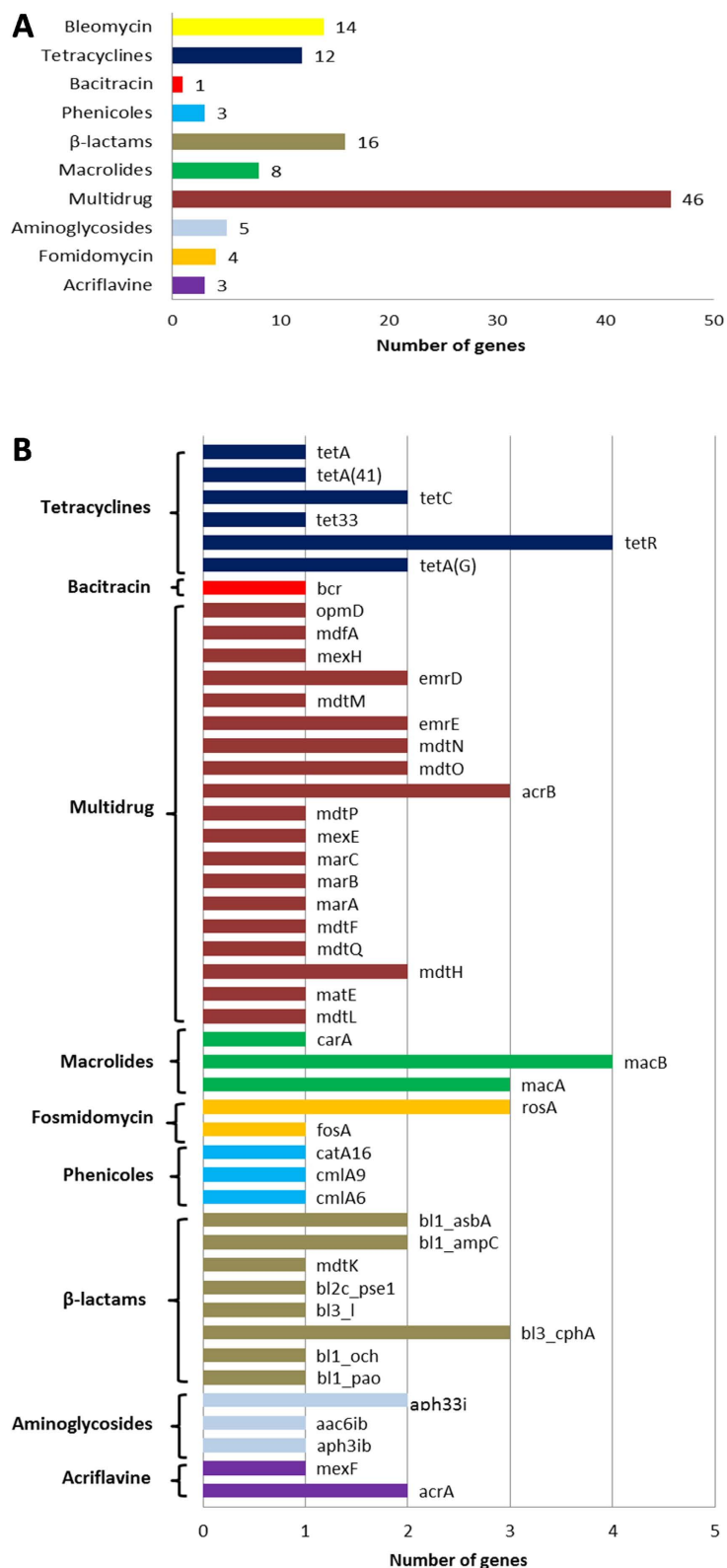
Concerning to  $\beta$ -lactamases, 16 corresponding genes were identified and classified in eight gene variants (see Fig. 3B). Identified  $\beta$ -lactamases show 40 to 100% identity to reference enzymes. These results suggest the presence of putative novel lactamases. Lactamases are of particular interest since  $\beta$ -lactams are the most widely used class of antibiotics, and MGEs, mainly plasmids, are the vectors mediating their dissemination<sup>51</sup>.

Aminoglycosides are among the most commonly used broad-spectrum antibiotics. Resistance is mainly conferred by enzymes that modify aminoglycosides through phosphorylation, adenylation or acetylation<sup>52</sup>. Two putative genes for phosphotransferases were found, one of which corresponds to *aphA* that is present in the plasmid pRK2 (GenBank accession number CP002187). Next, the adjacent CDSs in the contig were analyzed and the presence of *traABC* genes from pRK2 was confirmed.

