Cox Murray (Orcid ID: 0000-0003-1936-0236)
Sitter Thomas (Orcid ID: 0000-0002-7160-6240)

Title: Evolution of virulence in a novel family of transmissible mega-plasmids

Thomas L. Sitter[12ϕ] ORCID:0000-0002-7160-6240,

Amy L. Vaughan[1,2] ORCID:0000-0003-0309-8851,

Marion Schoof[1,2] ORCID:0000-0001-9788-8947,

Simon A. Jackson[3] ORCID:0000-0002-4512-3093,

Travis R. Glare[2] ORCID:0000-0001-7795-8709,

Murray P. Cox[2,4] ORCID 0000-0003-1936-0236

Peter C. Fineran[2,3] ORCID 0000-0002-4639-6704

Paul P. Gardner[2,5] ORCID 0000-0002-7808-1213

Mark R. H. Hurst[1,2] ORCID:0000-0001-5826-5253

[1] Forage Science, AgResearch, Lincoln Research Centre, Christchurch, New Zealand

[2] Bio-Protection Research Centre, Lincoln, New Zealand

[3] Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand

[4] Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North, New Zealand

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15595

[5] Department of Biochemistry, University of Otago, Dunedin, New Zealand

Φ Corresponding Author: Thomas L. Sitter

E-mail: lesleysitter@hotmail.com, lesleysitter@gmail.com

Phone: +45 1573 7895508

Address: Prager Str. 61, 04317, Leipzig, Germany

Summary

Some *Serratia entomophila* isolates have been successfully exploited in biopesticides due to their ability to cause amber disease in larvae of the Aotearoa (New Zealand) endemic pasture pest, *Costelytra giveni*. Anti-feeding prophage and ABC toxin complex virulence determinants are encoded by a 153-kb single-copy conjugative plasmid (pADAP; amber disease-associated plasmid). Despite growing understanding of the *S. entomophila* pADAP model plasmid, little is known about the wider plasmid family. Here, we sequence and analyze mega-plasmids from 50 *Serratia* isolates that induce variable disease phenotypes in the *C. giveni* insect host. Mega-plasmids are highly conserved within *S. entomophila*, but show considerable divergence in *Serratia proteamaculans* with other variants in *S. liquefaciens* and *S. marcescens*, likely reflecting niche adaption. In this study to reconstruct ancestral relationships for a complex mega-plasmid system, strong co-evolution between *Serratia* species and their plasmids were found. We identify twelve distinct mega-plasmid genotypes, all sharing a conserved gene backbone, but encoding highly variable accessory regions including virulence factors, secondary metabolite biosynthesis, Nitrogen fixation genes and toxin-antitoxin systems. We show that the variable pathogenicity of *Serratia* isolates is largely caused by presence/absence of virulence clusters on the mega-plasmids, but notably, is augmented by external chromosomally encoded factors.

Importance

We identified a new family of mega-plasmids related to a previously identified amber disease associated plasmid (pADAP) of *Serratia entomophila*. Members of this family are single copy and encode a highly conserved backbone while bearing highly diverse auxiliary genes that likely reflect plasmid and therefore bacterial host fitness. We observed that plasmids can evolve at different rates based on the carrying species.

Despite being trans-species conjugatable we showed the ability of the chromosome to modulate disease phenotype caused by plasmid encoded virulence determinants. Evidence presented suggests that various plasmids have likely co evolved to their host strain from where their further transfer to other species was likely stymied by plasmid handcuffing, a process where two plasmids dimerize as a mechanism to maintain a low copy-number.

The evolution of plasmids is an understudied area with few well-studied models. This study resolves the in-depth relationship between all members of the newly identified mega-plasmid family, defined novel elements acquired through horizontal gene transfer and correlated genotype with disease phenotype. Additionally, a new R tool called Roary_stats, was developed that can rapidly process protein orthology output from the Roary software, and automatically generate meta-data as well as absence/presence matrices, to allow a quick visual comparison between genomes. Using the results from this study, we were able to construct an evolutionary map of the plasmid family, which suggests the plasmids have evolved from a common ancestor but followed divergent pathways.

Keywords: Mega-plasmid, pADAP, pathogenicity, insecticidal, plasmidome

Introduction

Bacterial plasmids are a major driving force behind the rapid dissemination of antibiotic (Dolejska and Papagiannitsis, 2018) and pesticide (Rangasamy *et al.*, 2017) resistance. Additionally, plasmids can quickly evolve through horizontal gene transfer and recombination (Norberg *et al.*, 2011), and be transferred between bacteria via conjugative machinery (Hu *et al.*, 2019), making them reservoirs of transmissible genes (Stalder *et al.*, 2019). Plasmid replication and the expression of plasmid-based genes can impose a metabolic burden on cells (Raymond and Bonsall, 2013; San Millan and MacLean, 2017) and should therefore be negatively selected for, while beneficial genes should eventually be integrated into the chromosome (Bergstrom *et al.*, 2000). This is often referred to as the 'plasmid paradox'. Yet plasmids are often not negatively selected for (Carroll and Wong, 2018) and persistence has been observed even under neutral selection (Harrison and Brockhurst, 2012; MacLean and San Millan, 2015). This persistence is most likely because plasmids can undergo compensatory evolution to mitigate the burden to their bacterial host (Zwanzig *et al.*, 2019), increasing plasmid stability (Wein *et al.*, 2019) and sometimes contribute to bacterial fitness (Dionisio *et al.*, 2005) as exemplified by the pSymA and pSymB mega-plasmids carrying genes that facilitate microbe-plant interactions in rhizobia (Maróti and Kondorosi, 2014).

Plasmids can also be key drivers of virulence, as is the case for pFra/pMT1 and pPla/pPCP plasmids found in *Yersinia pestis* (Demeure *et al.*, 2019). These plasmids encode the type III secretion system-derived *Yersinia* outer proteins (Yops), which suppress the host immune response (Grabowski *et al.*, 2017). Other host organisms, such as members of the *Vitis* genus, commonly known as grapevines, can be affected by the tumor-inducing (Ti) and root-inducing (Ri) plasmids of the *Agrobacterium tumefaciens* which cause crown gall and hairy root (Ridé *et al.*, 2000). Some strains of the insecticidal Cry toxin-encoding bacterium *Bacillus thuringiensis* have been exploited in biopesticides (Brar *et al.*, 2006). Distinct Cry toxins can affect various invertebrate hosts (Palma *et al.*, 2014) and are often encoded on lineage-specific plasmids (Méric *et al.*, 2018).

In Aotearoa (New Zealand), the Gram-negative bacterium *Serratia entomophila* (Adeolu *et al.*, 2016; Grimont *et al.*, 1977), is used in biopesticides against larvae of the endemic pasture pest *Costelytra giveni* (Coleoptera: Scarabaeidae) (Johnson *et al.*, 2001; Townsend *et al.*, 2004), commonly known as grass grub or tūtae ruru in Māori. The anti-insect properties of this bacterium are carried by a 153-kb amber disease-associated plasmid (pADAP) (Glare *et al.*, 1993; Hurst and Glare, 2002) (Fig. 1A), which encodes two virulence determinants, an insect-active ABC toxin complex (Bowen *et al.*, 1998) called the *S. entomophila* pathogenicity (Sep) complex (Hurst *et al.*, 2000) (Fig. 1B), and a contractile injection system known as the anti-feeding prophage (Afp) (Hurst *et al.*, 2004) (Fig. 1C). Together, Sep and Afp particles cause amber disease in *C. giveni* larvae (Hurst *et al.*, 2004; Hurst *et al.*, 2000), a chronic condition characterized by voidance of the gut and cessation of feeding that can last several months and results in larval death (Jackson *et al.*, 1993). Apart from these insecticidal virulence determinants, pADAP has a core backbone encoding replication and stability genes, as well as a type IV conjugative pilus (Pil) and a putative fimbria (Sef) (Hurst *et al.*, 2011).

Beyond *S. entomophila*, some field-isolated *S. proteamaculans* isolates also display varying pathogenicity towards *C. giveni* larvae (Dodd, 2003; Hurst *et al.*, 2018). Restriction enzyme profiling suggested that insecticidal *S. entomophila* isolates all share a similar plasmid, whereas insecticidal *S. proteamaculans* isolates carry more diverse plasmids (Dodd, 2003). So far several variants of pADAP, as well as other mega-plasmids with regions homologous to pADAP, have been discovered. For example, a non-pADAP plasmid p49, found in *Yersinia frederiksenii* isolate 49, encodes a Sep ortholog termed TcYF (Dodd *et al.*, 2006); a pADAP variant pU143, found in *S. proteamaculans* strain 143, lacks the Afp-encoding region and additionally encodes a Sep variant designated as Spp (*S. proteamaculans* pathogenicity) (Hurst *et al.*, 2011); and the pADAP variant pAGR96X from *S. proteamaculans* strain AGR96X encodes an Afp variant, termed AfpX (Hurst *et al.*, 2018). The presence of these variants suggests that the pathogenicity clusters found on pADAP are diverse and widely distributed. Additionally, the *S. proteamaculans* strain AGR96X causes rapid death in *C. giveni* larvae instead of inducing chronic amber disease. *S. proteamaculans* strain AGR96X is also bioactive

against larvae of the Aotearoa endemic scarab species *Pyronata festiva* and *Pyronata setosa*, which are not affected by pADAP carrying *S. entomophila* isolates (Hurst *et al.*, 2018).

In combination, these studies suggest that pADAP variants are diverse in terms of their gene content and virulence mechanisms. Here, we investigated the genetic variation of *Serratia*-based pADAP mega-plasmids and identify how their genetic makeup correlates with disease phenotypes. All 52 analyzed plasmids, including pADAP itself and one *Serratia marcescens* plasmid found in GenBank, contain a ~63-kb conserved backbone, encode highly diverse accessory clusters, suggesting that they originated from a common ancestral plasmid that diverged into quite different evolutionary pathways. Based on the conserved nature of this backbone, this collection of plasmids are designated STAMPs (*Serratia* transmissible adaptive mega-plasmids). We demonstrate that the high variability of accessory determinants directly correlates to the differing disease phenotypes. Additionally, we show that the genome background can augment the pathogenic responses of plasmid encoded virulence determinants, revealing a higher level of complexity in the pADAP model system than previously recognized.

