

# SCIENTIFIC REPORTS



OPEN

## Transposase interaction with the $\beta$ sliding clamp: effects on insertion sequence proliferation and transposition rate

Received: 28 May 2015

Accepted: 23 July 2015

Published: 26 August 2015

Héctor Díaz-Maldonado<sup>1</sup>, Manuel J. Gómez<sup>2,†</sup>, Mercedes Moreno-Paz<sup>1</sup>, Patxi San Martín-Úriz<sup>1</sup>, Ricardo Amils<sup>1,2</sup>, Víctor Parro<sup>1</sup> & Francisco J. López de Saro<sup>1</sup>

Insertion sequences (ISs) are ubiquitous and abundant mobile genetic elements in prokaryotic genomes. ISs often encode only one protein, the transposase, which catalyzes their transposition. Recent studies have shown that transposases of many different IS families interact with the  $\beta$  sliding clamp, a DNA replication factor of the host. However, it was unclear to what extent this interaction limits or favors the ability of ISs to colonize a chromosome from a phylogenetically-distant organism, or if the strength of this interaction affects the transposition rate. Here we describe the proliferation of a member of the IS1634 family in *Acidiphilium* over ~600 generations of cultured growth. We demonstrate that the purified transposase binds to the  $\beta$  sliding clamp of *Acidiphilium*, *Leptospirillum* and *E. coli*. Further, we also demonstrate that the *Acidiphilium* IS1634 transposase binds to the archaeal sliding clamp (PCNA) from *Methanosarcina*, and that the transposase encoded by *Methanosarcina* IS1634 binds to *Acidiphilium*  $\beta$ . Finally, we demonstrate that increasing the strength of the interaction between  $\beta$  and transposase results in a higher transposition rate *in vivo*. Our results suggest that the interaction could determine the potential of ISs to be mobilized in bacterial populations and also their ability to proliferate within chromosomes.

Insertion sequences (ISs) are the simplest mobile genetic elements as they often contain just a single gene encoding the transposase required for transposition. ISs are ubiquitous in bacterial genomes, where they play a key role in genome evolution by promoting chromosomal rearrangements and gene inactivation and mobilization<sup>1,2</sup>. A characteristic feature of ISs is their ability to cross species barriers and transpose actively in new hosts, which makes ISs essential players in the process of horizontal gene transfer (HGT). IS-mediated HGT has been documented extensively among habitat-sharing organisms for Bacteria and Archaea<sup>3,4</sup>. ISs have often been observed flanking recently-incorporated foreign DNA in genomic islands<sup>5</sup>, or in smaller gene clusters in, for example, pathogens<sup>6</sup>, extremophiles<sup>7</sup>, or endosymbionts<sup>8</sup>. However, the invasion of a new host relies on the compatibility of the IS with the new molecular environment: the transposase gene has to be transcribed and translated, and any interactions of the transposase with host functions ought to be maintained<sup>9</sup>. Indeed, since most ISs do not confer a direct selective advantage to the host, IS persistence in populations implies that they must be constantly imported into genomes by HGT<sup>2,3,10</sup>. It could be predicted that an active IS has a short time-window to proliferate in a chromosome before inactivating mutations accumulate<sup>11–14</sup>. ISs have been classified into a number of families depending on their organization, terminal inverted repeats and mechanism of transposition<sup>1</sup>. While some IS families are widely distributed in Bacteria and Archaea, others seem restricted

<sup>1</sup>Centro de Astrobiología (INTA-CSIC), Ctra. de Ajalvir, Km. 4, Torrejón de Ardoz, 28850 Madrid, Spain. <sup>2</sup>Centro de Biología Molecular Severo Ochoa (UAM-CSIC), 28049 Madrid, Spain. <sup>†</sup>Current address: Manuel J. Gómez, Bioinformatics Unit, Centro Nacional de Investigaciones Cardiovasculares (CNIC), 28029 Madrid, Spain. Correspondence and requests for materials should be addressed to F.J.L.S. (email: lopezsfj@cab.inta-csic.es)

to specific groups. Similarly, some IS families can proliferate explosively in a given bacterial lineage and account for a large fraction of the genome, while others are always found in one or few copies in the chromosome<sup>15</sup>. IS abundance has been correlated with lifestyle or with episodes of rapid evolution, such as the colonization of new ecological niches<sup>16–18</sup>.

Although ISs are highly autonomous, there are several examples of interaction with host enzymes and processes, especially chromosomal replication (for a review, see Ref. 1). The association of transposition to replication was first observed for *IS10*, which is activated by the passing of the replication fork<sup>19</sup>. Recently it has been demonstrated that transposases of various IS families interact with the host replication machinery via the  $\beta$  sliding clamp, an essential factor of DNA replication and repair<sup>20–22</sup>.  $\beta$  is a homodimeric protein that encircles DNA and serves as an anchoring platform for diverse enzymes. In addition to conferring processivity to DNA polymerases,  $\beta$  targets enzymes to the chromosome and the replication fork. These enzymes, including the transposases, often contain a short  $\beta$ -binding motif at their C-terminal end, consisting of the consensus sequence Q-L-x-L-F<sup>21,23,24</sup>. The fact that transposases with very different transposition pathways and mechanisms bind to  $\beta$  suggests convergent evolution for a critical contact with the host<sup>21</sup>. The ability of a transposase to bind to  $\beta$  could possibly determine its fate during the invasion of a new host. If binding is weak, the transposase may not have a chance to compete with the many other host factors (or other transposases) that also bind to  $\beta$ , transposition could fail and the IS would become inactivated by mutation. Alternatively, if the transposase binds strongly to  $\beta$ , competition with host enzymes would likely interfere with essential host functions and be detrimental to the cell. Depending on the role of  $\beta$  in transposition, a strong interaction could either favor or block transposase activity.

The goal of this study was to determine whether the interaction of transposases with  $\beta$  is a possible requirement that limits IS movement across species, and whether the strength of the interaction could influence the frequency of transposition events. We analyzed the dynamics of IS populations during 600 generations in a recently-isolated acidophilic bacterium, *Acidiphilium* sp. PM, and identified those ISs that had changed in copy number during that period of time. We detected proliferation of a member of the *IS1634* family, a commonly observed IS element in the genome of acidophiles. We demonstrated that the *IS1634* transposase can physically interact with the *Acidiphilium*  $\beta$  sliding clamp and with  $\beta$  from other acidophiles. We then tested the ability of the *Acidiphilium* *IS1634* family member to proliferate in *E. coli*, an organism in which this IS family has never been detected, and obtained direct evidence suggesting that the affinity of the transposase for  $\beta$  could directly determine transposition rate. Finally, we also found that peptides containing the  $\beta$  binding motif of the transposase encoded by the *Acidiphilium* *IS1634* can bind to an archaeal processivity factor, PCNA, from *Methanosarcina*, and that peptides derived from a *Methanosarcina* *IS1634* transposase can bind to the *Acidiphilium*  $\beta$  sliding clamp.

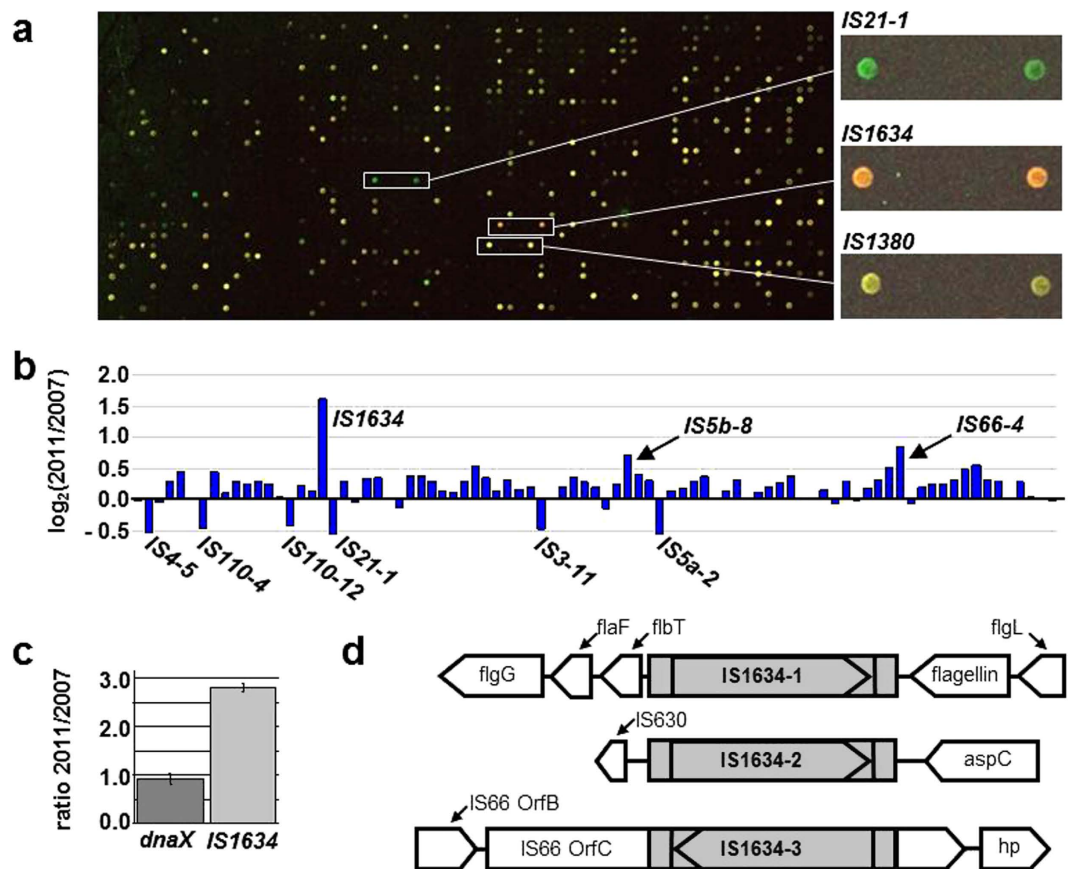
## Results

### A member of the *IS1634* family is active in *Acidiphilium* sp. PM, and its transposase contains a $\beta$ -binding motif.