Three putative chloramphenicol resistance genes were detected. Two of them (*cmlA9* and *cmlA6*) are identical to corresponding genes located on the variable region of class 1 integrons identified in different species. On one contig, the typical structure *cmlA9-tetR-tetA-lysR* was found. Neither the *intI1* gene nor any other elements associated with class 1 integrons were detected on this contig.

Macrolide and glycopeptide resistance genes are homologous to previously described ones in environmental samples. The gene *macB*, which encodes a macrolide-specific efflux system, was detected. These genes can easily be transferred from one host to another since they are usually located on mobile elements such as plasmids<sup>53</sup> and transposons<sup>54</sup>. Finally, tetracycline resistance genes were detected and all of them show a high percentage of identity with the already described ones (*tetA*, *tetC*, *tet33*, and *tetR*) in plant-associated bacteria, human and animal pathogens.

Overall, *in silico* analysis showed that BPS bacteria are a reservoir of novel and unknown genes that have not yet been isolated from other specimens and environments. The advantage that resistance genes confer to microorganisms exposed to anthropogenic antibiotic pressure is undisputed. However, the role of resistance genes in bacteria in environments that are not directly affected by the application of antibiotics is less clear<sup>55</sup>, and not limited to antimicrobial defense. Bacteria can have proteins that provide antibiotic resistance beyond their primary function<sup>56</sup>.



**Figure 3.** (A) Classes of antibiotic resistance genes presents in the plasmid dataset. (B) Antibiotic resistance genes present in the BPS plasmid dataset.

*Presence of heavy metal resistance genes.* Transposons carrying mercury resistance determinants (*mer*) are widely distributed in clinical and environmental bacteria<sup>57</sup>. A contig that carries a truncated narrow spectrum *mer* resistance transposon inserted upstream of a *tnpA* gene was detected in the plasmid dataset. The *mer* gene



cluster encodes a regulatory protein (MerR), a mercuric ion transport protein (MerT), a periplasmic mercuric ion binding protein (MerP), mercuric reductase (MerA), a secondary regulatory protein (MerD), and proteins of unknown functions (MerE and ORF-2). This cluster is similar to the corresponding operons present on Tn4378 of the *Cupriavidus metallidurans* CH34 megaplasmid pMOL28, on Tn501 of the *Pseudomonas aeruginosa*, the *Pseudomonas stutzeri* mercury resistance plasmid pPB, and the Tn501 remnant in pJP4.

In addition, genes related to transport/resistance to arsenic, cobalt, zinc, copper, cadmium, iron, nickel, tellurite, selenium and manganese were detected in the plasmid contigs (see Supplementary Table S6).

**Detection of genes involved in xenobiotic and other organic toxic compounds degradation.** Since our sample was derived from a BPS exposed to pesticides, the microorganisms that inhabit this niche were under a strong selective pressure, may have developed degradation abilities that allow them to survive in this environment. Since plasmids and other MGEs play a key role in bacterial adaptation to environmental changes, could be possible to associate the genes responsible for the degradation of xenobiotic compounds to MGEs. Recently, Dunon *et al.*<sup>15</sup> reported that the increased prevalence of MGEs in a BPS microcosm was accompanied by an increase in the capacity to mineralize the applied pesticides. In general, functions involved in pesticide and other xenobiotic compound resistances are difficult to attribute to particular genes. The biodegradation of specific compounds often is complex requiring a series of biochemical reactions that involve different enzymes whose encoding genes remain unknown. However, preliminary identification of some genes involved in xenobiotic degradation was possible. Genes coding for the main enzymes implicated in the degradation of linuron, 2,4-dichlorophenoxyacetic acid (2,4-D), 3-fluorobenzoate, 4-fluorobenzoate, hexachlorocyclohexane (HCH), dichlorodiphenyltrichloroethane (DDT), atrazine (ATZ), 1,4-dichlorobenzene, dibenzofuran, parathion, endosulfan and methyl viologens were investigated. BLASTp analyses were done using the reference sequences listed in Table S7 against the assembled plasmid database within the GenDB platform. Two putative genes encoding the two main enzymes for 2,4-D degradation (*tfdA* and *tfdB*), the gene for the enzyme responsible for the first step of atrazine degradation (*atzA*), two genes encoding the alpha- and beta-subunits of benzoate 1,2-dioxygenase for 3-fluorobenzoate degradation, and two genes whose products participate in dibenzofuran degradation were identified (see Table S8). The major facilitator system (MFS) transporter for methyl viologen resistance as well as the enzyme required for parathion degradation were also predicted to be encoded in the plasmid dataset.