Results and Discussion

A robust dataset of *Serratia* plasmids

To investigate the genetic variation and evolutionary history of pADAP variants, and their varying bioactivity towards *C. giveni* larvae, 50 *Serratia* isolates from different species, with different disease phenotypes (Table S1) and geographical origins (Fig. S1), were sequenced and compared with the pADAP reference sequence (Hurst *et al.*, 2011) (Fig. 1). Through comparative genomics, phylogenetic analysis, conjugation and plasmid retention experiments, these plasmids were analyzed to gain insight into the causes underlying variance in pathogenicity, as well as to establish the evolutionary origin of the pADAP-like class of mega-plasmids.

pADAP are single copy plasmids with close relatives in the *Serratia* genus

To first establish the phylogenetic relationship between pADAP variants and other plasmids, we compared the replication proteins (Rep) (del Solar *et al.*, 1998) of the *S. entomophila* pADAP and *S. proteamaculans* pAGR96X and p1137 to that of other plasmids in GenBank (Fig. S2AB, Table S4). The proteins most closely related to pADAP RepA were observed in *Serratia*. The only RepA encoding isolates exhibiting insecticidal properties were those isolated from Aotearoa. Importantly, a separate RepA phylogeny revealed that all *S. entomophila* RepA proteins cluster together uniquely with the RepA from two *S. proteamaculans* plasmids, p145 and p142 (Fig. S2D).

Proteins of the autoregulating RepA family bind to 12–18-bp repeat motifs, termed iterons, positioned 5' of *repA* (Chattoraj, 2000) (Fig. S2C), and employ the chromosomal primosome for plasmid replication (Messer, 2002). Twenty copies of the 16-bp iteron repeat motif variants (5'-CAkrAAAtNNNTCCaC-3') (Hurst *et al.*, 2011) were observed in a conserved pattern in each pADAP variant, and at varying numbers in other *repA*-associated plasmids (Fig. S2C). Iteron-bound RepA from two plasmids can dimerize, termed 'handcuffing', blocking further *repA* transcription (Schumacher *et al.*, 2014) and plasmid replication (Molina-García *et al.*, 2016) thus regulating plasmid copy number. Previous observations that *S. liquefaciens* 377, upon introduction of pADAP, lost its RepA encoding p377 (Grkovic *et al.*, 1995), suggests that these RepA plasmids form their own incompatibility group (Novick, 1987). This is corroborated through sequence comparison of read coverage between chromosome and plasmid contigs for several PacBio assemblies, which indicate that the pADAP variants were single-copy (Fig. S3). These plasmids appear to employ RepA and iterons to mediate replication, maintain a low plasmid copy number, reduce horizontal transmission of these plasmids, and therefore decreasing the chance to recombine with similar RepA encoding plasmids.

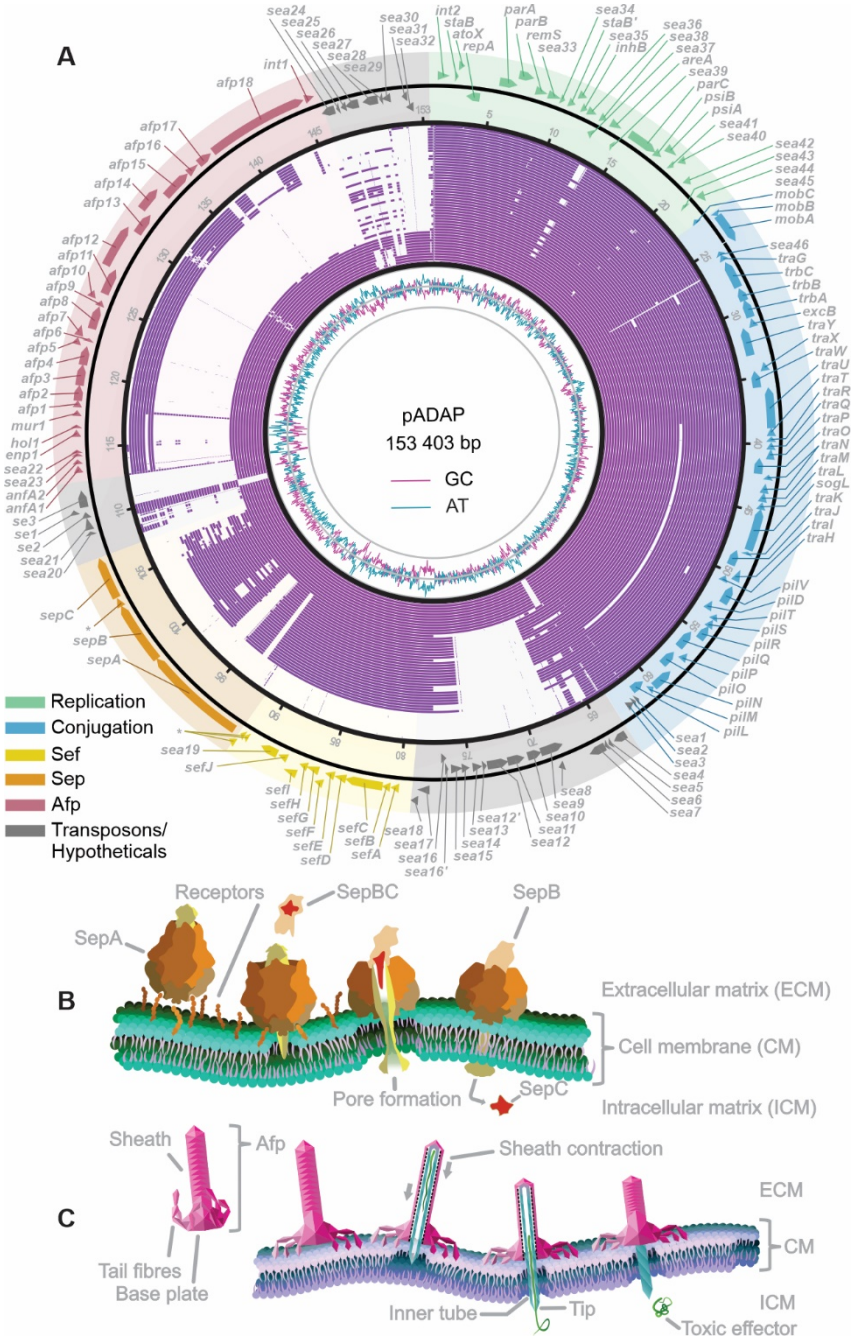

Fig. 1: Variants of *Serratia entomophila* strain A1MO2 pADAP all exhibit homology to the backbone, but diverge at the Sep and Afp encoding regions. A) The outer ring represents the previously annotated pADAP features (GenBank accession NC_002523) (Hurst *et al.*, 2011). Shaded on the outer ring are the replication region (green), the region of conjugation (blue), the Sef fimbria cluster (yellow), the Sep operon (brown), and the Afp features (burgundy). Grey regions are clusters of hypothetical and transposon-like elements. The purple middle rings show homology between the 50 sequenced pADAP variant plasmids compared to the pADAP reference plasmid, which were sorted manually based on visual similarity. The inner ring represents the percent GC content of the pADAP

plasmid. Asterisks denote hypothetical coding sequences. B) An illustration of the mode of action of the pADAP derived Sep proteins based on models by Gatsogiannis *et al.* (2013). C) An illustration of the mode of action of Afp proteins produced by pADAP based on models by Desfosses *et al.* (2019).

Boundaries of the pADAP plasmid backbone

To compare the 50 sequenced plasmids against the pADAP reference, it was important to first define their backbone. The gene *int2* encodes a XerD-family tyrosine recombinase (Colloms, 2013), and was previously designated the start of the pADAP backbone (Hurst *et al.*, 2011). XerD is involved in resolving chromosomal and plasmid dimers during replication in *Escherichia coli* to ensure DNA segregation during cell division (Castillo *et al.*, 2017). Regions surrounding *int2* were aligned, which revealed a clear point 524-bp 5′ of *int2* where pADAP variants diverged (Fig. S4). This point of demarcation was subsequently designated as the first nucleotide of each plasmid. A deletion-induced filamentation (*dif*) recombination site, required for site-specific Xer-mediated recombination, was identified 464-bp 5′ of *int2*. The *int2* gene was co-located with *repA* as well as a stability/partitioning locus containing *staAB* stability genes (Guynet *et al.*, 2011), partition genes *parABC* (Gerdes *et al.*, 2000), and two putative SOS inhibitors *psiA* (Bagdasarian *et al.*, 1986) and *psiB* (Petrova *et al.*, 2010).

We identified three sequences in GenBank that had similarities to the pADAP replication region: the *S. marcescens* isolate B3R3 unnamed1 plasmid, the *S. marcescens* isolate FZSF02 (Lin *et al.*, 2019) and the *S. marcescens* WVU-005 isolate WVU-005-1 plasmid, which encodes most of this region except for the *int2*/*repA* locus (Fig. S5). Similarly, sequence identity of the pADAP conjugation region was observed to the WVU-005-1 plasmid, with lesser nucleotide similarity to the *Salmonella enterica* serovar Infantis plasmids R64 and pESI (Fig. S6). The co-location of these Tra/Pil clusters on diverse plasmids from different bacterial species is suggestive of a strong association between the Tra and Pil operons, which likely reflects their structural association and their importance in plasmid dissemination through conjugation (Hu *et al.*, 2019). Unexpectedly, the *S. proteamaculans* plasmid p1769 has a 23-kb truncation of this conjugation region (from *traT* to *pilL*; Fig. S7).

A DNA alignment was made of the ~63-kb backbone of all 52 pADAP variants, including the *S. marcescens* WVU-005-1 plasmid as it contained a ~44-kb homologous region to the pADAP backbone with ~79% DNA similarity. This backbone spanned from the point of demarcation and including *repA* and *int2*, up to the anti-sense *pilL* gene and revealed a high level of nucleotide sequence identity (78%–100%) (Fig 1, Fig. S7).