The sequencing of the genome of *Acidiphilium* sp. PM, an acidophilic  $\alpha$ -proteobacterium isolated from the Tinto river in the south of Spain, revealed that up to 2.4% of its chromosome is composed of repeated regions<sup>25</sup>. 85 ORFs related to insertion sequences (transposases and associated factors) were detected and classified by comparative analysis against the Pfam and ISfinder databases (See Ref. 21 for details). Although some of these ORFs corresponded to full-size transposases or associated proteins, many of them were likely inactivated remnants and gene fragments that are no longer capable of transposition. In order to identify active transposable elements, we performed a long-term cultivation of *Acidiphilium* sp. PM, and analyzed changes in IS abundance that had become fixed in the population at the end of the experiment. The sequenced strain was used to start a culture which was then grown for ~600 generations over four years (See Methods). To identify changes in the number of ISs in the population, we used an oligonucleotide microarray consisting of probes representing 1,358 genes associated to ISs, phage and plasmid mobility functions present in the genomes of selected acidophilic organisms, including *Acidiphilium* sp. PM (see Methods). Although this methodology would not detect transposition events that did not change IS copy number in the chromosome (i.e., conservative transposition) or chromosomal rearrangements, we expected that it would detect IS proliferation or loss.

DNA obtained from the founder strain ('2007') and from the 4-year-old population ('2011') were differentially labelled with fluorescent dyes Cy3 and Cy5 and hybridized together against the microarray (Fig. 1a). We observed changes ( $\log_2[2011/2007] > 0.5$ ) in nine elements (Fig. 1b and Table S1). Changes implied the apparent loss of the IS ORF in six cases (*IS4-5*, *IS110-4*, *IS110-12*, *IS21-1*, *IS3-11*, *IS5a-2*), and a small gain in copy number in three cases (*IS1634-1*, *IS5b-8* and *IS66-4*).

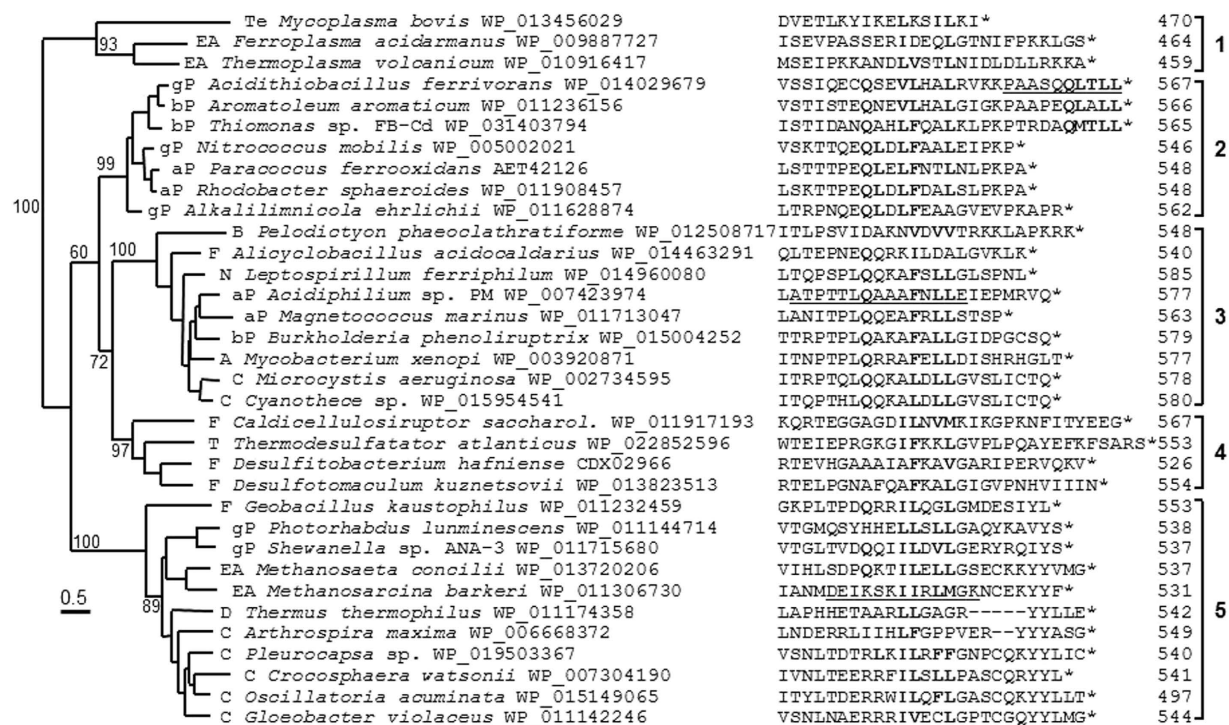
A member of the *IS1634* family, *IS1634-1*, present in single copy in the '2007' culture, showed an increase to three copies in the '2011' culture, and we focused on this element for the rest of the study. *IS1634-1* in the sequenced '2007' culture (Fig. S1, Accession number: WP\_007423974) could be a recent insertion in the chromosome, as suggested by its anomalous %GC and GC skew (Fig. S2). The copy number increase was confirmed by qPCR, inverse PCR, and sequencing (Fig. 1c,d). Our analysis of the insertion sites of the new copies (*IS1634-2* and *IS1634-3*) revealed the characteristic inverted repeats generated by true transposition events, but did not detect obvious adaptive benefits for the evolving culture of *Acidiphilium*.



**Figure 1. Identification and quantification of IS changes in a 600-generation culture of *Acidiphilium* sp. PM.** (a) Oligonucleotide microarray to identify and quantify changes in ISs copy number in chromosomes. Three examples of are shown: for a IS that was lost (IS21-1), increased in copy number (IS1634-1), and remained without change (IS1380-1). (b) Plot of the relative change ( $\log_2[2011/2007]$ ) of transposases (or fragments) observed with the oligonucleotide microarray. Each bar represents the average change for each transposase, calculated from the fluorescence of two different oligos replicated three times in each microarray. The microarrays were assayed three independent times. Relevant ISs are indicated (see Table S1 for the full list). (c) qPCR experiment to quantify the relative change of IS1634 from the '2007' to the '2011' culture. DnaX, encoding the  $\gamma$  subunit of DNA polymerase III, and present in single copy in the chromosome, was chosen as a reference. (d) Inverse PCR sequencing showed the genomic context of IS1634-1 in the '2007' culture and the two additional copies present in the '2011' culture. IS1634-1 and 2 were located in the contig NZ\_AFP01000455 (10547 bp), while IS1634-3 was located in contig NZ\_AFP01000454 (3091 bp). 'hp', hypothetical protein. The inverted repeats were 'ACTAGT' for IS1634-1, 'TCTAAA' for IS1634-2, and 'GATAGA' for IS1634-3.

The IS1634 family is a poorly characterized family of insertion sequences related to IS4, ISH3 and IS701<sup>26</sup>. A similarity tree for the transposases encoded by IS1634 members (Fig. 2) shows that this IS family can be found in Euryarchaea and in numerous groups of Bacteria. Interestingly, IS1634 appears in organisms that share acidic, heavy-metal rich environments, and is present, among others, in the chromosomes of most species of *Leptospirillum* (Nitrospira), *Acidithiobacillus* ( $\gamma$ -proteobacteria) and *Acidiphilium* ( $\alpha$ -proteobacteria), as well as in *Paracoccus ferrooxidans* ( $\alpha$ -proteobacteria), *Ferroplasma acidarmanus* (Euryarchaea), *Alicyclobacillus acidocaldarius* (Firmicutes), *Thiomonas* sp. ( $\beta$ -proteobacteria), and *Desulfobacca acetoxidans* ( $\delta$ -proteobacteria) (Fig. 2). IS-sharing has been observed before for co-habiting organisms<sup>3</sup>.

The interaction of transposases with the host  $\beta$  sliding clamp occurs almost always at the C-terminus of the protein via a short sequence with a weak consensus sequence ( $Q_1-L_2-x-L_4-F_5$ )<sup>21</sup>. A search of the  $\beta$ -binding motif among IS1634 transposases shows that a putative motif can be found at the C-terminus of these proteins (Fig. 2). Despite high variation in sequence context, the alignment of this region shows conservation of the Gln at position 1 of the putative  $\beta$ -binding motif plus conservation of hydrophobic or aromatic residues at positions 4 and 5, suggesting a possible interaction of IS1634 transposases with the host  $\beta$  sliding clamp via this region.