In addition, a search for aromatic hydrocarbon degrading enzymes was carried out using GenDB annotations. Different putative genes that could participate in xylene, toluene, and/or benzene utilization were detected (Table S8). Two contigs show high similarity to regions of the TOL plasmid pWW0 of *Pseudomonas putida*. This plasmid encodes the metabolic pathways for degradation of toluene, xylene and their alcohol and carboxylate derivatives<sup>58</sup>. A 9 kb contig contained four genes involved in the TOL upper degradation pathway (*xylC-xylM-xylA-xylB*) and another 2 kb contig contained three genes involved in the TOL lower degradation pathway (*xylI-xylH-xylS*).

**Presence of other MGEs in the plasmid metagenome.** Plasmids often carry other MGEs such as IS elements and/or transposons. IS elements are simple MGE consisting of the gene required for transposition of the element and inverted repeats at the ends of the element. Transposons are more complex, in that they also harbor accessory genes not involved in translocation of the element. BLAST analyses of contigs against the transposon and IS database IS finder<sup>59</sup> revealed a total of 120 hits. Among the transposable elements present in the plasmid contigs, those belonging to the Tn3 family were the most abundant ones, followed by IS3 and IS110 family members (Table S9 and Supplementary Fig. S1).

## Conclusions

Plasmid and HGT studies on bacteria colonizing natural or semi-natural environments are important to foster our understanding of the mechanisms that facilitate bacterial adaptation to specific environmental conditions<sup>60,61</sup>. Due to recent developments in high-throughput sequencing technologies, different plasmid pool sequencing approaches have been conducted to investigate and characterize diverse soil microbial communities<sup>12</sup>. In the present study, we isolated and characterized the pool of large plasmids from a bacterial collection previously obtained from an on-farm BPS<sup>14</sup>. This biofilter imposes a strong and long-lasting pesticide selective pressure on the associated microbiome<sup>2</sup>. Moreover, previous reports have shown a high abundance of IncP-like plasmids and other MGEs<sup>15,16</sup> making this environment a highly relevant target site to address bacterial adaptation processes. A limiting factor of large plasmids is the plasmid DNA purification step. Here, we used a successful method comprising cell lysis followed by CsCl density gradient centrifugation. Application of this method enabled the isolation of large amounts of HMW plasmid DNA.

Plasmid sequencing and subsequent *in silico* analysis allowed us to determine replicon types and the genetic potential regarding plasmid mobility. Moreover, information on accessory plasmid functions being of adaptive value for the host bacteria was obtained. In order to explore the host range of plasmids as well as their ability to be mobilized and maintained in different bacteria, plasmid functions were studied in detail. As mentioned above, most of the identified *rep* genes in the dataset (grouped into 13 Pfam families) corresponded to reference genes present on plasmids whose hosts represent different species of the Alpha-, Beta- and Gamma-proteobacteria, *Firmicutes* and *Actinobacteria*. This indicates that a large diversity of plasmid replicons is present in the BPS microbiome. Additionally, the detection of transfer and mobility modules suggested that this bacterial community actively exchanged information *via* plasmid transfer. Experimental evidence for transfer of plasmids of the BPS community was obtained recently<sup>14</sup>. Moreover, other MGEs, i.e. IS-elements and transposons, were identified in the plasmid contigs, suggesting that the flow of information among community members of the BPS can occur by HGT mechanisms.

We analyzed the presence of catabolic genes involved in degradation pathways regarding xenobiotic and other toxic compounds, since such genes are often located on MGEs<sup>62,63</sup>. Different genes whose products could

participate in the degradation of 2,4-D, atrazine, 3-fluorobenzoate, dibenzofuran, parathion, methyl viologen and xylene/toluene/benzene were identified. Although many enzymes that potentially catalyze the degradation of pesticides were detected, full understanding of the underlying biodegradation pathways requires additional investigations. The present study revealed the genetic potential of biodegradation pathways for persistent pesticides/xenobiotics. However, the activity of the genes at the transcript and/or proteome levels remains to be determined. Therefore, further research involving more promising integrated metatranscriptome analyses should be conducted in order to shed light on the ecological and metabolic significance of the identified genes. Discovery of new genes of biotechnological interest is expected from such metatranscriptome studies. Future work will address functional analyses also comprising genes that currently were annotated as hypothetical ones.