Manual curation of backbone alignments revealed only minor variations (Fig. S8, Table S5). A 5 339-bp insertion between *traG* and *trbC*, designated the divergence marker, was found in 31 *S. proteamaculans* and the two *S. liquefaciens* plasmids, but in none of the *S. entomophila* pADAP plasmids (Fig. S9, Table S6). Additionally, *S. proteamaculans* plasmid p465 contained a ~37-kb insertion between *pilO* and *pilN*, which encodes phage head, baseplate, tail, DNA packaging and terminase components (Fig. S7, Fig. S10A, Table S7), suggestive of a functional bacteriophage designated *S. proteamaculans* bacteriophage (Spb). Additional plasmidome meta-data (Fig. S11D) generated using a custom build R tool called Roary_stats (https://github.com/IamIamI/pADAP_project/), revealed that there are 39 core backbone genes, which are conserved among ≥95% of all plasmids. Overall, the core pADAP backbone comprises the highly conserved replication and conjugation regions, which typically span ~63-kb, with occasional variances due to horizontal gene transfer.

Plasmidome comparison reveals 12 distinct genotypes and novel accessory clusters

Based on the presence of the conserved backbone, found in all 52 *Serratia* plasmids, including pADAP and the *S. marcescens* WVU-005-1, these plasmids were designated STAMPs (*Serratia* transmissible adaptive mega-plasmids). Despite encoding a conserved backbone, these STAMPS encode divergent accessory regions 5′ of the *dif* recombination site (Fig. 2, Fig. S11E). A highly divergent horizontal gene transfer hotspot was observed 5′ of *pilL* between the backbone and the accessory clusters, which variously encodes a wide range of transposons, insertion sequence (IS) elements and other genes (Fig. S12, Table S8). Though limited in number, orthologous sequences to that of STAMP

members are found from different parts of the world, similar to findings of other studies such as the comprehensive analysis on the pT26-2 plasmid family by Badel *et al.* (2019).

In addition to the known Sep and Afp cluster variants, Spp (Hurst *et al.*, 2011) and AfpX (Hurst *et al.*, 2018), respectively, several novel accessory clusters were identified. One such region is a ~39-kb later outlined region of unknown function (RUF) (Fig. S10B) that was near-identical (~99% nt id) between *S. proteamaculans* plasmids p149, p336, p465, p1769, p1770, p1772, p12a, p12d, p28F and pM (Fig. 2, genotypes C, H, J, and K). This high conservation suggest it may encode traits that are beneficial to the plasmid. The RUF in STAMPs belonging to genotypes C and K is in the opposing orientation relative to that of genotype H and J plasmids (Fig. S10B), likely either due to independent acquisition or a rearrangement event. In addition, p299, p1457 and pD (genotype I) encode a truncated RUF, while pK (genotype C) does not encode a RUF.

The diverse nature of the accessory gene clusters identified in this study suggests functional specialization of the plasmids. Studies such as those by Sentchillo *et al.* (2013) show that through unbiased sequencing of plasmids from separate locations, one can find plasmids encoding similar maintenance functions but diverse ancillary genes, similar to our finding. Based on the gene content of the accessory regions, we classified STAMPs into 12 distinct genotypes (type A–L) (Fig. 2).

Assessments of codon bias (Fig. S11A), amino acid frequency (Fig. S11B) and stop codon bias (Fig. S11C), revealed no discernible differences among STAMPs. Stop codon analysis showed a low affinity for TAG but equal affinity for TAA and TGA, correlating with %GC in agreement with models by Povolotskaya *et al.* (2012). These results suggest a long coevolutionary relationship between STAMPs and the *Serratia* genus.

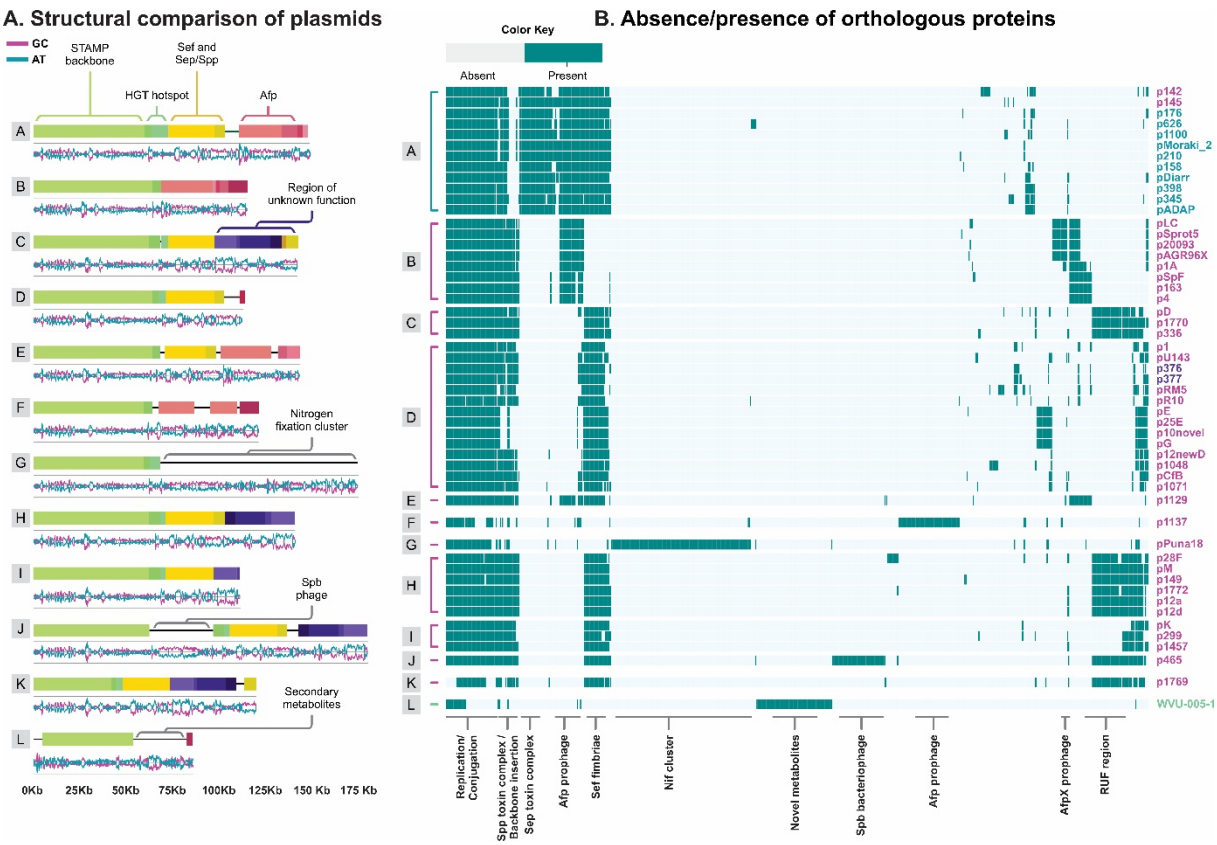

Fig. 2 Plasmidome comparison and protein orthology analysis revealed 12 distinct STAMP genotypes. A) The *Serratia entomophila* pADAP reference plasmid (GenBank accession NC_002523) (Hurst *et al.*, 2011) is depicted as type A, while the plasmid obtained from *Serratia proteamaculans* strain AGR96X (Hurst *et al.*, 2018) is depicted as type B. Colors in the synteny map symbolize regions of interest, including the STAMP backbone (light green), horizontal gene transfer hotspot (dark green), Sef (light blue), Sep/Spp (dark blue), Afp/AfpX (purple), and a novel region of unknown function (red). Synteny blocks were based on manual curation of the progressiveMauve (Darling *et al.*, 2010) alignment of all pADAP-type plasmids. The %GC plot was generated with a 50-bp sliding window. B) A protein orthology analysis was performed using Roary (Page *et al.*, 2015), and processed through our custom Roary_stats tool (https://github.com/IamIamI/pADAP_project/) to generate an absence/presence matrix of orthologous proteins among STAMPs. Isolate names are indicated on the right of the protein orthology matrix. Blue labels are *S. entomophila*, green *Serratia marcescens*, purple *Serratia liquefaciens*, and pink *S. proteamaculans*. Refer to Fig. S13 for more detail of the absence/presence matrix shown in this figure.

STAMPs are transferred by conjugation and are highly stable

Next, we tested whether the different STAMPs (genotypes A, D, E, and H) were transferable among *Serratia* species by conjugation. We tagged *S. proteamaculans* p142, pU143, p145, p149 and p1129, along with *S. entomophila* p210, p398 and p1100 (Table S2), with antibiotic resistance markers and used these isolates as plasmid donors. We detected conjugation into *S. entomophila* 5.6 (pADAP cured) (Glare *et al.*, 1993) and naturally plasmid-free *S. proteamaculans* isolate 3041 (Table S9). These results corroborate previous studies on pADAP (Glare *et al.*, 1996; Grkovic *et al.*, 1995)..

To determine if STAMPs are easily lost, a 10-transfer stability experiment was performed, using various isolates carrying antibiotic-marked STAMPs (genotypes A-K). Patched colonies from serial dilutions of passaged cultures at day-10, revealed 100% plasmid retention under L Broth Base (LB), mitomycin C (MitC) induction and minimal media growth conditions (Table S3B). Additionally, using flow cytometry, GFP-marked STAMPs showed an average ~90% STAMP retention for all culture conditions (Table S3C). Comparison of *S. entomophila* A1MO2 (pADAP+) to *S. entomophila* 5.6 (pADAP-) in several stress conditions, such as mitomycin C-mediated SOS-response (Fig. S14), revealed that the plasmid imparts some growth benefits to host bacteria in stress conditions. These findings are in line with studies of conjugatable plasmids such as the *E. coli* R1 plasmids, which shows that plasmids can evolve from being a burden to its bacterial host cell to conveying bacterial fitness over a relatively short time (Dionisio *et al.*, 2005). STAMPs appear highly stable in their associated isolate but also, importantly, in cross-species settings.