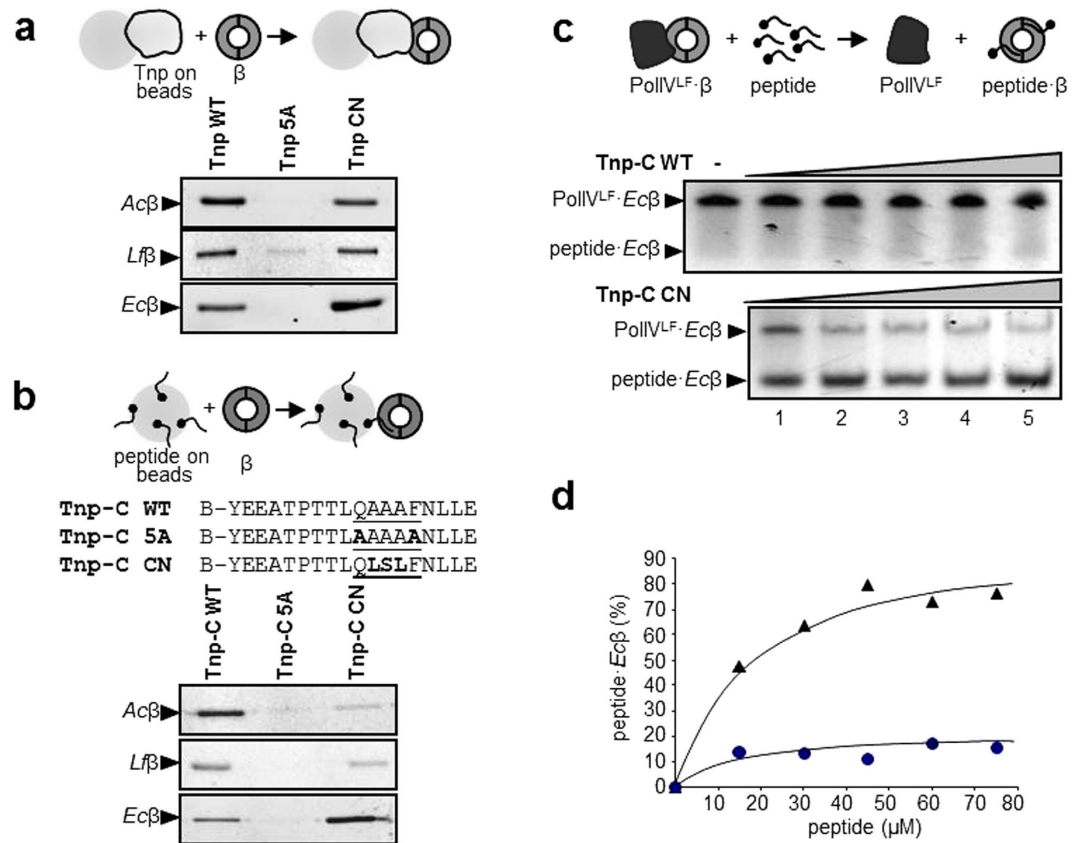


**Figure 2. Unrooted similarity tree of IS1634 transposases and alignment of C-terminal region.** IS1634 transposase sequences were obtained from a previous study<sup>21</sup>, in which 2,216 genomes downloaded from the National Centre for Biotechnology Information (NCBI) Genome database on October 2012, were systematically scanned to identify and classify ISS. The collection of IS1634 transposase sequences was made non-redundant at a sequence similarity level of 90% and by choosing one representative sequence per genus. The tree was computed with PhyML using a JTT model and a bootstrap of 500 replicates, and the results visualized with Seaview<sup>61</sup>. Bootstrap values are shown for the main branches. The analysis reveals five distinct groups of IS1634 transposases. The code preceding each organism name is as follows: aP, bP, gP and dP stand for  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  Proteobacteria, respectively; F, Firmicutes; Te, Tenericutes; N, Nitrospira; C, Cyanobacteria; T, Thermodesulfobacteria; D, *Deinococcus-Thermus*; A, Actinobacteria; EA, Euryarchaeota. For each transposase the sequence of the C-terminus of the protein is presented and residues involved in the putative  $\beta$ -binding motif are aligned and in bold. Underlined sequences are those of peptides used for biochemical analysis in this work.

**The IS1634 transposase from *Acidiphilium* can bind  $\beta$  from diverse organisms.** To determine whether *Acidiphilium* IS1634 transposase (Tnp) could bind *Acidiphilium*  $\beta$ , we purified both proteins from overexpressing *E. coli* strains (See Methods) and probed their interaction biochemically. We purified the full-size, wild-type transposase and, for comparative purposes, two engineered mutant proteins, one in which the Q<sub>1</sub> and the F<sub>5</sub> of the  $\beta$ -binding motif had been mutated to Alanine (named 5A), and one containing a full consensus sequence (QLSLF, named CN). Also, in addition to *Acidiphilium*  $\beta$  protein (*Ac* $\beta$ ), we purified  $\beta$  from *Leptospirillum ferrooxidans* (*Lf* $\beta$ ) and *Escherichia coli* (*Ec* $\beta$ ) (See Methods). We chose *Leptospirillum* (phylum Nitrospira) because, although distant from *Acidiphilium* phylogenetically, it shares the same acidic habitats, where it is usually the most common organism. *E. coli* was chosen because its  $\beta$  sliding clamp and its ligands have been well described, facilitating the realization of comparative and genetic experiments as described below, and because no IS1634 member has been yet detected in Enterobacteria. We assayed the interaction by covalently coupling the transposase to tosyl-activated magnetic beads, probing them with fluorescently-labelled  $\beta$ , and extensive washing. As shown in Fig. 3a, wild type Tnp can bind and retain  $\beta$  from *Acidiphilium*, *Leptospirillum* and *Escherichia* but the double mutant (Tnp 5A) could not. The mutant with the consensus sequence motif (CN) could bind to  $\beta$  from the three species.

In order to further demonstrate that the C-terminal regions of Tnp contained a  $\beta$ -binding motif, we synthesized peptides containing these sequences. The peptides, which carried biotin at the N-terminus, were coupled to streptavidin-coated magnetic beads and probed against fluorescently-labelled  $\beta$  from the three species. The results (Fig. 3b), resemble those obtained with full size proteins, and confirm that the C-terminus of the transposase is responsible for the interaction with  $\beta$ .

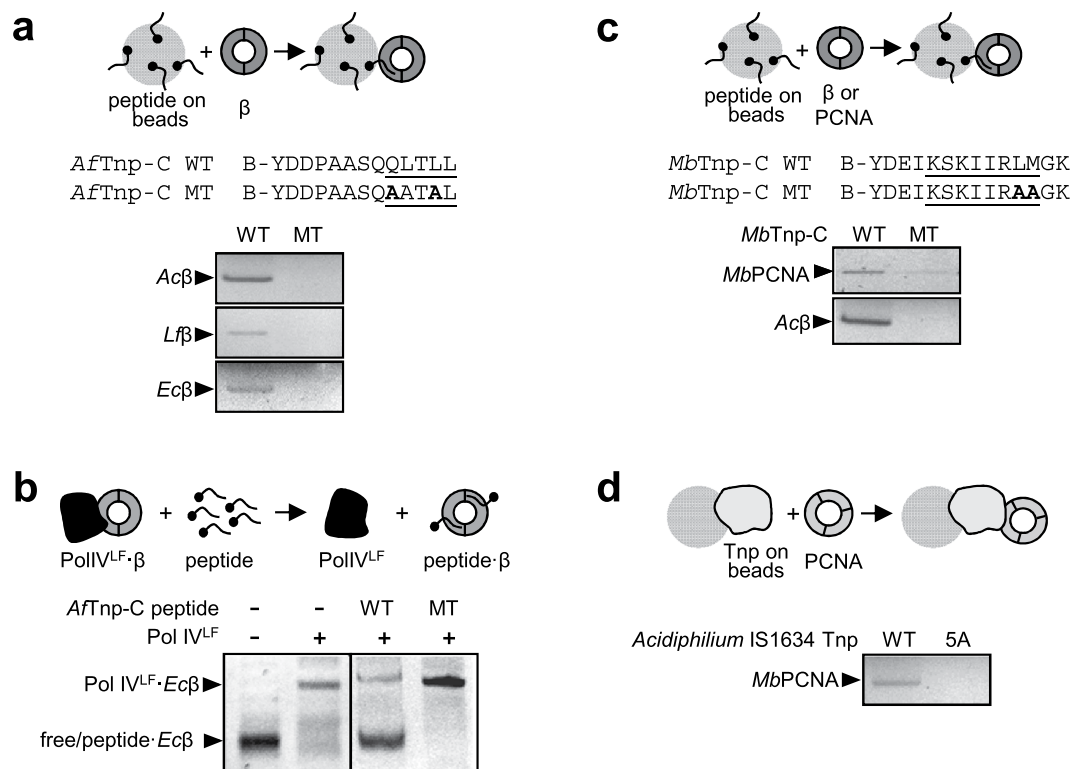




**Figure 3. Interaction between IS1634 transposases and sliding clamps.** (a) Binding assay of wild-type (WT) and mutants 5A and CN of *Acidiphilium* IS1634 Tnp to labelled  $\beta$  from, top to bottom, *Acidiphilium*, *Leptospirillum* and *Escherichia coli*. Tnp was covalently coupled to tosyl-activated magnetic beads and any  $\beta$  retained analyzed by SDS-PAGE. (b) Sequence and binding assay of peptides derived from C-terminal sequences of wild-type and mutants 5A and CN of *Acidiphilium* IS1634 Tnp. The amino acids putatively comprising the  $\beta$ -binding motif are underlined, and those changed from the WT sequence are in bold type; B, biotin; the sequence of the C-terminus of *Acidiphilium* IS1634 Tnp begins at residue 4 of the synthetic peptide, see Fig. 2. The N-terminally biotinylated peptides were bound to streptavidin beads and assayed with labelled  $\beta$  from *Acidiphilium*, *Leptospirillum* and *Escherichia coli*.  $\beta$  retained was analyzed by SDS-PAGE. (c) Competition assay of complexes of *E. coli*  $\beta$  bound to the C-terminal domain ('little finger', LF) of DNA polymerase IV (PolIV<sup>LF</sup>). The WT (top panel) and CN (bottom panel) peptides derived from *Acidiphilium* IS1634 Tnp were used to disrupt PolIV<sup>LF</sup>-Ec $\beta$  complexes. Concentration of peptides is as follows for lanes 1–5: 15, 30, 45, 60, 75  $\mu$ M. See Ref. 21 for a detailed description of the methodology. (d) Quantification of PolIV<sup>LF</sup>-Ec $\beta$  and peptide-Ec $\beta$  complexes of (c) was done by densitometry and plotted as percentage of peptide-Ec $\beta$  bound. Tnp-C WT data is represented with circles and Tnp-C CN is represented as triangles. See Figure S5 for full size gels corresponding to the cropped gels shown in this figure.