Finally, our study shows the technical possibilities to conduct specific HMW plasmid metagenome analyses and highlights the importance of plasmids for local adaptation of complex microbial communities. The plasmid metagenome of the BPS community now provides the basis for continuing efforts to further understand the role of large plasmids under local selective conditions.

## Materials and Methods

**Sampling and plasmid DNA extraction.** Plasmids were obtained from 35 plasmid-containing bacterial isolates previously characterized by Martini *et al.*<sup>14</sup>. The list of all pesticides added to the BPS (Kortrijk, Belgium) and their quantification have been previously reported<sup>16</sup>. Plasmids were obtained from independent 250 ml bacterial-Luria-Bertani (LB) cultures<sup>64</sup> incubated at 28 °C to late exponential phase (approximately 10<sup>8</sup> cells/ml). Plasmids were extracted using a modified alkaline lysis method as previously described by Jouanin *et al.*<sup>65</sup>. Plasmid DNA purification was performed individually to avoid some technical issues, such as differential DNA extraction efficiency from different microorganisms. At least 50 HMW plasmids were purified from other forms of nucleic acids using isopycnic density centrifugation in a classical CsCl-ethidium bromide gradient or QIAGEN Large Construction Kit following the manufacturer's instructions. The concentration of DNA sample was measured with Quant-iT™ PicoGreen<sup>®</sup> dsDNA (Invitrogen). Then the samples were pooled considering the number and size of the plasmids in each isolate in order to add approximately equal amounts of each plasmid. Finally, plasmidic DNA preparations were pooled together for sequencing.

**High-throughput sequencing and computational analysis.** Sequencing was performed at CeBiTec (Bielefeld, Germany) using the Illumina Miseq technology. The reads obtained from the sequencing were assembled using Newbler 2.6. Contigs were imported into the genome annotation system GenDB version 2.0<sup>17</sup> for gene analysis. Individual (short) sequences were analyzed and functionally annotated. An automated function prediction was computed using a combination of standard bioinformatic tools such as Blast and InterPro<sup>66</sup>. After the automatic annotation, the sequence information was refined manually. This approach leads to consistent gene annotation, assigning gene names, EC numbers, COG and GO numbers. Protein identity values were determined using BLASTp.

Plasmid finishing was accomplished using the CONSED<sup>67</sup> software package as described recently<sup>68</sup>. Briefly, reads protruding contig ends were used to identify contigs that flank a certain source contig. These manually approved reads were used for an in silico based gap closure approach. Based on this approach, the completeness of the plasmids was established. Complete replicons were first automatically annotated using the GenDB system<sup>17</sup>. Then circular contigs were manually curated and genes were annotated using the best BLASTp hit at National Center for Biotechnology Information (NCBI) database (<http://www.ncbi.nlm.nih.gov/>). The sequences of these replicons were submitted to GenBank under the following accession numbers: pMC1 (LT158601), pMC2 (LT158602), pMC3 (LT158603), pMC4 (LT158604), pMC5 (LT158605) and pMC6 (LT158606).

For Pfam analysis, unassembled data was uploaded to the MGX platform (Jaenicke, unpublished). Pfam annotations were created using a BlastX search against seed sequences of the Pfam using an e-value cutoff of  $1 \times 10^{-5}$  with disabled sequence complexity filter; the Pfam family of the best blast hit was used to annotate the query sequence.

Plasmid dataset relaxases BJP\_1037, BJP\_380 AP\_265, BJP\_6246, BJP\_4415, BJP\_1723, pMC3\_7, BJP\_1498, AP\_669, BJP\_441, BJP\_847, BJP\_634, BJP\_2740, AP\_394, AP\_113, BJP\_237, BJP\_387, BJP\_198, and pMC6\_1 were used as queries in BLAST searches<sup>69</sup>. For each MOB family, queries and hits were trimmed to keep only the N-terminal relaxase domain (300 amino acids for MOB<sub>B</sub>, MOB<sub>Q</sub> and MOB<sub>V</sub> families and 500 amino acid for MOB<sub>P</sub>) and aligned with MUSCLE<sup>70</sup>. The phylogenetic reconstruction was carried out by maximum likelihood (ML), using RAxML version 7.2.7<sup>71</sup>. In total 20 ML trees were executed using the JTTGAMMA model. Bootstrap trees were then inferred using the autoFC stop criterion to obtain the confidence values for each node of the best ML tree. MOB families and subfamilies were defined as in Garcillan-Barcia *et al.*<sup>32</sup>.