AfpX encoding STAMPs carry toxin-antitoxin systems.

Plasmids can further enforce their maintenance in a bacterial population by employing toxin-antitoxin systems (Diaz-Orejas *et al.*, 2017). These systems inhibit the growth of progeny that do not inherit a stable plasmid copy. Two toxin-antitoxin clusters were identified downstream of all AfpX variant clusters. The *S. proteamaculans* p4, p163, p1129, p1A and pSpF plasmids (Fig. 2, genotypes B and E) encoded a colicin M-like lipid II-degrading bacteriocin (El Ghachi *et al.*, 2006) and resistance

protein (Ghequire *et al.*, 2017) (Fig. S10E, Table S10), which likely function as secreted anti-microbial that ensures stable inheritance of the plasmid (Fig. S15A).

In *S. proteamaculans* pAGR96X, pSprot5, pLC and p20093, a ~13-kb region downstream of the AfpX cluster encoded a ribonuclease PilT N-terminal domain, which contains a VapC toxin and a VapB antitoxin (Deep *et al.*, 2017; Matelska *et al.*, 2017) (Fig. S10F, Table S10). VapC cleaves initiator transfer RNA N-formylmethionine (Cruz *et al.*, 2015), required for initiation of translation and is proposed to lead to a persister cell state (Roy *et al.*, 2020) (Fig. S15B). However, a stability assay between *S. proteamaculans* AGR96X strains carrying pAGR96X::Cm^R and a *vapBC* deleted pAGR96XΔVapBC mutant showed no difference in growth or plasmid retention (Table S11). In summary, multiple toxin-antitoxin systems are present on AfpX encoding STAMPS, and potentially play a role in the evolutionary stability of these plasmids (Bardaji *et al.*, 2019).

The region of unknown function does not contribute to pathogenicity

Serratia isolates carrying RUF encoding STAMPs exhibit varied virulence towards *C. giveni*. The first observed RUF region was observed on *S. proteamaculans* p149, which encodes only RUF and the Spp toxin complex. The p149 RUF region encodes 37 putative proteins including products that have limited amino acid similarity to partitioning and replication components, IS elements, a tyrosine recombinase (Table S12). Two predicted operons were identified, which were designated ruf1 and ruf2 (Fig. 10B). The ruf1 cluster encodes several putative fimbrial elements and two exotoxins, whereas ruf2 mostly encodes hypotheticals and two putative RHS core domain-containing proteins (Table S12). The homologous protein sequences listed in Table S12 are the most closely related gene products in the NCBI Reference Sequence database (RefSeq; O'Leary *et al.*, 2016). Although there are some genes with *Serratia* based origins (Table S12), BLASTN assessment of the two identified operons yielded no substantial hits.

The presence of exotoxins in ruf1 prompted the hypothesis that there was some association with bioactivity in *C. giveni*. To test this, bioassays were performed on *C. giveni* using *S. proteamaculans*

isolate 149 derivatives with deletions of ruf1, ruf2 and the A and B components of Spp, designated p149Δruf1, p149Δruf2 and p149ΔSppAB, respectively. Twelve days post-challenge, larvae fed p149Δruf1 or p149Δruf2 showed similar pathology to wild-type isolate 149 (Table S13A), whereas larvae challenged with p149ΔSppAB remained healthy. We then tested whether the RUF was associated with bioactivity in different insect hosts by inoculating *Adoryphorus couloni*, *Odontria* spp., *Acrossidius tasmaniae*, *P. festiva* and *P. setosa* larvae with *S. proteamaculans* isolate 149. However, all challenged individuals remained healthy (Table S13B). Together, these data suggest that RUF is unlikely to be directly involved with inducing amber disease in *C. giveni* or related insect pests.

Accessory clusters with alternate host associations

Two STAMPs did not encode any of the known pADAP virulence determinants but instead encode elements that may enable the survival of the bacterium in other niches than *C. giveni* larvae. *S. proteamaculans* pPuna18 (genotype G) shared ~63-kb (94.1% nt id) with the pADAP backbone. In place of virulence determinants, pPuna18 encodes genes involved in fixation of nitrogen (N_2) to ammonia (NH_3), including three separate ATP-binding cassette operons, two flavodoxin proteins (Freigang *et al.*, 2002), and a ~24-kb *nif* cluster (Poudel *et al.*, 2018) (Fig. S10C, Table S14). The pPuna18 *nif* cluster, containing *nifAB,E,L-N,Q,S-Z*, was similar (90.4% nt id) to pRahaq202 (GenBank accession CP003246.1) in *Rahnella aquatilis*, a plant growth promoting symbiont of grapevines (Chen *et al.*, 2007).

The *S. marcescens* WVU-005-1 STAMP (genotype L) harbors a ~31-kb region, encoding a range of hypothetical proteins, transposon-like elements, and genes of unknown function (Fig. S10D, Table S15). A ~7 Kb subregion was homologous to *Enterobacter hormaechei* subsp. *hoffmannii* isolate Eh1 plasmid p1 (GenBank accession CP034755.1) (99% nt id). Both *S. marcescens* and *E. hormaechei* subsp. *hoffmannii* species are associated with nosocomial infections (Khanna *et al.*, 2013; Townsend *et al.*, 2008). Based on the functional descriptions of genes encoded on WVU-005-1 and pPuna18 STAMPs, these STAMPs are likely not associated with *C. giveni* and may have helped their host enter

alternate niches. Certainly, it is clear that pADAP-like plasmids play a wide range of functional roles, not solely related to pathogenicity.

No clear correlation between STAMP encoded proteins and virulence

Apart from the aforementioned *S. proteamaculans* pPuna18 and *S. marcescens* WVU-005-1, all remaining isolates encoded either Sep/Spp, Afp/AfpX or both virulence determinants. Isolates associated with Sep/Spp and Afp/AfpX encoding plasmids exhibit divergent disease phenotypes, even among those with similar gene clusters. *In vitro* and *in vivo* stability experiments revealed that the Sep and Afp regions themselves are stable, with no detected mobility (gain or loss) of marked pathogenicity clusters (Table S3A). Attempts to correlate genotypes with pathogenicity traits, using a Roary derived orthology matrix, (Fig. 2) proved unsuccessful. The orthology matrix revealed homogeneity among putative proteins encoded by *S. entomophila* STAMPs but more heterogeneity among the putative proteins encoded by *S. proteamaculans*-derived plasmids (Fig. 2).

To investigate the role of the chromosome in disease responses, we challenged *C. giveni* larvae with the transconjugant *S. entomophila* 5.6 and *S. proteamaculans* 3041 carrying p210, and *S. entomophila* 5.6 carrying p1100. All transconjugants displayed disease symptoms similar to those induced by the plasmid donor isolate (Table S13C). However, transconjugants of *S. entomophila* 5.6 and *S. proteamaculans* 3041 carrying p145 showed a lower disease response ($p \leq 0.05$) than the *S. proteamaculans* 145 wildtype isolate. These findings suggest that genes on the chromosome might have a much larger effect on pathogenicity augmentation than previously assumed.

Phylogenetic inference of STAMPs shows lineage-specific plasmids

As shown in the previous section, STAMPs can be trans-conjugated across *Serratia* species. It was therefore important to assess whether there was evidence of STAMPs transferring to other hosts. To determine this, a 16S gene-based phylogenetic tree was generated along with a phylogenetic tree

summary constructed from 13 individual core STAMP backbone genes (*repA*, *parA*, *parB*, *parC*, *mobA*, *mobB*, *mobC*, *trbA*, *trbB*, *trbC*, *traU*, *traW*, *traX)* (Fig. 3). Similar to studies performed on Cry-encoding plasmids from *Bacillus thuringiensis* (Méric *et al.*, 2018) and the *Rickettsia* pRICO derivatives (El Karkouri *et al.*, 2016), we identified lineage-dependent plasmids in the STAMP dataset, such as pADAP STAMPs that were almost exclusively associated with *S. entomophila*. Two direct homologs of the pADAP STAMP p142 (98.5% nt id) and p145 (97.7% nt id) were carried by *S. proteamaculans* isolates 142 and 145 (Fig. 2, genotype A), and show that pADAP is not confined to *S. entomophila*. The pADAP homologs are largely confined to their own cluster, which mirrors the *S. entomophila* 16s cluster, in line with the findings of Dimitriu et al. (2019) who showed that plasmid transfer is more likely between bacteria of the same clone. However, there is poor correlation between the *S. proteamaculans* 16S tree and the clustering of their corresponding plasmids. This suggests a higher frequency of horizontal transmission of *S. proteamaculans* STAMPs compared with *S. entomophila*, in which there is a more vertical transmission pattern. Additionally, no *S. proteamaculans* or *S. liquefaciens* STAMPs were observed in *S. entomophila* isolates, indicating that their transfer to *S. entomophila* could be more tightly regulated, potentially through some chromosomally encoded mechanism. The *S. liquefaciens* plasmids p376 and p377 branch off from *S. proteamaculans* STAMPs, which indicates a more recent plasmid acquisition by *S. liquefaciens* from *S. proteamaculans* and is corroborated by the RepA phylogeny (Fig. S2D). The comparison between the *S. marcescens* WVU-005-1 plasmid and the other STAMPs confirmed a shared deep ancestry but also suggests a prolonged separation. Though limited in phylogenetic resolution, comparison of the 16S and STAMP backbone tree (Fig. 3) revealed no clue as to why related *S. proteamaculans* isolates often carry diverse STAMPs as opposed to *S. entomophila* which only carries pADAP homologs.