Since experiments in Fig. 3a,b are mostly qualitative, we used a competition assay to analyze the relative strength of the interaction between WT and CN peptides and Ec $\beta$ . This assay makes use of the gel-shift generated by DNA polymerase IV (LF domain, PolIV<sup>LF</sup>) bound to  $\beta$  in native gel electrophoresis<sup>21</sup>. These two proteins bind strongly and their interaction has been described in atomic detail<sup>27</sup>. PolIV<sup>LF</sup> binds to Ec $\beta$  at the canonical hydrophobic pocket on the C-side of the ring which is the binding site for all other proteins studied to date. Therefore a displacement of this complex by an excess of peptide, observable in the native gel by a change in the mobility of  $\beta$ , would indicate that the peptide binds strongly to the same pocket on  $\beta$ . A titration of the WT and CN peptides on the preformed complex (See Methods for details) indicate that the CN peptide binds to Ec $\beta$  with higher affinity than the WT peptide, as expected (Fig. 3c,d).

The alignment of Fig. 2 showed that in some species (*Aromatoleum*, *Thiomonas*, *Desulfobacca* and *Acidithiobacillus*) a potential second motif can be found at the C-terminal end of some IS1634 family transposases. To determine whether these sequences could bind to  $\beta$ , we synthesized a peptide containing the C-terminal sequence of *Acidithiobacillus ferrivorans* IS1634 transposase. As shown in Fig. 4a, this peptide (AfTnp-C WT) can bind and retain *Acidiphilium*, *Leptospirillum* and *E. coli*  $\beta$ , while a double mutant in which the Q and L in first and fourth positions of the motif were mutated cannot



**Figure 4. Binding of *Acidithiobacillus* and *Methanosarcina* IS1634 transposase to sliding clamps.** (a) Interaction of peptides derived from the *Acidithiobacillus ferrivorans* IS1634 transposase (WP\_014029679) with  $\beta$  from *Acidiphilium*, *Leptospirillum* and *E. coli*. Peptide sequences are shown for the wild-type sequence (WT) and mutant (MT), and the putative  $\beta$ -binding motif underlined. The sequence of the C-terminus of *Acidithiobacillus* transposase starts at residue 4 of the synthetic peptide (see Fig. 2). The biotinylated (B) peptides were bound to streptavidin-coated magnetic beads and assayed with the different  $\beta$ 's. (b) Native PAGE was used to separate PolIV<sup>LF</sup>· $\beta$  complexes from free or peptide-bound  $\beta$ , as in Fig. 3C, demonstrating that the AfTnp-C peptide binds to the same hydrophobic pocket on  $\beta$  as PolIV<sup>LF</sup>. (c) Interaction of peptides derived from the *Methanosarcina barkeri* IS1634 transposase (WP\_011306730) with *Methanosarcina barkeri* PCNA and *Acidiphilium*  $\beta$ . Biotinylated peptides were bound to streptavidin-coated magnetic beads and assayed with fluorescently-labelled MbPCNA or Ac $\beta$ . (d) Interaction between *Methanosarcina barkeri* PCNA and *Acidiphilium* IS1634 Tnp (WT and 5A mutant). Wild-type and 5A mutant were coupled to tosyl-activated magnetic beads and probed with labelled *Methanosarcina* PCNA. See Figure S5 for full size gels corresponding to the cropped gels shown in this figure.

(AfTnp-C MT). Further, AfTnp-C WT can displace PolIV<sup>LF</sup> from a complex with Ec $\beta$ , revealing that the *Acidithiobacillus* peptide binds to the same region of Ec $\beta$  as other peptides (Fig. 4b). Finally, we have also detected a sliding clamp-binding motif in the transposases of two IS1634 members from Euryarchaea (*Methanosarcina* and *Methanosaeta*) compatible with the consensus PCNA-binding motif (Q-x-x-I-x-x-F-F)<sup>28</sup>. PCNA (Proliferating Cell Nuclear Antigen) is the Archaeal and Eukaryotic sliding clamp, functionally homologous to bacterial  $\beta$ <sup>29</sup>. We purified *Methanosarcina barkeri* PCNA and demonstrated binding to a peptide derived from the IS1634 transposase present in this organism and containing the putative PCNA binding motif (Fig. 4c). Given the relatedness and overlap between the  $\beta$  and PCNA binding motifs, we also tested whether interactions could be maintained across phylogenetic domains. Certainly, we could observe the interaction between the peptide derived from the *Methanosarcina* transposase and *Acidiphilium*  $\beta$  (Fig. 4c). Then, we tested whether peptides derived from *Acidiphilium* Tnp could interact with *Methanosarcina* PCNA. As shown in Fig. 4d, these proteins interact, showing that the interaction could potentially be maintained, bidirectionally, across the Bacteria-Archaea boundary.

**A stronger  $\beta$ -binding motif increases transposition *in vivo*.** Although the  $\beta$ -binding motif is universal, the established consensus sequence (QLxLF) was experimentally validated using *Escherichia coli*  $\beta$ <sup>23,30,31</sup>. A comparison of several  $\beta$ -binding motifs found in *E. coli* and *Acidiphilium* DNA replication

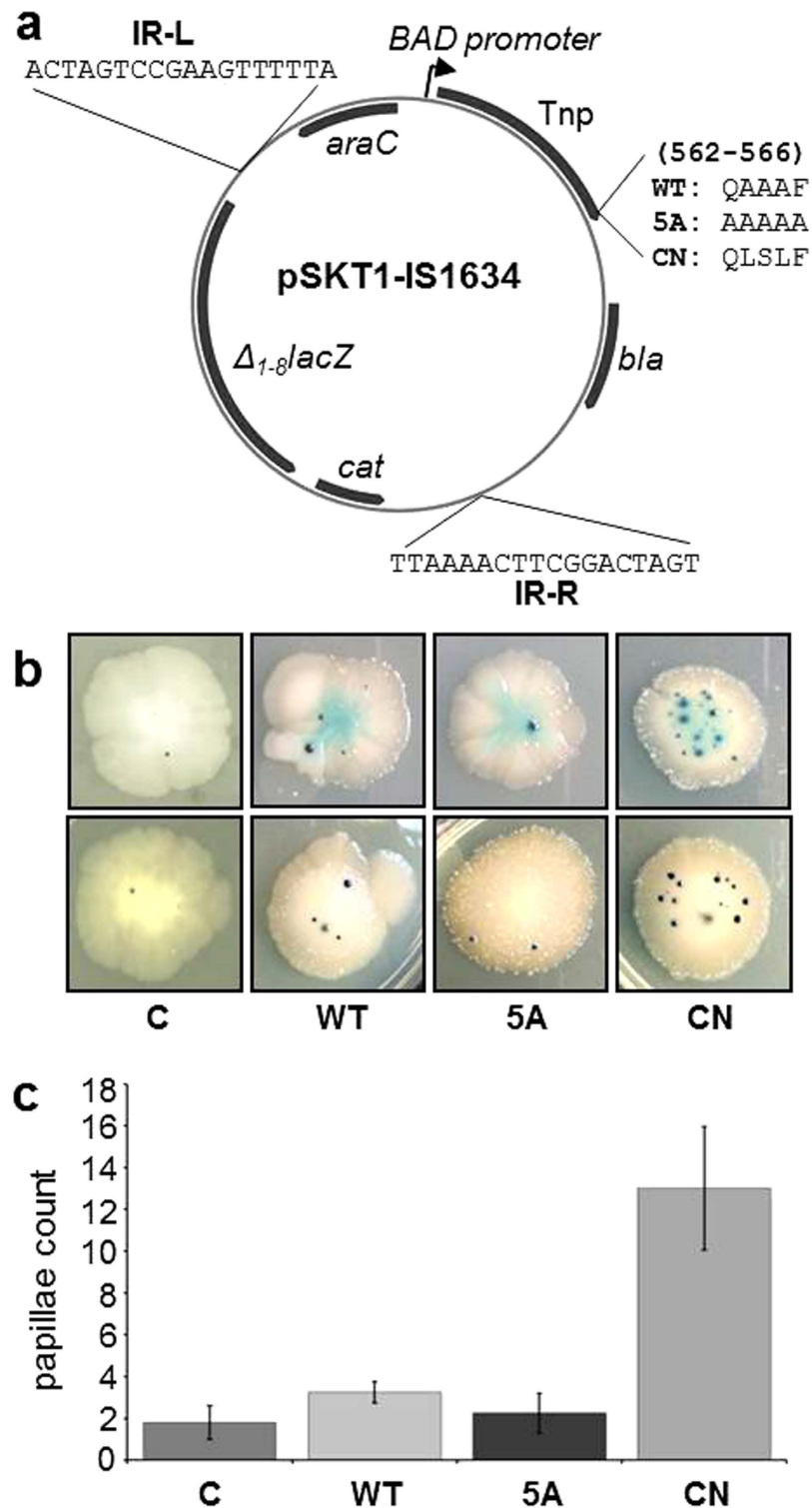
and DNA repair enzymes shows that those of *E. coli* better conform to the consensus (Fig. S4). On the other hand, the identity between *Acidiphilium* and *E. coli*  $\beta$  is only 35.4%. Although IS1634 Tnp binds *E. coli*  $\beta$  (Fig. 3 and crosslinking, Fig. S3), especially the variant containing the  $\beta$ -binding consensus (CN), it was uncertain whether this higher affinity (Fig. 3c,d) would result in more efficient transposition. In order to determine whether the interaction detected *in vitro* with  $\beta$  has an effect in the ability of transposases to be functional in other organisms, we performed an *in vivo* transposition assay for the three variants of the *Acidiphilium* IS1634 transposase in *E. coli*. We used a recently-developed vector that generates genomic insertions of the *lacZ* gene flanked by the IS inverted repeats<sup>32</sup>. The *Acidiphilium* IS1634 transposase gene was placed under the transcriptional control of the BAD promoter, allowing for the modulation of its expression by addition of arabinose (Fig. 5a). We transformed *E. coli* DH5 $\alpha$  cells with the three plasmid variants containing wild-type Tnp and the mutants 5A and CN, plus a control plasmid with no transposase cloned, and incubated for 15 days at 30 °C. We observed papillae reflecting transposition events in WT colonies, although clearly occurring at a low frequency in our experimental conditions. Mutation 5A showed lower generation of papillae than WT but further experimentation will be required to investigate whether this mutation decreases or eliminates transposition. However, cells harbouring the plasmid with the CN mutation clearly had a significant increase of transposition with respect to WT, generating ~4–6-fold more papillae (Fig. 5b,c). Our results therefore suggest that stronger binding of the transposase to  $\beta$  increases the frequency of transposition events in the cell.