The analysis of VirB4-like proteins of the plasmid dataset included 13 proteins belonging to the CagE-TrbE-VirB pfam ID (recorded in Supplementary Table S3) and representatives of the eight mating pair formation systems defined in [http://conjdb.web.pasteur.fr/conjdb/\\_design/conjdb/index.html](http://conjdb.web.pasteur.fr/conjdb/_design/conjdb/index.html)<sup>35</sup> (GenBank Acc. No. BAA78009.1: TraU\_R64 and AAN87691.1: TraU\_pCTX-M3 for MPF<sub>C</sub>; BAB78290.1: Alr7206\_pCC7120alpha for MPF<sub>C</sub>; WP\_014326569.1: Tfc16\_ICEHin10810 for MPF<sub>C</sub>; AAB60017.1: Orf16\_Tn916 for MPF<sub>FA</sub>; YP\_002790923.1: TrsE\_pGO1, YP\_195783.1: PrgJ\_pCF10, YP\_001086885.1: CD418\_Ctn2 and CAD47019.1: Gbs1360\_ICESaNEM316 for MPF<sub>FATA</sub>; AAC44180.1: TraC\_F for MPF<sub>B</sub>; NP\_809006.1: TraG\_CtnDOT for MPF<sub>B</sub>; AA043554.1: TrbE\_p42a, YP\_009077466.1: TrwK\_R388, CAC82751.1: TraE\_pIPO2T, NP\_065362.1: TraE\_R721; CAJ85691.1: TrbE\_RP4; NP\_053384.1: VirB4\_pTi for MPF<sub>T</sub>). All proteins were aligned with MUSCLE<sup>70</sup> and the maximum-likelihood phylogenetic reconstruction was performed with FastTree 2<sup>72</sup>.

For analysis of antibiotic resistance genes, those genes that were automatically annotated as resistance genes at the GenDB platform were confirmed and annotated according to the best BLASTp hit at NCBI and then blasted against the ARDB database<sup>48</sup> for their classification at gene level. All insertion sequences (IS) were downloaded

from ISfinder<sup>59</sup> in order to generate a database. For such analysis, assembled contigs were used as query for BLAST against such database, using an e-value cutoff of 1e-20. The IS of the best BLAST hit was used to annotate the query sequence. For the analysis of genes involved in pesticide degradation, a BLASTp search was carried out using reference sequences for such genes (listed in Supplementary Table S1) against contigs and the GenDB platform BLASTp tool. Identification of genes involved in metal resistance and degradation of xenobiotics<sup>73</sup> and aromatic hydrocarbons was done with the GenDB platform using automatic annotation and confirmed using the best BLASTp hit at NCBI.