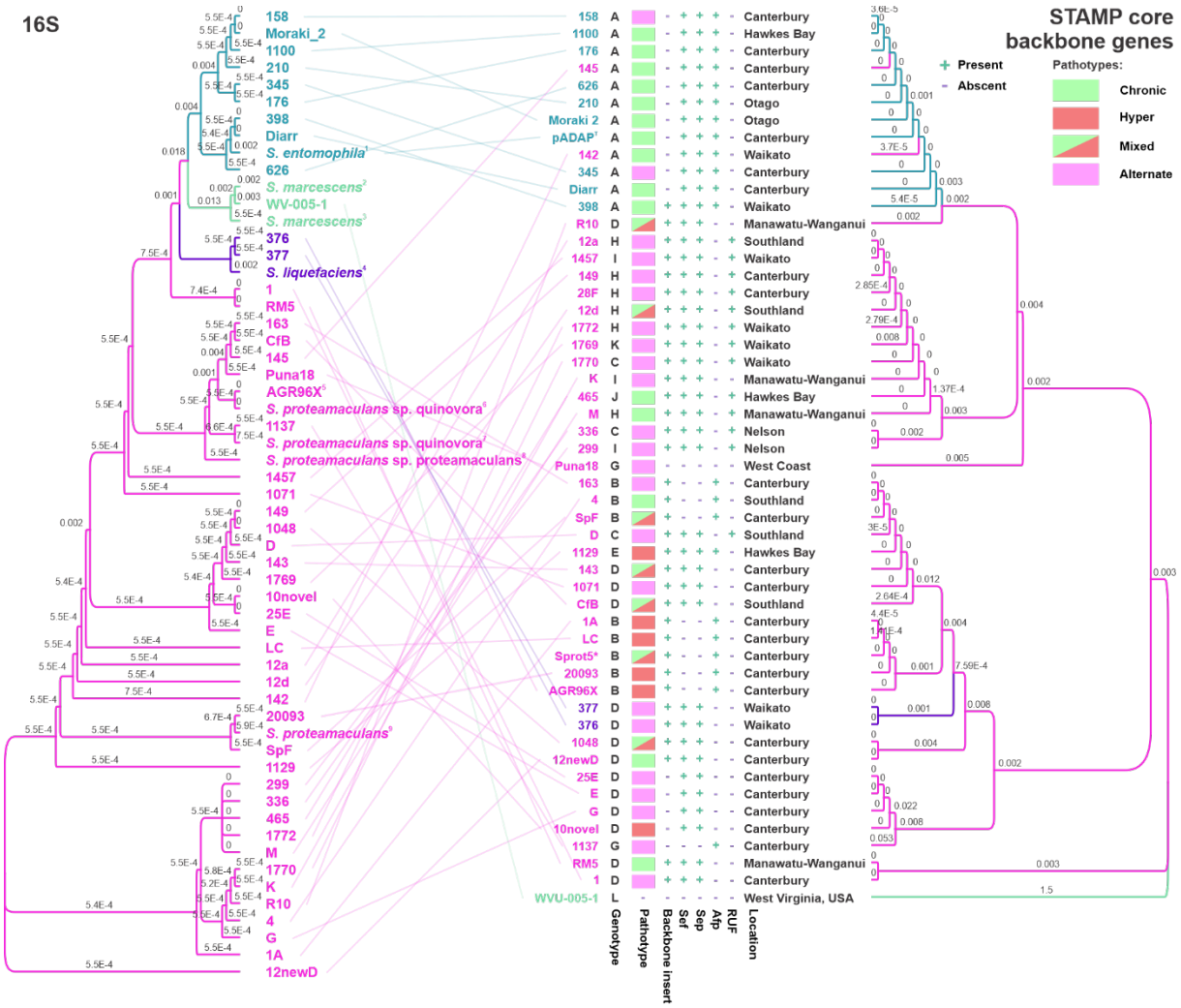

Fig. 3 The *Serratia proteamaculans* based STAMPs appear less associated with specific *S. proteamaculans* isolates as opposed to *Serratia entomophila* pADAP homolog STAMPs. 16S gene-based maximum likelihood cladogram tree tangled with the STAMP core backbone gene summary cladogram tree. Isolates labeled in blue are *S. entomophila*, green are *Serratia marcescens*, purple are *Serratia liquefaciens,* and pink are *S. proteamaculans*. Additional isolate information shows a lack of correlation between any identifier and a specific plasmid genotype. Nucleotide substitutions per site are represented on the branches. References used for 16S tree construction: [1] *S. entomophila* [AJ233427], [2] *S. marcescens* [AJ233431], [3] *S. marcescens* [AB061685], [4] *S. liquefaciens* [AJ306725], [5] *S. proteamaculans* sp. quinovora [KY934181], [6] *S. proteamaculans* sp. quinovora [AJ279044], [7] *S. proteamaculans* sp. proteamaculans [AJ233435], [8] *S. proteamaculans* [AJ2333434]. An untransformed maximum likelihood backbone gene summary tree can be observed in Fig. S16.

An evolutionary model of STAMPs

Based on the relatively conserved composition of STAMPs backbones an attempt was made to reconstruct the evolution of the STAMP family genotypes A-L (Fig. 4). Slight differences in the backbone regions, and the conserved region of demarcation associated with *int2* and its associated *dif* site (Fig. S4) both suggest that site-specific recombination might reflect a significant mode of acquisition of accessory clusters as shown in other systems (Blakely *et al.*, 1993; Brovedan *et al.*, 2019). However, the *S. marcescens* WVU-005-1 plasmid encodes neither the *int2* nor the *repA* genes, but instead encodes for a RepB family replication protein, indicating that it is a more distant relative to the other STAMPs and suggesting that the other backbone elements could have recombined before the replication genes became attached. The presence of RUF in two orientations makes it challenging to determine whether acquisition is a result of an independent recombination event or a reorientation event. The presence of the conjugation machinery consisting of a relaxosome (*tra*) and a Type IV pili (*pil*) operon (Hu *et al.*, 2019) most likely propagated STAMP persistence in the population. These STAMPs share a conserved backbone, yet evolved into quite different pathways from that of the apparent co-evolution between pADAP and *C. giveni*. This finding is in agreement with other studies that identified plasmids with similar maintenance functions but diverse ancillary genes based on unbiased sequencing of plasmids from separate locations (Ho *et al.*, 2011; Sentchilo *et al.*, 2013).

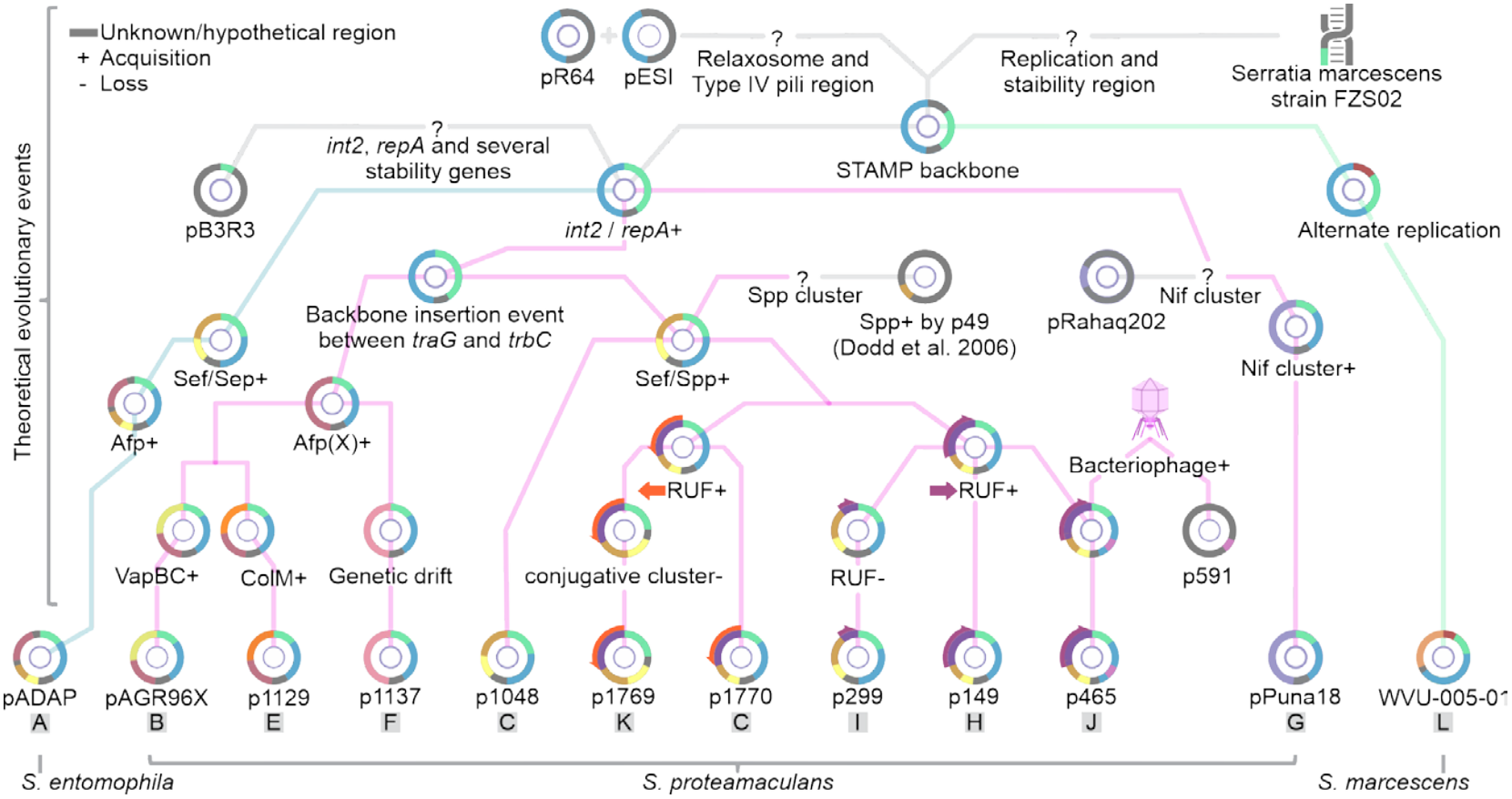

Fig. 4 A Pedigree chart of the STAMP family of plasmids reveals theorized evolutionary events leading up to the current day STAMPS. The manually constructed chart is based on the assumption that regions observed more frequently predate those observed less frequently. Blue branches denote *Serratia entomophila*, green *Serratia marcescens*, pink *Serratia proteamaculans*. Representative plasmids from each genotype are presented at the bottom. Refer to Fig. 2 for additional details on subtle inter-genotype differences.