## Discussion

The objective of this study was to identify an active insertion sequence from a natural environment and explore its interaction with the host  $\beta$  sliding clamp regarding its ability to proliferate within the chromosome and its potential for dispersal to other species. We searched for changes in IS abundance in the genome of *Acidiphilium* sp. PM that had become fixed in the population during 4 years of laboratory cultivation, and identified one IS that had increased its copy number during this period. We demonstrated that the transposase encoded by this IS interacts with  $\beta$  and mapped the interaction motif to the C-terminus of the protein. Next, we showed that the transposase could bind to  $\beta$  from various species, some of them distant phylogenetically, and identified additional examples of how the interaction between transposases and processivity factors ( $\beta$  and PCNA) can occur even across large phylogenetic distances. Finally, we demonstrated that optimizing the binding site on the transposase to fit the consensus  $\beta$ -binding motif increases transposition in a new host.

Our 600-generation laboratory evolution experiment with *Acidiphilium* sp. PM, aimed at detecting changes in IS copy number that became fixed in the culture by comparative microarray hybridization, found three cases of proliferation and six deletions. Other IS transposition events could have occurred and remained undetected because they were detrimental to the host and were selected against, because they were lost by genetic drift, or because they involved IS relocation in the chromosome (in the case of cut and paste transposition). As it is the case with other studies that have analyzed spontaneous global transposition activity in genomes, our study suggests that only a small fraction of the ISs detected by sequencing are active<sup>33</sup>. The detectable transposition activity can vary greatly with IS element and strain (from  $10^{-3}$  to  $10^{-7}$  per generation, according to some estimates)<sup>34–36</sup>. Unlike the well-defined and relatively predictable point mutation rates resulting from the combined action of DNA polymerases and repair genes, transposition rates are likely the result of the combination of diverse factors such as IS sequence variation, transposase expression and activity, chromosomal location effects, or various host regulatory mechanisms. Additionally, it is likely that the process of isolation and adaptation to laboratory conditions change transposition rates. For example, the observed transposition events in *Ferroplasma* were substantially more frequent in culture than in environmental samples of the same organism<sup>37</sup>. Although transposase gene expression in laboratory conditions has not been explored, metatranscriptomic and metaproteomic analysis of bacterial populations have also detected relatively quick changes in transposase gene expression in response to environmental stimuli<sup>38,39</sup>, although the underlying mechanisms are unknown.

The generally low rates of transposition observed in nature suggest that they are determined by regulatory mechanisms that are intrinsic to IS elements, perhaps as a strategy to limit any negative effects on the host genome<sup>9</sup>. However, the interaction of ISs with the essential replication and repair factor  $\beta$ , first demonstrated for transposon Tn<sup>722</sup>, has now been generalized to nine IS families with very different transposition pathways<sup>21</sup>. This extreme case of functional evolutionary convergence for a highly conserved and universal element suggests that  $\beta$  binding could be an integral aspect of transposition regulation. The  $\beta$  binding motif is present in a large number of proteins involved in DNA synthesis and repair. Since  $\beta$  is limiting in the cell<sup>24</sup>, the concentration of transposases and the strength of their interaction could determine the efficiency of transposition. All proteins so far studied in detail interact with  $\beta$  on the same face of the ring, binding competitively to the same conserved 'hydrophobic pocket'. Using peptides, it was shown that different types of transposases also bind competitively to the same site on  $\beta$ <sup>21</sup> and we show here that the same principle applies to the transposase of several members of the IS1634 family. The affinity with which enzymes bind  $\beta$  is likely to be finely-tuned in the cell to accommodate the various processes for which  $\beta$  is an essential component<sup>24,40</sup>. Since  $\beta$ -binding by a transposase could be a burden on the host, it would be unlikely to find strong motifs in transposases. Also, the relatively simplicity of



**Figure 5.** *In vivo* assay to study the effect of different  $\beta$  binding motifs in IS1634 transposition.

(a) Plasmid design of vector pSKT1-IS1634 (See Methods). The sequences for the left (IR-L) or right (IR-R) inverted repeats is shown. The sequences of the wild type and mutants 5A and CN is shown (amino acids 562-566 of IS1634 Tnp). (b) Papillation assay of transposition. The pictures show representative examples of *E. coli* colonies for the three versions of the transposase (WT and 5A or CN mutants), and a negative control C (pSKT1 containing the inverted repeats but no transposase gene). (c) Quantification of the papillation assay for 8 colonies (C =  $1.51 \pm 0.6$ ; WT =  $3.15 \pm 0.5$ ; 5A =  $2.05 \pm 0.85$ ; CN =  $13.0 \pm 2.95$ ).



the  $\beta$  binding motif, which often resides in unstructured sequences, could be created *de novo* by random mutation and selection. This could explain the appearance of  $\beta$  motifs at distinct locations in homologous transposases<sup>21</sup>, as it is the case here with the secondary motif identified at the C-terminal sequence of *Acidithiobacillus ferrivorans* IS1634 (Fig. 4a). Likewise, the fact that the motif appears consistently in non-homologous transposases with very different transposition pathways seems to favour the idea that it has appeared independently many times in the evolutionary history of this diverse group of proteins<sup>21</sup>.

Our results show that interaction with  $\beta$  should not be a strong barrier for IS exchange among phylogenetically-distant but habitat-sharing organisms<sup>3,41</sup>. Because ISs are subject to inactivation by mutation, horizontal transfer is an essential aspect of their lifestyle and critical for the persistence of ISs in bacterial populations<sup>2,12</sup>. *Acidiphilium* IS1634 Tnp can bind to *Leptospirillum* or *Escherichia*  $\beta$ , and even to an archaeal sliding clamp, PCNA from *Methanosarcina*. Although  $\beta$  and PCNA share no detectable sequence similarity, these proteins are structurally and functionally very similar, and the binding motifs of the proteins interacting with them are highly related<sup>23,24,29</sup>. Our results suggest that transposases could interact relatively easily and bidirectionally with the replication machinery of bacterial or archaeal hosts. With a few exceptions (e.g., ISH6, found only in Archaea), most IS families can be found in bacterial and archaeal genomes, and genome sequencing suggest that movement of these mobile elements between the two domains is fluid<sup>42</sup>. However, a recent survey of prokaryotic IS elements in eukaryotic genomes detected few events of IS transfer in recent eukaryotic evolutionary history, with the possible exception of cyanobacterial IS607<sup>43</sup>. Although HGT events from prokaryotic to eukaryotes have been documented extensively<sup>44,45</sup>, it remains to be investigated if mobile elements are involved. In eukaryotic organisms, PCNA-binding has been detected for Pogo, a *Drosophila* transposase<sup>28,46</sup> and, recently, for the endonuclease/reverse transcriptase of LINE-1, where it has been shown to be critical for retrotransposition<sup>47</sup>.