## References

- De Wilde, T. *et al.* Overview of on-farm bioremediation systems to reduce the occurrence of point source contamination. *Pest Manag Sci.* **63**, 111–128 (2007).
- Sniegowski, K. *et al.* Improvement of pesticide mineralization in on-farm biopurification systems by bioaugmentation with pesticide-primed soil. *FEMS Microbiol Ecol.* **76**, 64–73 (2011).
- Frost, L. S., Leplae, R., Summers, A. O. & Toussaint, A. Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol.* **3**, 722–732 (2005).
- Nakamura, Y., Itoh, T., Matsuda, H. & Gojobori, T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet.* **36**, 760–766 (2004).
- Sentchilo, V. *et al.* Community-wide plasmid gene mobilization and selection. *ISME J.* **7**, 1173–1186 (2013).
- Jones, B. V. & Marchesi, J. R. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. *Nat Methods.* **4**, 55–61 (2007).
- Palenik, B., Ren, Q., Tai, V. & Paulsen, I. T. Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol.* **11**, 349–359 (2009).
- Ma, Y., Paulsen, I. T. & Palenik, B. Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. *Environ Microbiol.* **14**, 453–466 (2012).
- Brown Kav, A., Benhar, I. & Mizrahi, I. A method for purifying high quality and high yield plasmid DNA for metagenomic and deep sequencing approaches. *J Microbiol Methods.* **95**, 272–279 (2013).
- Norman, A. *et al.* An improved method for including upper size range plasmids in metatranscriptomes. *PLoS One.* **9**, e104405 (2014).
- Jorgensen, T. S., Xu, Z., Hansen, M. A., Sorensen, S. J. & Hansen, L. H. Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metatranscriptome. *PLoS One.* **9**, e87924 (2014).
- Jorgensen, T. S., Kiil, A. S., Hansen, M. A., Sorensen, S. J. & Hansen, L. H. Current strategies for mobilome research. *Front Microbiol.* **5**, 750 (2015).
- Smalla, K., Jechalke, S. & Top, E. M. Plasmid detection, characterization, and ecology. *Microbiol Spectr.* **3**, 1 (2015).
- Martini, M. C. *et al.* Characterization of a collection of plasmid-containing bacteria isolated from an on-farm biopurification system used for pesticide removal. *Plasmid.* **80**, 16–23 (2015).
- Dunon, V. *et al.* High prevalence of IncP-1 plasmids and IS1071 insertion sequences in on-farm biopurification systems and other pesticide-polluted environments. *FEMS Microbiol Ecol.* **86**, 415–431 (2013).
- Dealtry, S. *et al.* Shifts in abundance and diversity of mobile genetic elements after the introduction of diverse pesticides into an on-farm biopurification system over the course of a year. *Appl Environ Microbiol.* **80**, 4012–4020 (2014).
- Meyer, F. *et al.* GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* **31**, 2187–2195 (2003).
- Taylor, D. E., Gibreel, A., Lawley, T. D. & Tracz, D. M. in *Plasmid Biology*. (eds Funell, B. E. *et al.*) Ch. 23, 473–491. (ASM Press, 2004).
- Maeda, K. *et al.* Complete nucleotide sequence of carbazole/dioxin-degrading plasmid pCAR1 in *Pseudomonas resinovorans* strain CA10 indicates its mosaicity and the presence of large catabolic transposon Tn4676. *J Mol Biol.* **326**, 21–33 (2003).
- Li, W. *et al.* Complete nucleotide sequence and organization of the naphthalene catabolic plasmid pND6-1 from *Pseudomonas* sp. strain ND6. *Gene.* **336**, 231–240 (2004).
- Kohler, K. A. *et al.* Complete genome sequence of *Pseudomonas* sp. strain VLB120 a solvent tolerant, styrene degrading bacterium, isolated from forest soil. *J Biotechnol.* **168**, 729–730 (2013).
- Kunz, D. A. & Chapman, P. J. Isolation and characterization of spontaneously occurring TOL plasmid mutants of *Pseudomonas putida* HS1. *J Bacteriol.* **146**, 952–964 (1981).
- Soler Bistue, A. J. *et al.* *Klebsiella pneumoniae* multiresistance plasmid pMET1: similarity with the *Yersinia pestis* plasmid pCRY and integrative conjugative elements. *PLoS One.* **3**, e1800 (2008).
- Pansegrau, W. *et al.* Complete nucleotide sequence of Birmingham IncP alpha plasmids. Compilation and comparative analysis. *J Mol Biol.* **239**, 623–663 (1994).
- Han, J. E. *et al.* Prevalence of tet gene and complete genome sequencing of tet gene-encoded plasmid (pAHH01) isolated from *Aeromonas* species in South Korea. *J Appl Microbiol.* **112**, 631–638 (2012).
- Götz, A. *et al.* Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR. *Appl Environ Microbiol.* **62**, 2621–2628 (1996).
- Tennstedt, T., Szczepanowski, R., Krahn, I., Pühler, A. & Schluter, A. Sequence of the 68,869 bp IncP-1alpha plasmid pTB11 from a waste-water treatment plant reveals a highly conserved backbone, a Tn402-like integron and other transposable elements. *Plasmid.* **53**, 218–238 (2005).
- de la Cruz, F. & Davies, J. Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol.* **8**, 128–133 (2000).
- Garcillán-Barcia, M. P. & de la Cruz, F. Ordering the bestiary of genetic elements transmissible by conjugation. *Mob Genet Elements.* **3**, e24263 (2013).
- Francia, M. V. *et al.* A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev.* **28**, 79–100 (2004).
- Shintani, M., Sanchez, Z. K. & Kimbara, K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in microbiology* **6**, 242 (2015).
- Garcillán-Barcia, M. P., Francia, M. V. & de la Cruz, F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol Rev.* **33**, 657–687 (2009).
- Guglielmini, J., Quintais, L., Garcillán-Barcia, M. P., de la Cruz, F. & Rocha, E. P. The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet.* **7**, e1002222 (2011).
- Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. & de la Cruz, F. Mobility of plasmids. *Microbiol Mol Biol Rev.* **74**, 434–452 (2010).
- Guglielmini, J. *et al.* Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res.* **42**, 5715–5727 (2014).
- Garcillán-Barcia, M. P., Alvarado, A. & de la Cruz, F. Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiol Rev.* **35**, 936–956 (2011).
- Li, L. L., Norman, A., Hansen, L. H. & Sorensen, S. J. Metatranscriptomics—expanding our knowledge on the pool of plasmid encoded traits in natural environments using high-throughput sequencing. *Clin Microbiol Infect.* **18**(4), 5–7 (2012).