In addition to potential fitness benefits, pADAP provides the host cell with Sep and Afp, which allows *C. giveni* larvae invasion. The range of pathogenic responses induced by the STAMP carrying *Serratia* isolates is most likely a result of plasmid encoding only one of the two known *C. giveni* pathogenic virulence determinant variants, Afp/AfpX or Sep/Spp, but is influenced in addition by chromosomal factors, as noted in the *S. entomophila* 5.6 and *S. proteamaculans* 3041 carrying the *S. proteamaculans* p145 transconjugant. It is assumed that pathogenicity clusters that induce amber disease are in a predator-prey co-evolutionary relationship with *C. giveni.* If true, these insecticidal protein encoding STAMPs would only be attuned to *C. giveni* or other related endemic beetles as a result of geographical isolation.

The *S. entomophila* A1MO2 (pADAP+) type-strain (Grimont *et al.*, 1988) and *S. proteamaculans* AfpX encoding AGR96X (Hurst *et al.*, 2018) were bioactive against larvae of the Aotearoa endemic *C. giveni* and, in the instance of AGR96X, also bioactive against larvae of the endemic *P. festiva* and *P. setosa*. The closely related invasive pests *A. tasmaniae* and *A. couloni*, originating from Australia, are unaffected by all tested isolates. These findings allude to evolution through geographical isolation and is further strengthened by the discovery of novel accessory clusters such as those encoded by the *S. proteamaculans* isolate Puna18 plasmid pPuna18 and the *S. marcescens* WVU-005 isolate WVU-005-1 found in the United States. Additional analysis on plasmid free *Serratia* isolates, and *Serratia* isolates that carry non-STAMP plasmids, will be necessary to understand the transmission of STAMPs and STAMP accessory determinants, and to determine how pathogenicity is augmented by chromosomal components.

Materials and Methods

Bacterial strains, vectors and culture conditions

Fifty distinct *S. entomophila*, *S. proteamaculans* and *S. liquefaciens* isolates were sequenced (Table S1, Fig. S1). All *Serratia* isolates were cultured in Lennox L Broth Base (LB; Invitrogen) liquid medium, Miller Luria-Bertani base agar (Merck) solid medium or M9 minimal liquid or solid medium (Elbing and Brent, 2019) containing 0.4% casamino acids (carbon source) at 30°C. *Escherichia coli* strains were cultured in LB liquid or solid medium at 37°C. Bacterial strains and vector systems used in this study are listed in Table 1.

Table 1: Strains and vectors

Type	Reference	Cassette	Description
Plasmids			
pACYC184	(Chang and Cohen, 1978)	Cm^R	Chloramphenicol O-acetyltransferase
		Tc^R	Tetracycline efflux MFS transporter
pACYC177	(Chang and Cohen, 1978)	Km^R	APH(3')-I family aminoglycoside O-phosphotransferase
		Ap^R	Beta-lactamase TEM-1 variant
pHP45Ω	(Prentki and Krisch, 1984)	Sm^R / Sp^R	Streptomycin 3'-adenylyltransferase
pUCP30T-GFPmut3	(Barbier and Damron, 2016)	GFPmut3	Gm^R, Green fluorescent protein mut3.1
pJP5608	(Riedel *et al.*, 2013)		Tc^R, Suicide vector
XΔNOVa	Hurst [unpublished data]		pAGR96X mutant with *vapBC* complex knock out
E. coli strain			
ST18	(Thoma and Schobert, 2009)		Δ*hemA*, requires 5-aminolevulinic acid complementation
Serratia isolates			
A1MO2	(Grimont *et al.*, 1988)		*S. entomophila* A1MO2 carrying pADAP (pADAP+)
5.6	(Glare *et al.*, 1993)		pADAP heat-cured (pADAP-) *S. entomophila* A1MO2 derivative strain 5.6
3041 / Tukino	(Glare *et al.*, 1993)		Naturally plasmid-free *S. proteamaculans* isolate 3041

Ap, ampicillin; Cm, chloramphenicol; Gm, gentamicin; Km, kanamycin; Sm, streptomycin; Sp, spectinomycin; Tc, tetracycline

^{R} Resistance

Genome assembly and bioinformatic analysis

DNA was extracted using an ISOLATE II Genomic DNA Kit (Bioline) as per the manufacturer's instructions and sequenced by Macrogen (South Korea) using either 100-bp paired-end reads on the HiSeq 2500 platform (Illumina) or the PacBio RSII system (Pacific Biosciences). Illumina short reads were trimmed using Trim Galore v0.6.1 (Krueger, 2019) and assembled using A5-miseq v20160825 (Coil *et al.*, 2015). Sequence assemblies were scaffolded using SSPACE v3.0 (Boetzer *et al.*, 2011) and ambiguous base-calls resolved using GapFiller v1-10 (Nadalin *et al.*, 2012). PacBio reads were assembled using Canu v2.0 (Koren *et al.*, 2017) and base-corrected with PILON v1.23 (Walker *et al.*, 2014). Circlator v1.5.5 (Hunt *et al.*, 2015) was used to resolve overhangs on circular consensus reads.

The previously sequenced *S. entomophila* strain A1MO2 pADAP plasmid (GenBank accession NC_002523) (Hurst *et al.*, 2011) was used as the reference plasmid. Plasmid contigs were identified using nucleotide BLAST (basic local alignment search tool) (Camacho *et al.*, 2009) searches against the pADAP sequence or elements thereof. Assemblies yielded 50 plasmids that were confirmed to contain homologous pADAP regions. The first ten contigs were checked with PCR and DNA sequencing to validate the assembly pipeline (Table S1). Additionally, the publicly available WVU-005-1 plasmid from *S. marcescens* isolate WVU-005 (GenBank accession PRJNA545504) was included in the study.

Plasmid annotation was performed using Prokka v1.13 (Seemann, 2014) against the non-redundant microbial RefSeq protein database. Gene orthology analysis was undertaken using Roary v3.11.2 (Page *et al.*, 2015), using a 90% amino acid identity cut-off. Roary output was processed using our custom R tool called Roary_stats found at https://github.com/IamIamI/pADAP_project/ . Homologs of the pADAP RepA gene were extracted from the GenBank database (Clark *et al.*, 2016), limited to ≥50% RepA amino acid identity, resulting in 23 homologs from various non-redundant bacterial species. A maximum likelihood phylogenetic tree of RepA (100 bootstrap replicates) was constructed using PhyML v3.1 (Guindon *et al.*, 2010) with nucleotide substitutions per site represented on the tree branches. Initial comparison of pADAP to the 50 sequenced plasmids was performed using

BLAST Ring Image Generator (BRIG) v0.95 (Alikhan *et al.*, 2011). Alignments of regions of interest were generated using Clustal Omega v1.2.4 (Larkin *et al.*, 2007). A phylogenetic tree summary for the backbone was generated using SumTree, part of the DendroPy v4.0.0 library (Sukumaran and Holder, 2010). Multiple sequence synteny was analyzed using progressiveMauve v20150226 (Darling *et al.*, 2010) and pairwise synteny using Easyfig v2.2.5 (Sullivan *et al.*, 2011). As distant plasmids were included (most notably WVU-005-1), branch lengths of the more closely related clusters to appear very short / non-existent (such as the "conserved" pADAP types). We therefore use rooted proportional cladogram to visualize the plasmid phylogeny. The unmodified tree can be viewed in Figure S.16.

Bioassay assessments

Pathogenicity was determined based on maximum-dose oral-challenge bioassays using 12 healthy field-collected third instar *C. giveni* larvae per assay as described previously (Grkovic *et al.*, 1995), with an average of three replicates for each assay depending on seasonal availability of larvae. Larvae treated with the pADAP+ *S. entomophila* A1MO2 strain and untreated larvae were used as positive and negative controls, respectively. Virulence assessment was based on the presence of amber disease symptoms in the larva at day 12 post challenge (Jackson and Saville, 2000) (Table S1). Additional assays were performed using larvae of the endemic Scarabaeidae *P. festiva* and *P. setosa* larvae, and exotic *A. couloni*, *A. tasmaniae* and *Odontria* spp.

Construction of marked plasmid variants

DNA manipulations and cloning were performed as described previously (Sambrook and Russell, 2001). For targeted mutagenesis, a two-step fusion PCR approach was performed, as described by Szewczyk et al. (2006). The resulting amplicon was then cloned into the suicide vector pJP5608 (Riedel *et al.*, 2013). The suicide vectors were conjugated into target isolates, and double-

recombinants were selected based on antibiotic screening and confirmed using sequencing. Primers, clones and double-recombinant mutants constructed using this method are listed in Table S2.

In vitro and *in vivo* STAMP stability assessment

Isolates bearing antibiotic-marked plasmid-mutants of *S. proteamaculans* pU143 and p145, and *S. entomophila* p210, p398, p626, and p1100 (Table S2), were cultured overnight at 30°C, shaking at 200 rpm in LB broth without antibiotics. Cultures were then passaged for 10 days using 1:100 dilutions of each overnight culture into fresh medium. Following the final incubation, dilutions of each culture were plated on LB agar and incubated overnight at 30°C. A total of 200 colonies per isolate were then patched onto LB agar plates supplemented with chloramphenicol (90 µg/ml), spectinomycin (160 µg/ml), or kanamycin (50 µg/ml) and assessed for loss of resistance to the antibiotics (Table S3A). Additionally, these isolates were independently assessed via bioassays using *C. giveni* larvae. At day 12, larvae were externally sterilized and macerated (Trought *et al.*, 1982). Dilutions of the macerate were plated on LB agar supplemented with appropriate antibiotics to determine *in vivo* plasmid retention and virulence cluster stability (Table S3A).