By interacting with  $\beta$ , transposases are targeted to the replication fork, thus associating chromosomal replication with transposition. Coupling of transposition to replication would ensure, first, the possibility of recombination and repair with the sister chromosome, which would limit the potential damage to the host caused by IS excision and the generation of double-strand breaks. Second, IS elements with a conservative (cut and paste) mechanism of transposition would have a chance of jumping to the sister chromosome and therefore increase their number in that replicon. Third, transposition would co-localize with repair factors required in the later stages of transposition to fill in the gaps and ligate (e.g., both DNA polymerase I and DNA ligase interact with  $\beta$ <sup>48</sup>). Yet another possibility is that transposases couple  $\beta$ -binding with allosteric changes that initiate transposition, such as monomer-dimer transitions or binding to the end sequences of the IS. Transposases could require  $\beta$  to be released from a state of self-inhibition which has often been observed in these enzymes<sup>35</sup>. Formation of the transpososome involves dimerization and conformational changes which often involve C-terminal regions of the protein<sup>49,50</sup> which are, almost always, also the sites of interaction with  $\beta$ .

Although transposase binding to  $\beta$  assures integration with the host's chromosomal replication, it remains to be investigated if an excess of transposase, or a transposase with strong affinity for  $\beta$ , would be disruptive for DNA synthesis. It is remarkable that all enzymes involved in mutagenic processes, namely the DNA polymerases, the mismatch repair system, and transposition, directly interact with the  $\beta$  sliding clamp. Indeed, a recent experimental study demonstrated a direct 'conflict' between these processes, as mismatch repair mutator alleles, often present in bacterial populations, limit insertion sequence proliferation in the early stages of invasion in a new host<sup>14</sup>. Further, different IS families could directly compete among them for access to a limiting host resource, the  $\beta$  sliding clamp, as predicted by the genome ecosystem hypothesis<sup>51</sup>. The relation between IS and host genomes has been described as a dynamic equilibrium state, which may become transiently perturbed by the rapid expansion of IS populations and that can be explained just with two parameters: the duplication-deletion ratio and the LGT-deletion ratio<sup>52</sup>. We propose that the ability of transposases to interact with  $\beta$  is one of the molecular processes that contribute to the "duplication" component of the first parameter and, as such, it is subject to natural selection. Variations in the binding site that increase the affinity for  $\beta$  may be selected upon colonization of new hosts and cause explosive proliferation of IS. On the contrary, variations that decrease the affinity for  $\beta$  would reduce transposition rates, contributing to a reduction in IS populations. Future studies will be required to understand how these various processes contribute modulate genetic variability and to the adaptive potential of bacterial populations.

The *in vivo* assay reported here (Fig. 5) shows that transferring IS1634 from *Acidiphilium* sp. PM to *Escherichia coli* is greatly favoured by point mutations in the motif that make the interaction stronger (Fig. 3c,d). Our IS1634 transposase CN mutant defines a new category of hyperactive transposases: one with improved interface with the host. The search for hyperactive transposases has been based typically in genetic screens of random point mutants which are then combined in a single transposase to greatly increase its activity. For example, commercial Tn5 transposase contains three amino acid changes which increase activity by >1000-fold<sup>55</sup>, PiggyBac mutations in seven amino acids generate a 17-fold increase in activity<sup>53</sup>, and a combination of multiple mutations in Sleeping Beauty can generate up to a 100-fold increase in transposition<sup>54</sup>. More recent studies have shown that mariner transposases could be easily made hyperactive by mutation of the dimerization interface, thus disrupting the autorregulatory allosteric mechanism between the subunits that limits transposition<sup>55,56</sup>. The principles derived from our designed hyperactive IS1634 mutant could potentially open the door for development of novel hyperactive transposases for use in biotechnology, or for the design of transposases with an expanded range of hosts.

## Methods

**Bacterial strains and long-term cultivation.** *Acidiphilium* sp. PM (DSM 24941) was originally isolated from the Tinto river in Huelva (Spain)<sup>57</sup> and its genome sequenced<sup>25</sup>. The sequenced strain was used to initiate a 100 ml culture in GYE medium, which is composed of a mineral salt solution (0.2% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.01% KCl, 0.033% K<sub>2</sub>HPO<sub>4</sub>·3H<sub>2</sub>O, 0.025% MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.0014% Ca(NO<sub>3</sub>)<sub>2</sub>·4H<sub>2</sub>O) supplemented with 0.2% (w/v) glucose and 0.01% (w/v) yeast extract<sup>58</sup>. The pH was adjusted to 2.5 with 1 N H<sub>2</sub>SO<sub>4</sub> prior to autoclaving (111 °C, 0.5 atm, 30 min). Cultivation took place at 30 °C with vigorous shaking. Serial transfers (1:50 dilution) took place in periods of 14 days. Under these conditions the cultures reached stationary phase in 5 days and the culture experienced 9-day-periods in nutrient-deprived medium. The 1:50 dilution allowed ~5.6 generations (log<sub>2</sub>50) per serial dilution, or 584 generations in 4 years. Samples from populations were taken periodically and stored at –80 °C.

**Oligonucleotide microarray of IS-related genes.** To design an oligonucleotide microarray that could detect changes in copy number and gene expression of mobile elements in acidophilic organisms, we selected 12 fully sequenced acidophilic genomes (including *Acidiphilium* sp. PM) (manuscript in preparation). We then used RPS Blast (NCBI) to scan 40,231 proteins identified in these genomes with a set of 149 position-specific scoring matrix (PSSM) profiles, selected from Pfam and corresponding to mobile element proteins. This procedure identified a set of 1,358 transposases and associated proteins. The corresponding gene sequences were progressively clustered by means of a customized pipeline that used UCLUST<sup>59</sup> and Cons<sup>60</sup> to generate a set of 769 consensus sequences characterized by having at least two conserved blocks of minimal length equal to 50 bp and zero ambiguities. The program Array Designer 4 (Premier Biosoft International) was then used to design two oligonucleotides for each consensus sequence (1,538 oligonucleotides) with an average length of 40 nt and a constant estimated melting temperature (72 °C). See Table S1 for sequences corresponding to *Acidiphilium* IS-related ORFs. Importantly, any IS containing more than one ORF (e.g., one that contains the transposase and an accessory element), will be represented by oligonucleotides from each of these ORFs. In addition to the oligonucleotides representing the mobile genes present in the 12 genomes of acidophiles, we included also three reference genes (*rpoB*, *dnaX* and *gyrB*) for each of these genomes. Spotting was carried out with the MicroGrid-TAS II Arrayer (Genomic Solutions, Huntingdon, UK) at 22 °C and 50–60% relative humidity on epoxy-substrate slides (Arrayit Corp.) according to the manufacturer's instructions. An array containing 10,752 spots (including three replicas) was constructed.

**qPCR.** DNA was extracted from both the initial and final time-points of the long-term *Acidiphilium* sp. pure culture by using Gnome<sup>®</sup> DNA Isolation Kit (MP Biomedicals). The concentration and quality of extracted DNA was measured by a Qubit<sup>®</sup> Fluorometer (Life Technologies) according to manufacturer's recommendations. The primer sequence used in the qPCR were oligo #1 and oligo #2 for the *Acidiphilium* IS1634 transposase gene; and oligo #3 and #4 for the *dnaX* gene (See Table S2 for oligonucleotide sequences). PCR reactions (25 µL) were performed with 0.4 µM primers (each), *Acidiphilium* sp. PM '2011' target genomic DNA at two different final concentrations (40 pg/µL and 8 pg/µL) and 12.5 µL of iQ SYBR<sup>®</sup> Green Supermix (Life Technologies) according to the manufacturer's instructions. Two biological and three technical replicates were prepared for each gene and each DNA concentration. Reactions were carried out with a MyiQ<sup>®</sup> Single-Colour Real-Time PCR Detection System (Bio-Rad). Cycling parameters comprised an initial cycle of 3 minutes at 95 °C followed by 40 cycles of 30 seconds at 95 °C, 30 seconds at 60 °C and 30 seconds at 72 °C. A melting curve analysis was performed for each reaction to rule out non specific reaction products or primer dimers. '2007' genomic DNA was serially diluted in 5-fold increments from 5 ng to 1, 6 pg. These serial dilutions with three technical replicates, were used to create a standard curve for the reference *dnaX* gene ( $r = 0.992$ ) and another curve for IS1634 transposase gene ( $r = 0.997$ ). The genomic copy numbers of *Acidiphilium* IS1634 and *dnaX* in '2011' time-point was determined from their corresponding '2007' standard curves where both genes are present in a single copy per chromosome, using the comparative threshold cycle method (CT). Results in Fig. 1D represent the average of '2011' *dnaX* and *Acidiphilium* IS1634 chromosomal gene copy number and the standard deviation (SD) of the replicates.

**Inverse PCR.** Inverse PCR was used to amplify the flanking regions of the *Acidiphilium* IS1634 insertion sites detected in the '2011' culture. For this purpose 1 µg of '2007' and '2011' *Acidiphilium* sp. genomic DNA, extracted as previously described, were digested with EcoRI (NEB), an enzyme that does not cut within the transposase sequence, for 2 h at 37 °C in 20 µL. Digestions were ligated in 200 µL reaction volume with T4 ligase (NEB) according to the manufacturer's recommendation. A PCR was performed using 1 µL of ligation products and 1 µM of each divergent primers #5 and #6 (See Table S2). Inverse PCR reaction products of '2007' and '2011' time-points were resolved and visualized in a 1% agarose gel. DNA fragments were extracted of the gel with QIAquick Gel Extraction Kit (QIAGEN) and sequenced.