38. Cabezon, E., Ripoll-Rozada, J., Pena, A., de la Cruz, F. & Arechaga, I. Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev.* **39**, 81–95 (2015).
39. Trokter, M., Felisberto-Rodrigues, C., Christie, P. J. & Waksman, G. Recent advances in the structural and molecular biology of type IV secretion systems. *Curr Opin Struct Biol.* **27**, 16–23 (2014).
40. Dealtry, S. *et al.* Cultivation-independent screening revealed hot spots of IncP-1, IncP-7 and IncP-9 plasmid occurrence in different environmental habitats. *PLoS One.* **9**, e89922 (2014).
41. Krasowiak, R. *et al.* PCR primers for detection and characterisation of IncP-9 plasmids. *FEMS Microbiol Ecol.* **42**, 217–225 (2002).
42. Fernández-López, R. *et al.* Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. *FEMS Microbiol Rev.* **30**, 942–966 (2006).
43. Lorenzo-Díaz, F., Fernández-López, C., Garcillán-Barcia, M. P. & Espinosa, M. Bringing them together: plasmid pMV158 rolling circle replication and conjugation under an evolutionary perspective. *Plasmid.* **74**, 15–31 (2014).
44. Jechalke, S., Dealtry, S., Smalla, K. & Heuer, H. Quantification of IncP-1 plasmid prevalence in environmental samples. *Appl Environ Microbiol.* **79**, 1410–1413 (2013).
45. de Toro, M., Garcillán-Barcia, M. P. & de La Cruz, F. Plasmid diversity and adaptation analyzed by massive sequencing of *Escherichia coli* plasmids. *Microbiol Spectr.* **2** (2015).
46. Hartley, R. W. Barnase and barstar: two small proteins to fold and fit together. *Trends Biochem Sci.* **14**, 450–454 (1989).
47. Nishida, H. Comparative analyses of base compositions, DNA sizes, and dinucleotide frequency profiles in archaeal and bacterial chromosomes and plasmids. *International Journal of Evolutionary Biology* 342482 (2012).
48. Liu, B. & Pop, M. ARDB-Antibiotic Resistance Genes Database. *Nucleic Acids Res.* **37**, D443–447 (2009).
49. Dantas, G. & Sommer, M. O. Context matters—the complex interplay between resistome genotypes and resistance phenotypes. *Curr Opin Microbiol.* **15**, 577–582 (2012).
50. Levy, S. B. Active efflux, a common mechanism for biocide and antibiotic resistance. *J Appl Microbiol.* **92**, 65S–71S (2002).
51. Brolund, A. & Sandegren, L. Characterization of ESBL disseminating plasmids. *Infect Dis. (Lond)*, 1–8 (2015).
52. Mingeot-Leclercq, M. P., Glupczynski, Y. & Tulkens, P. M. Aminoglycosides: activity and resistance. *Antimicrob Agents Chemother.* **43**, 727–737 (1999).
53. Liu, Y. F. *et al.* Presence of plasmid pA15 correlates with prevalence of constitutive MLS(B) resistance in group A streptococcal isolates at a university hospital in southern Taiwan. *J Antimicrob Chemother.* **59**, 1167–1170 (2007).
54. Okitsu, N. *et al.* Characterization of *ermB* gene transposition by Tn1545 and Tn917 in macrolide-resistant *Streptococcus pneumoniae* isolates. *J Clin Microbiol.* **43**, 168–173 (2005).
55. Sengupta, S., Chattopadhyay, M. K. & Grossart, H. P. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Front Microbiol.* **4**, 47 (2013).
56. Monk, B. C. & Goffeau, A. Outwitting multidrug resistance to antifungals. *Science.* **321**, 367–369 (2008).
57. Mindlin, S. *et al.* Mercury resistance transposons of gram-negative environmental bacteria and their classification. *Res Microbiol.* **152**, 811–822 (2001).
58. Assinder, S. J. & Williams, P. A. The TOL plasmids: determinants of the catabolism of toluene and the xylenes. *Adv Microb Physiol.* **31**, 1–69 (1990).
59. Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32–36 (2006).
60. Smets, B. F. & Barkay, T. Horizontal gene transfer: perspectives at a crossroads of scientific disciplines. *Nat Rev Microbiol.* **3**, 675–678 (2005).
61. Thomas, C. M. Paradigms of plasmid organization. *Mol Microbiol.* **37**, 485–491 (2000).
62. Top, E. M. & Springael, D. The role of mobile genetic elements in bacterial adaptation to xenobiotic organic compounds. *Curr Opin Biotechnol.* **14**, 262–269 (2003).
63. Nojiri, H., Shintani, M. & Omori, T. Divergence of mobile genetic elements involved in the distribution of xenobiotic-catabolic capacity. *Appl Microbiol Biotechnol.* **64**, 154–174 (2004).
64. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual.* (1989).
65. Jouanin, L., De Lajudie, P., Bazetoux, S. & Huguet, T. DNA sequence homology in *Rhizobium meliloti* plasmids. *Mol Gen Genet.* **182**, 189–195 (1981).
66. Mulder, N. & Apweiler, R. InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol Biol.* **396**, 59–70 (2007).
67. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998).
68. Wibberg, D. *et al.* Complete genome sequencing of *Agrobacterium* sp. H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J Biotechnol.* **155**, 50–62 (2011).
69. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
70. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
71. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* **22**, 2688–2690 (2006).
72. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* **5**, e9490 (2010).
73. Fang, H. *et al.* Metagenomic analysis reveals potential biodegradation pathways of persistent pesticides in freshwater and marine sediments. *Sci Total Environ.* **470**, 983–992 (2014).