Assessment of STAMP stability under stressors

Isolates bearing a chloramphenicol resistance marked (Cm^R)- or GFPMut3-marked (GFP^+) (Cormack *et al.*, 1996) plasmid-mutant (Table S2), representative of the identified plasmid genotypes A–L (see main text), were cultured overnight at 30°C in triplicate in 3 ml LB broth, M9 + 0.4% casamino acids medium or LB broth + 0.2 µg/ml SOS activator Mitomycin C (MitC) (Janion, 2008), without antibiotics. Cultures were then passaged for 10 days using 1:100 dilutions. Dilutions of the final Cm^R cultures were plated on LB agar and 200 subsequent colonies were patched onto LB agar plates containing chloramphenicol (90 µg/ml) as a means to assess plasmid retention (Table S3B). Aliquots of the final GFP^+ cultures were diluted 1:10 in phosphate-buffered saline (MP Biomedical) and analyzed using a FACSCanto™ II cytometer (BD). Approximately 10 000 cells were gated using forward scatter (488/10

nm) and side scatter (488/10 nm), and GFP$^+$ cells were gated based on fluorescence (530/30 nm). The gates were set using overnight cultures of a wild-type *S. entomophila* isolate 626 (pADAP+) and GFP$^+$ *S. entomophila* isolate 626 (pADAP+) mutant as controls. Data were analyzed using FlowJo v10.6.1 (2019). Values presented in Table S3C are the mean values of three replicates. Raw data are provided in Data S1.

Data Availability

All processed sequencing data generated in this study have been submitted to the NCBI Reference Sequence database (RefSeq; https://www.ncbi.nlm.nih.gov/refseq/) under accession number MT039142–MT039228 (Table S1). R v3.5.3 (2019) scripts written with RStudio v1.1.463 (2015), including Roary_stats, used to generate figures and process data, are available at https://github.com/IamIamI/pADAP_project/

Acknowledgements

This work was supported by the New Zealand Tertiary Education Commission under the Centre for Research Excellence Programme. We thank Richard Townsend for help with *C. giveni* larvae collection and Amy Beattie and Renee Abigail Watson for help with gDNA preparation.

References

FlowJo. v10.6.1 ed. Oregon, United States of America: FlowJo, LLC; 2019.

R: A language and environment for statistical computing. v3.6.1 ed. Vienna, Austria: R Foundation for Statistical Computing; 2019.

RStudio: Integrated Development for R. v1.2.1335 build1379 ed. Boston, United States of America: RStudio, Inc; 2015.

Adeolu M, Alnajar S, Naushad S, R SG. 2016. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morganellaceae fam. nov., and Budviciaceae fam. nov. 66: 5575-5599.

Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12: 402.

Badel C, Erauso G, Gomez AL, Catchpole R, Gonnet M, Oberto J, et al. 2019. The global distribution and evolutionary history of the pT26-2 archaeal plasmid family. Environ Microbiol 21: 4685-4705.

Bagdasarian M, Bailone A, Bagdasarian MM, Manning PA, Lurz R, Timmis KN, et al. 1986. An inhibitor of SOS induction, specified by a plasmid locus in *Escherichia coli*. Proc Natl Acad Sci U S A 83: 5723-5726.

Barbier M, Damron FH. 2016. Rainbow vectors for broad-range bacterial fluorescence labeling. PLoS One 11: e0146827.

Bardaji L, Añorga M, Echeverría M, Ramos C, Murillo J. 2019. The toxic guardians — multiple toxin-antitoxin systems provide stability, avoid deletions and maintain virulence genes of Pseudomonas syringae virulence plasmids. Mobile DNA 10: 7.

Bergstrom CT, Lipsitch M, Levin BR. 2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. Genetics 155: 1505-1519.

Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, Lovett ST, et al. 1993. Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. Cell 75: 351-361.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579.

Bowen D, Rocheleau TA, Blackburn M, Andreev O, Golubeva E, Bhartia R, et al. 1998. Insecticidal toxins from the bacterium *Photorhabdus luminescens*. Science 280: 2129-2132.

Brar SK, Verma M, Tyagi RD, Valéro JR. 2006. Recent advances in downstream processing and formulations of *Bacillus thuringiensis* based biopesticides. Process Biochem 41: 323-342.

Brovedan M, Repizo GD, Marchiaro P, Viale AM, Limansky A. 2019. Characterization of the diverse plasmid pool harbored by the blaNDM-1-containing Acinetobacter bereziniae HPC229 clinical strain. PLOS ONE 14: e0220584.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Carroll AC, Wong A. 2018. Plasmid persistence: costs, benefits, and the plasmid paradox. Can J Microbiol 64: 293-304.

Castillo F, Benmohamed A, Szatmari G. 2017. Xer site specific recombination: double and single recombinase systems. Front Microbiol 8: 453-453.

Chang AC, Cohen SN. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol 134: 1141-1156.

Chattoraj DK. 2000. Control of plasmid DNA replication by iterons: no longer paradoxical. Mol Microbiol 37: 467-476.

Chen F, Guo YB, Wang JH, Li JY, Wang HM. 2007. Biological control of grape crown gall by *Rahnella aquatilis* HX2. Plant Dis 91: 957-963.

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2016. GenBank. Nucleic Acids Res 44: D67-D72.

Coil D, Jospin G, Darling AE. 2015. A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31: 587-589.

Colloms SD. 2013. The topology of plasmid-monomerizing Xer site-specific recombination. Biochem Soc Trans 41: 589.

Cormack BP, Valdivia RH, Falkow S. 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173: 33-38.

Cruz JW, Sharp JD, Hoffer ED, Maehigashi T, Vvedenskaya IO, Konkimalla A, et al. 2015. Growth-regulating *Mycobacterium tuberculosis* VapC-mt4 toxin is an isoacceptor-specific tRNase. Nat Commun 6: 7480-7480.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5: e11147.

Deep A, Kaundal S, Agarwal S, Singh R, Thakur KG. 2017. Crystal structure of *Mycobacterium tuberculosis* VapC20 toxin and its interactions with cognate antitoxin, VapB20, suggest a model for toxin–antitoxin assembly. The FEBS Journal 284: 4066-4082.

del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R. 1998. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev 62: 434-464.

Demeure CE, Dussurget O, Mas Fiol G, Le Guern A-S, Savin C, Pizarro-Cerdá J. 2019. Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. Genes Immun 20: 357-370.

Desfosses A, Venugopal HP, Joshi T, Felix J, Jessop M, Jeong H, et al. 2019. Atomic structures of an entire contractile injection system in both the extended and contracted states. Nature Microbiology.

Diaz-Orejas R, Espinosa M, Yeo CC. 2017. The importance of the expendable: toxin-antitoxin genes in plasmids and chromosomes. Front Microbiol 8: 1479.

Dimitriu T, Marchant L, Buckling A, Raymond B. 2019. Bacteria from natural populations transfer plasmids mostly towards their kin. Proc Biol Sci 286: 20191110.

Dionisio F, Conceição IC, Marques ACR, Fernandes L, Gordo I. 2005. The evolution of a conjugative plasmid and its ability to increase bacterial fitness. Biol Lett 1: 250-252.

Dodd SJ. Horizontal transfer of plasmidborne insecticidal toxin genes of *Serratia* species (doctoral thesis). Dunedin: University of Otago, Dunedin, New Zealand; 2003.

Dodd SJ, Hurst MRH, Glare TR, O'Callaghan M, Ronson CW. 2006. Occurrence of sep insecticidal toxin complex genes in *Serratia* spp. and *Yersinia frederiksenii*. Appl Environ Microbiol 72: 6584-6592.

Dolejska M, Papagiannitsis CC. 2018. Plasmid-mediated resistance is going wild. Plasmid 99: 99-111.

El Ghachi M, Bouhss A, Barreteau H, Touze T, Auger G, Blanot D, et al. 2006. Colicin M exerts its bacteriolytic effect via enzymatic degradation of undecaprenyl phosphate-linked peptidoglycan precursors. J Biol Chem 281: 22761-22772.

El Karkouri K, Pontarotti P, Raoult D, Fournier P-E. 2016. Origin and evolution of *Rickettsial* plasmids. PLoS One 11: e0147492-e0147492.

Elbing KL, Brent R. 2019. Recipes and tools for culture of *Escherichia coli*. Curr Protoc Mol Biol 125: e83.

Freigang J, Diederichs K, Schäfer KP, Welte W, Paul R. 2002. Crystal structure of oxidized flavodoxin, an essential protein in *Helicobacter pylori*. Protein Sci 11: 253-261.

Gatsogiannis C, Lang AE, Meusch D, Pfaumann V, Hofnagel O, Benz R, et al. 2013. A syringe-like injection mechanism in *Photorhabdus luminescens* toxins. Nature 495: 520-523.

Gerdes K, Møller-Jensen J, Jensen RB. 2000. Plasmid and chromosome partitioning: surprises from phylogeny. Mol Microbiol 37: 455-466.

Ghequire MG, Kemland L, De Mot R. 2017. Novel immunity proteins associated with colicin M-like bacteriocins exhibit promiscuous protection in *Pseudomonas*. Front Microbiol 8: 93.

Glare TR, Corbett GE, Sadler TJ. 1993. Association of a large plasmid with amber disease of the New Zealand grass grub, *Costelytra zealandica*, caused by *Serratia entomophila* and *Serratia proteamaculans*. J Invertebr Pathol 62: 165-170.

Glare TR, Hurst MRH, Grkovic S. 1996. Plasmid transfer among several members of the family Enterobacteriaceae increases the number of species capable of causing experimental amber disease in grass grub. FEMS Microbiol Lett 139: 117-120.

Grabowski B, Schmidt MA, Rüter C. 2017. Immunomodulatory Yersinia outer proteins (Yops)-useful tools for bacteria and humans alike. Virulence 8: 1124-1147.

Grimont PA, Grimont F, De Rosnay HL. 1977. Taxonomy of the genus Serratia. J Gen Microbiol 98: 39-66.