**Protein purifications.** *E. coli* β, *Methanosarcina barkeri* PCNA and GST-PolIV<sup>LF</sup> were cloned, overexpressed and purified as described<sup>21</sup>. *Acidiphilium* sp. PM β was amplified by PCR using oligonucleotides

#7 and #8 (See Table S2), cloned into pET16b (Novagen) as a NcoI/BamHI fragment, and overexpressed in *E. coli* BL21(DE3). 10 g of cells (dry weight) were resuspended in Buffer G (100 mM Tris-HCl, 1 M NaCl, 2 mM EDTA, 10% glycerine, 1 mM  $\beta$ -mercaptoethanol, 1 mM PMSF, pH 8) and processed four times with a French press at 4 °C. The lysed cells were diluted to 150 mL and centrifuged (15000 g, 30 m, 4 °C).  $\beta$  was mostly found in pelleted fraction in the form of inclusion bodies. The pellet was washed three times with 20 mL of buffer G + 1% Triton<sup>®</sup> X-100 (Bio-Rad) to remove cell membranes and residual membrane proteins, and then centrifuged (15000 g, 30 m, 4 °C). The inclusion body pellet was then washed three times with 20 mL of buffer G to remove residual Triton<sup>®</sup> X-100. In a subsequent step, inclusion bodies were solubilized in 5 mL of buffer H (100 mM Tris-HCl, 100 mM NaCl, 2 mM EDTA, 10% glycerine, 1 mM  $\beta$ -mercaptoethanol, 1 mM PMSF, pH 8) + 6 M guanidine-HCl and incubated (25 °C, 15 m). Insoluble material was removed by centrifugation 15 minutes at 15000 g. The solubilised inclusion bodies were slowly drop by drop diluted in 400 mL of buffer H with constant stirring at 4 °C and allowed to refold for 1 hour. The solution was centrifuged 20 minutes at 15000 g to remove insoluble material. This clarified solution was applied on a 30 mL Q sepharose FF (GE Healthcare) ion-exchange chromatography column equilibrated in Buffer H. Protein was eluted with Buffer H over a NaCl gradient (120 mL, 0.1–1.0 M). Fractions containing  $\beta$  were pooled and dialyzed against 2 L of Buffer I (50 mM Tris-HCl, 50 mM NaCl, 1 mM EDTA, 10% glycerine, 1 mM DTT, pH 8). This fraction (1.3 mg mL<sup>-1</sup>) was aliquoted and stored at -80 °C.

To amplify the  $\beta$  gene from *Leptospirillum ferrooxidans* by PCR we used oligos #9 and #10 (See Table S2). These oligos introduced a sequence encoding a FLAG epitope at the N-terminus of the gene. The PCR product was cloned as a BglII/NheI fragment into pET11a (Novagen). *L. ferrooxidans*  $\beta$  formed inclusion bodies and was purified following the same protocol as for *Acidiphilium* sp. PM  $\beta$ . The final yield was 1.4 mg mL<sup>-1</sup>.

The gene for the IS1634 transposase from *Acidiphilium* sp. PM (Tnp) was amplified by PCR from genomic DNA with oligonucleotides #11 and #12 (See Table S2), cloned in vector pET16b, and sequenced. The 5A and CN Tnp mutants were created by site-directed mutagenesis of this plasmid by QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies). Wild-type (WT) and mutant Tnps were overexpressed in Rossetta<sup>®</sup> 2 (DE3) pLysS (Merck Millipore) with 1 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) overnight at 25 °C. After centrifugation, the cell paste was resuspended in Buffer J (50 mM Tris-HCl, 400 mM NaCl, 2 mM EDTA, 10% glycerine, 1 mM  $\beta$ -mercaptoethanol, 0.1% Triton X-100, pH 8) and lysed in a French press. Cells were diluted to 150 mL in the same buffer and centrifuged (15,000 g, 30 min, 4 °C). All three transposases formed inclusion bodies. The pellet was washed three times with 20 mL of buffer J supplemented with 1% Triton X-100, and then washed another three times with 20 mL of buffer J. Afterwards, the pellet was dissolved for 30 minutes at 25 °C in 5 mL of Buffer J + 6 M guanidine-HCl, insoluble material was removed by centrifugation (30 m, 15,000 g). This solution was diluted drop by drop in 400 mL of buffer J with constant stirring at 4 °C and allowed to refold for 1 h. Soluble protein was applied to a 15 mL Heparin Sepharose FF (GE Healthcare) ion-exchange chromatography column equilibrated in the same Buffer. Transposases were eluted with Buffer J and a NaCl gradient (120 mL, 0.4–1.0 M), 1.5 mL fractions were collected and those of them containing transposase were dialyzed against 2 L of Buffer J. Tnp was obtained at concentrations of 0.8 mg mL<sup>-1</sup> for WT; 0.5 mg mL<sup>-1</sup> for 5A, and 0.5 mg mL<sup>-1</sup> for CN.

**Acidiphilium IS1634 Tnp pull-down assay.** The binding assay to test the interaction of IS1634 Tnp (WT, 5A and CN) with  $\beta$  clamp of *Escherichia coli*, *Acidiphilium* sp. and *Leptospirillum ferrooxidans*, used Dynabeads M-280 Tosyl-activated (Invitrogen). The reactions (50  $\mu$ L) contained 12  $\mu$ M of Tnp covalently coupled to 1.2 mg of magnetic beads in Binding Buffer A (phosphate-buffered saline, 0.1% Tween-20, 0.1% BSA, pH 7.2). 5  $\mu$ M of each  $\beta$  labeled with Alexa Fluor 350 C5-maleimide (Life Technologies) were incubated (15 m, 25 °C) with the Tnp-coated beads, and washed three times with Binding Buffer A to remove unbound  $\beta$ . Reactions were stopped with 1% SDS, subjected to SDS-PAGE electrophoresis and visualized on a UV transilluminator.

**IS1634 transposase peptides pull-down assay.** Peptides were obtained from ProteoGenix SAS (Schiltigheim, France). IS1634 transposase derived peptides from *Acidiphilium* sp. PM, *Acidithiobacillus ferrivorans* and *Methanosarcina barkeri* were assayed for the interaction with  $\beta$  clamp or PCNA. 400  $\mu$ M of biotinylated peptides were mixed in 50  $\mu$ L in Binding Buffer B (50 mM Tris, 50 mM NaCl, 5% glycerine, 0.1% BSA) with 1 mg of Dynabeads M-270 Streptavidin (Invitrogen), incubated (30 m, 25 °C) and washed three times with the same buffer. 4  $\mu$ M of labeled *Eco* $\beta$ , *Lf* $\beta$ , *Ac* $\beta$  or 3  $\mu$ M labeled *Mba*PCNA were added to the beads in a reaction volume of 50  $\mu$ L with Binding Buffer B, incubated (15 m, 25 °C), and washed three times with the same buffer. Reactions were stopped with 1% SDS, loaded on a SDS-PAGE and analyzed on a UV transilluminator.

**Cloning of IS1634 into pSKT1.** The pSKT1 plasmid was generously provided by H. Savilahti (University of Turku, Finland) and used as described<sup>32</sup>. Oligos for cloning of right inverted repeat (RIR) of IS1634 were #13 and #14 (Table S2); cloning of the left inverted repeat (LIR) were #15 and #16 (Table S2); and cloning of the transposase as a NcoI/EcoRI PCR product were #11 and #17 (Table S2).