## Acknowledgements

This work was supported by the European Commission's 7th Framework Programme (project Metaexplore 222625), the National Scientific and Technical Research Council of Argentina (Consejo Nacional de Investigaciones Científicas y Técnicas—CONICET, Argentina) and Ministry of Science Technology and Productive Innovation (Ministerio de Ciencia Tecnología e Innovación Productiva—MinCyT, Argentina), projects PICT2013-0113, PICT2012-518 and PICT 2012-1719). MCM, FJA were supported by fellowships from CONICET. MFDP, MP, ML, GTT and AL are researchers at CONICET. The bioinformatics support of the BMBF-funded project (grant 031A533) within the German Network for Bioinformatics Infrastructure (de. NBI) is gratefully acknowledged. Work in FdlC group was supported by grant “Plasmid Offensive” BFU2014-55534-C2-1-P from Ministerio de Economía y Competitividad (MINECO, Spain), and Spanish Network for the Research in Infectious Diseases (REIPI RD12/0015/0019) from Instituto de Salud Carlos III (Spain)-co-financed by European Development Regional Fund. The authors are grateful to Paula Giménez and Silvana Tongiani for excellent technical assistance.

### Author Contributions

A.L., M.F.D.P. and M.P. conceived the experiments, M.C.M., M.L., G.T.T. and F.A. conducted the experiment. M.C.M., D.W., S.J., A.S., A.P., M.P.G.-B. and F.d.I.C. analyzed the data. M.P., D.W., S.J., A.S., A.P., M.P.G.-B. and F.d.I.C. assisted with the bioinformatics analysis. M.M., A.L. and M.F.D.P. wrote the main manuscript text. All authors contributed to the discussion and provided comments on the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Martini, M. C. *et al.* Genomics of high molecular weight plasmids isolated from an on-farm biopurification system. *Sci. Rep.* **6**, 28284; doi: 10.1038/srep28284 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>