Grimont PAD, Jackson TA, Ageron E, Noonan MJ. 1988. *Serratia entomophila* sp. nov. associated with amber disease in the New Zealand grass grub *Costelytra zealandica*. Int J Syst Bacteriol 38: 1-6.

Grkovic S, Glare TR, Jackson TA, Corbett GE. 1995. Genes essential for amber disease in grass grubs are located on the large plasmid found in *Serratia entomophila* and *Serratia proteamaculans*. Appl Environ Microbiol 61: 2218-2223.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59: 307-321.

Guynet C, Cuevas A, Moncalián G, de la Cruz F. 2011. The stb operon balances the requirements for vegetative stability and conjugative transfer of plasmid R388. PLoS Genet 7: e1002073.

Harrison E, Brockhurst MA. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends in Microbiology 20: 262-267.

Ho PL, Lo WU, Yeung MK, Lin CH, Chow KH, Ang I, et al. 2011. Complete sequencing of pNDM-HK encoding NDM-1 carbapenemase from a multidrug-resistant *Escherichia coli* strain isolated in Hong Kong. PLoS One 6: e17989.

Hu B, Khara P, Christie PJ. 2019. Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*. Proc Natl Acad Sci U S A 116: 14222-14227.

Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16: 294.

Hurst MRH, Beattie A, Jones SA, Laugraud A, van Koten C, Harper L. 2018. Characterization of *Serratia proteamaculans* strain AGR96X encoding an anti-feeding prophage (tailocin) with activity against grass grub (*Costelytra giveni*) and manuka beetle (*Pyronota* spp.) larvae. Appl Environ Microbiol 84: 02739-02717.

Hurst MRH, Becher SA, O'Callaghan M. 2011. Nucleotide sequence of the *Serratia entomophila* plasmid pADAP and the *Serratia proteamaculans* pU143 plasmid virulence associated region. Plasmid 65: 32-41.

Hurst MRH, Glare TR. 2002. Restriction map of the Serratia entomophila plasmid pADAP carrying virulence factors for Costelytra zealandica. Plasmid 47: 51-60.

Hurst MRH, Glare TR, Jackson TA. 2004. Cloning Serratia entomophila Antifeeding Genes—a Putative Defective Prophage Active against the Grass Grub Costelytra zealandica. J Bacteriol 186: 5116-5128.

Hurst MRH, Glare TR, Jackson TA, Ronson CW. 2000. Plasmid-located pathogenicity determinants of *Serratia entomophila*, the causal agent of amber disease of grass grub, show similarity to the insecticidal toxins of *Photorhabdus luminescens*. J Bacteriol 182: 5127-5138.

Jackson TA, Huger AM, Glare TR. 1993. Pathology of amber disease in the New Zealand grass grub *Costelytra zealandica* (Coleoptera: Scarabaeidae). J Invertebr Pathol 61: 123-130.

Jackson TA, Saville DJ. Bioassays of replicating bacteria against soil-dwelling pests. In: Navon A, Ascher KRS, editors. Bioassays of entomopathogenic microbes and nematodes. Wallingford, Oxon, UK; New York, NY, USA: CABI Pub.; 2000. p. 73-94.

Janion C. 2008. Inducible SOS response system of DNA repair and mutagenesis in *Escherichia coli*. Int J Biol Sci 4: 338-344.

Johnson VW, Pearson J, Jackson TA. 2001. Formulation of *Serratia entomophila* for biological control of grass grub. N Z Plant Prot 54: 125-127.

Khanna A, Khanna M, Aggarwal A. 2013. *Serratia marcescens*- a rare opportunistic nosocomial pathogen and measures to limit its spread in hospitalized patients. J Clin Diagn Res 7: 243-246.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27: 722-736.

Krueger F. Trim Galore. v0.6.2 ed. Cambridge, England: Babraham Bioinformatics; 2019.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

Lin C, Jia X, Fang Y, Chen L, Zhang H, Lin R, et al. 2019. Enhanced production of prodigiosin by *Serratia marcescens* FZSF02 in the form of pigment pellets. Electron J Biotechnol 40: 58-64.

MacLean RC, San Millan A. 2015. Microbial Evolution: Towards Resolving the Plasmid Paradox. Current Biology 25: R764-R767.

Maróti G, Kondorosi É. 2014. Nitrogen-fixing *Rhizobium*-legume symbiosis: are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? Front Microbiol 5: 326.

Matelska D, Steczkiewicz K, Ginalski K. 2017. Comprehensive classification of the PIN domain-like superfamily. Nucleic Acids Res 45: 6995-7020.

Méric G, Mageiros L, Pascoe B, Woodcock DJ, Mourkas E, Lamble S, et al. 2018. Lineage-specific plasmid acquisition and the evolution of specialized pathogens in *Bacillus thuringiensis* and the *Bacillus cereus* group. Mol Ecol 27: 1524-1540.

Messer W. 2002. The bacterial replication initiator DnaA. DnaA and *oriC*, the bacterial mode to initiate DNA replication. FEMS Microbiol Rev 26: 355-374.

Molina-García L, Gasset-Rosa F, Moreno-Del Álamo M, Fernández-Tresguerres ME, Moreno-Díaz de la Espina S, Lurz R, et al. 2016. Functional amyloids as inhibitors of plasmid DNA replication. Sci Rep 6: 25425-25425.

Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. BMC Bioinformatics 13 Suppl 14: S8.

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

Novick RP. 1987. Plasmid incompatibility. Microbiol Rev 51: 381-395.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44: 733-745.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31: 3691-3693.

Palma L, Muñoz D, Berry C, Murillo J, Caballero P. 2014. *Bacillus thuringiensis* toxins: an overview of their biocidal activity. Toxins 6: 3296-3325.

Petrova V, Satyshur KA, George NP, McCaslin D, Cox MM, Keck JL. 2010. X-ray crystal structure of the bacterial conjugation factor PsiB, a negative regulator of RecA. J Biol Chem 285: 30615-30621.

Poudel S, Colman DR, Fixen KR, Ledbetter RN, Zheng Y, Pence N, et al. 2018. Electron transfer to nitrogenase in different genomic and metabolic backgrounds. J Bacteriol 200: e00757-00717.

Povolotskaya IS, Kondrashov FA, Ledda A, Vlasov PK. 2012. Stop codons in bacteria are not selectively equivalent. Biol Direct 7: 30-30.

Prentki P, Krisch HM. 1984. *In vitro* insertional mutagenesis with a selectable DNA fragment. Gene 29: 303-313.

Rangasamy K, Athiappan M, Devarajan N, Samykannu G, Parray JA, Aruljothi KN, et al. 2017. Pesticide degrading natural multidrug resistance bacterial flora. Microb Pathog 114: 304-310.

Raymond B, Bonsall MB. 2013. Cooperation and the evolutionary ecology of bacterial virulence: The *Bacillus cereus* group as a novel study system. BioEssays 35: 706-716.

Ridé M, Ridé S, Petit A, Bollet C, Dessaux Y, Gardan L. 2000. Characterization of plasmid-borne and chromosome-encoded traits of Agrobacterium biovar 1, 2, and 3 strains from France. Appl Environ Microbiol 66: 1818-1825.

Riedel T, Rohlfs M, Buchholz I, Wagner-Döbler I, Reck M. 2013. Complete sequence of the suicide vector pJP5603. Plasmid 69: 104-107.

Roy M, Bose M, Bankoti K, Kundu A, Dhara S, Das AK. 2020. Biochemical Characterization of VapC46 Toxin from *Mycobacterium tuberculosis*. Mol Biotechnol 62: 335-343.
Sambrook JF, Russell DW. Molecular cloning: A laboratory manual2001.
San Millan A, MacLean RC. 2017. Fitness costs of plasmids: a limit to plasmid transmission. Microbiol Spectr 5: MTBP-0016-2017.
Schumacher MA, Tonthat NK, Kwong SM, Chinnam NB, Liu MA, Skurray RA, et al. 2014. Mechanism of *staphylococcal* multiresistance plasmid replication origin assembly by the RepA protein. Proc Natl Acad Sci U S A 111: 9121-9126.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30: 2068-2069.
Sentchilo V, Mayer AP, Guy L, Miyazaki R, Green Tringe S, Barry K, et al. 2013. Community-wide plasmid gene mobilization and selection. ISME J 7: 1173-1186.
Stalder T, Press MO, Sullivan S, Liachko I, Top EM. 2019. Linking the resistome and plasmidome to the microbiome. ISME J 13: 2437-2446.
Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26: 1569-1571.
Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27: 1009-1010.
Szewczyk E, Nayak T, Oakley CE, Edgerton H, Xiong Y, Taheri-Talesh N, et al. 2006. Fusion PCR and gene targeting in *Aspergillus nidulans*. Nat Protoc 1: 3111-3120.
Thoma S, Schobert M. 2009. An improved *Escherichia coli* donor strain for diparental mating. FEMS Microbiol Lett 294: 127-132.
Townsend RJ, Ferguson CM, Proffitt JR, Slay MWA, Swaminathan J, Day S, et al. 2004. Establisment of *Serratia entomophila* after application of a new formulation for grass grub control. N Z Plant Prot 57: 310-313.
Townsend SM, Hurrell E, Caubilla-Barron J, Loc-Carrillo C, Forsythe SJ. 2008. Characterization of an extended-spectrum beta-lactamase *Enterobacter hormaechei* nosocomial outbreak, and other *Enterobacter hormaechei* misidentified as *Cronobacter* (*Enterobacter*) *sakazakii*. Microbiology 154: 3659-3667.
Trought TET, Jackson TA, French RA. 1982. Incidence and transmission of a disease of grass grub (*Costelytra zealandica*) in Canterbury. New Zeal J Exp Agr 10: 79-82.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9: e112963.
Wein T, Hülter NF, Mizrahi I, Dagan T. 2019. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. Nat Commun 10: 2595-2595.
Zwanzig M, Harrison E, Brockhurst MA, Hall JPJ, Berendonk TU, Berger U. 2019. Mobile compensatory mutations promote plasmid survival. mSystems 4: e00186-00118.