## References

1. Siguier, P., Goubeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014).
2. Cerveau, N., Leclercq, S., Bouchon, D. & Cordaux, R. in *Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution* (ed. Pontarotti, P.) 291–312 (Springer-Verlag, Berlin Heidelberg, 2011).
3. Hooper, S. D., Mavromatis, K. & Kyrpides, N. C. Microbial co-habitation and lateral gene transfer: what transposases can tell us. *Genome Biol.* **10**, R45 (2009).
4. Toussaint, A. & Chandler, M. in *Bacterial Molecular Networks: Methods and Protocols* (ed. van Helden, J.) (Springer Science, 2012).
5. Juhas, M. *et al.* Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**, 376–93 (2009).
6. Cunneen, M. M. & Reeves, P. R. The *Yersinia kristensenii* O11 O-antigen gene cluster was acquired by lateral gene transfer and incorporated at a novel chromosomal locus. *Mol. Biol. Evol.* **24**, 1355–1365 (2007).
7. DiRuggiero, J. *et al.* Evidence of recent lateral gene transfer among hyperthermophilic Archaea. *Mol. Microbiol.* **38**, 684–693 (2000).
8. Degnan, P. H., Yub, Y., Sisneros, N., Wing, R. A. & Moran, N. A. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc. Natl. Acad. Sci. USA* **106**, 9063–9068 (2009).
9. Nagy, Z. & Chandler, M. Regulation of transposition in bacteria. *Res. Microbiol.* **155**, 387–98 (2004).
10. Iranzo, J., Gomez, M. J., Lopez de Saro, F. J. & Manrubia, S. Large-scale genomic analysis suggests a neutral punctuated dynamics of transposable elements in bacterial genomes. *PLoS Comput. Biol.* **10**, e1003680 (2014).
11. Le Rouzic, A. & Capy, P. The first steps of transposable elements invasion: parasitic strategy vs. genetic drift. *Genetics* **169**, 1033–1043 (2005).
12. Bichsel, M., Barbour, A. D. & Wagner, A. The early phase of a bacterial insertion sequence infection. *Theor. Popul. Biol.* **78**, 278–88 (2010).
13. Lohe, A. R., Moriyama, E. N., Lidholm, D. & Hartl, D. L. Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Mol. Biol. Evol.* **12**, 62–72 (1995).
14. Fehér, T. *et al.* Competition between transposable elements and mutator genes in bacteria. *Mol. Biol. Evol.* **29**, 3153–3159 (2012).
15. Treangen, T. J., Abraham, A. L., Touchon, M. & Rocha, E. P. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* **33**, 539–71 (2009).
16. Newton, I. L. G. & Bordenstein, S. R. Correlations between bacterial ecology and mobile DNA. *Curr. Microbiol.* **62**, 198–208 (2010).
17. Moran, N. A. & Plague, G. R. Genomic changes following host restriction in bacteria. *Curr. Opin. Genet. Dev.* **14**, 627–33 (2004).
18. Parkhill, J. *et al.* Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat. Genet.* **35**, 32–40 (2003).
19. Roberts, D., Hoopes, B. C., McClure, W. F. & Kleckner, N. IS10 transposition is regulated by DNA adenine methylation. *Cell* **43**, 117–130 (1985).
20. Chandler, M. Clamping down on transposon targeting. *Cell* **138**, 621–3 (2009).
21. Gómez, M. J., Díaz-Maldonado, H., González-Tortuero, E. & López de Saro, F. J. Chromosomal replication dynamics and interaction with the  $\beta$  sliding clamp determine orientation of bacterial transposable elements. *Genome Biol. Evol.* **6**, 727–740 (2014).
22. Parks, A. R. *et al.* Transposition into replicating DNA occurs through interaction with the processivity factor. *Cell* **138**, 685–95 (2009).
23. Dalrymple, B. P., Kongsuwan, K., Wijffels, G., Dixon, N. E. & Jennings, P. A universal protein-protein interaction motif in the eubacterial DNA replication and repair systems. *Proc. Natl. Acad. Sci. USA* **98**, 11627–11632 (2001).
24. López de Saro, F. J. Regulation of interactions with sliding clamps during DNA replication and repair. *Current Genomics* **10**, 206–215 (2009).
25. San Martín Úriz, P., Gómez, M. J., Arcas, A., Bargiela, R. & Amils, R. Draft genome sequence of the electricigen *Acidiphilium* sp. strain PM (DSM 24941). *J. Bacteriol.* **193**, 5585–5586 (2011).
26. De Palmenaer, D., Siguier, P. & Mahillon, J. IS4 family goes genomic. *BMC Evol. Biol.* **8**, 18 (2008).
27. Bunting, K. A., Roe, S. M. & Pearl, L. H. Structural basis for recruitment of translesion DNA polymerase Pol IV/DinB to the  $\beta$ -clamp. *EMBO J.* **22**, 5883–5892 (2003).
28. Warbrick, E. The puzzle of PCNA's many partners. *Bioessays* **22**, 997–1006 (2000).
29. Hedglin, M., Kumar, R. & Benkovic, S. J. Replication clamps and clamp loaders. *Cold Spring Harb. Perspect. Biol.* **5**, a010165 (2013).
30. Georgescu, R. E. *et al.* Structure of a small-molecule inhibitor of a DNA polymerase sliding clamp. *Proc. Natl. Acad. Sci. USA* **105**, 11116–21 (2008).
31. López de Saro, F. J., Georgescu, R. E., Goodman, M. F. & O'Donnell, M. Competitive processivity-clamp usage by DNA polymerases during DNA replication and repair. *EMBO J.* **22**, 6408–6418 (2003).
32. Pajunen, M. I. *et al.* Universal platform for quantitative analysis of DNA transposition. *Mobile DNA* **1**, 24 (2010).
33. Martusewitsch, E., Sensen, C. W. & Schleper, C. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J. Bacteriol.* **182**, 2574–81 (2000).
34. Kleckner, N. Regulating Tn10 and IS10 transposition. *Genetics* **124**, 449–454 (1990).
35. Reznikoff, W. S. Transposon Tn5. *Annual Review of Genetics* **42**, 269–286 (2008).
36. Sousa, A., Bourgard, C., Wahl, L. M. & Gordo, I. Rates of transposition in *Escherichia coli*. *Biol. Lett.* **9**, 20130838 (2013).
37. Allen, E. E. *et al.* Genome dynamics in a natural archaeal population. *Proc. Natl. Acad. Sci. USA* **104**, 1883–8 (2007).
38. Mueller, R. S. *et al.* Ecological distribution and population physiology defined by proteomics in a natural microbial community. *Mol. Syst. Biol.* **6**, 374 (2010).
39. Hewson, I. *et al.* In situ transcriptomic analysis of the globally important keystone N<sub>2</sub>-fixing taxon *Crocospaera watsonii*. *ISME J.* **3**, 618–631 (2009).
40. Maul, R. W., Ponticelli, S. K., Duzen, J. M. & Sutton, M. D. Differential binding of *Escherichia coli* DNA polymerases to the beta-sliding clamp. *Mol. Microbiol.* **65**, 811–827 (2007).
41. Wagner, A. & de la Chaux, N. Distant horizontal gene transfer is rare for multiple families of prokaryotic insertion sequences. *Mol. Genet. Genomics* **280**, 397–408 (2008).
42. Filé, J., Siguier, P. & Chandler, M. Insertion sequence diversity in Archaea. *Microbiol. Mol. Biol. Rev.* **71**, 121–157 (2007).
43. Gilbert, C. & Cordaux, R. Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. *Genome Biol. Evol.* **5**, 822–832 (2013).
44. Schonknecht, G. *et al.* Gene transfer from bacteria and archaea facilitated evolution of an extremophilic eukaryote. *Science* **339**, 1207–1210 (2013).



45. Hotopp, J. C. D. *et al.* Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**, 1753–1756 (2007).
46. Warbrick, E., Heatherington, W., Lane, D. P. & Glover, D. M. PCNA binding proteins in *Drosophila melanogaster*: the analysis of a conserved PCNA binding domain. *Nuc. Acids Res.* **26**, 3925–32 (1998).
47. Taylor, M. S. *et al.* Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* **155**, 1034–1048 (2013).
48. López de Saro, F. J. & O'Donnell, M. Interaction of the  $\beta$  sliding clamp with MutS, ligase, and DNA polymerase I. *Proc. Natl. Acad. Sci. USA* **98**, 8376–8380 (2001).
49. Braam, L. A. M., Goryshin, I. Y. & Reznikoff, W. S. A mechanism for Tn5 inhibition – carboxyl terminal dimerization. *J. Biol. Chem.* **274**, 86–92 (1999).
50. Dyda, F., Chandler, M. & Hickman, A. B. The emerging diversity of transpososome architectures. *Quart. Rev. Biophys.* **45**, 493–521 (2012).
51. Brookfield, J. F. Y. The ecology of the genome – mobile DNA elements and their hosts. *Nat. Genet. Rev.* **6**, 128–136 (2005).
52. Iranzo, J., Gómez, M. J., López de Saro, F. J. & Manrubia, S. Large-scale genomic analysis suggests a neutral punctuated dynamics of transposable elements in bacterial genomes. *PLoS Comput. Biol.* **10**, e1003680 (2014).
53. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. USA* **108**, 1531–1536 (2011).
54. Mátés, L. *et al.* Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.* **41**, 753–761 (2009).
55. Lampe, D. J. Bacterial genetic methods to explore the biology of mariner transposons. *Genetica* **138**, 499–508 (2010).
56. Liu, D. & Chalmers, R. Hyperactive mariner transposons are created by mutations that disrupt allosterism and increase the rate of transposon end synapsis. *Nuc. Acids Res.* **42**, 2637–2645 (2014).
57. Malki, M., De Lacey, A. L., Rodríguez, N., Amils, R. & Fernandez, V. M. Preferential use of an anode as an electron acceptor by an acidophilic bacterium in the presence of oxygen. *App. Environ. Microbiol.* **74**, 4472–4476 (2008).
58. Harrison, A. P. The acidophilic thiobacilli and other acidophilic bacteria that share their habitat. *Annu. Rev. Microbiol.* **38**, 265–292 (1984).
59. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
60. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
61. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).

## Acknowledgments

The authors gratefully thank Harri Savilahti (University of Turku, Finland) for vector pSKT1, and to Marina Postigo (Centre for Astrobiology, Spain) for sequencing all genetic constructs. We also thank Enrique González-Tortuero (Centre for Astrobiology, Spain) for discussion and bioinformatics support. This work was funded by the Subdirección General de Proyectos de Investigación of the Spanish Ministry for Economy and Competitiveness Grants No. CGL2010-17384 and AYA2011-24803, and by the European Research Council (ERC) Advanced Grant No. 250350 (IPBSL). HDM has a FPI fellowship from the Spanish Government.

## Author Contributions

H.D.M. performed most experimental work. M.J.G. performed most of the bioinformatics work. M.M.P. and V.P. constructed the oligonucleotide microarray. P.S.M.U. and R.A. performed the laboratory evolution experiment with *Acidiphilium* and provided expertise with acidophilic organisms. F.J.L.S. conceived and designed all research, and performed some experimental work. H.D.M., M.J.G. and F.J.L.S. wrote the manuscript. All authors read and approved the final version of this 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:** Díaz-Maldonado, H. *et al.* Transposase interaction with the  $\beta$  sliding clamp: effects on insertion sequence proliferation and transposition rate. *Sci. Rep.* **5**, 13329; doi: 10.1038/srep13329 (2015).



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/>