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

Thomas L. Sitter (1,2,*) Amy L. Vaughan (1,2) Marion Schoof (1,2) Simon A. Jackson (3) Travis R. Glare (1,2) Murray P. Cox (2,4) Peter C. Fineran (2,3) Paul P. Gardner (2,5) and Mark R. H. Hurst (1,2)

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.
5 Department of Biochemistry, University of Otago, Dunedin, New Zealand.

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 analyse 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 12 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.

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 (Marót 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 tumour-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 tūtāe ruru in Māori. The anti-insect properties of this bacterium are carried by a 153-kb amber disease-associated plasmid (pADAP) (Glace 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 result 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 (Sep) (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 orthologue 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 Sp (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 analysed 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 analysed 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.
Evolution of virulence in mega-plasmids

Fig. 1. Legend on next page.
Proteins of the autoregulating RepA family bind to 12–18-bp repeat motifs, termed iterons, positioned 5’ of repA (Chatteraj, 2000) and employ the chromosomal primosome for plasmid replication (Messer, 2002). Twenty copies of the 16-bp iteron repeat motif variants (5’-CArAAATNNNTCCaC-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.

**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 colocated with repA as well as a stability/partitioning locus containing staAB stability genes (Guyenet 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 *tra* 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 5339-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 phase 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/lamlml/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 and colleagues (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 Sentchilo and colleagues (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 (Fig. S11C) coverage with %GC in agreement with models by Povolotskaya and colleagues (2012). These results suggest a long coevolutionary relationship between STAMPs and the Serratia genus.

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, p1413, 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.

© 2021 Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology, 23, 5289–5304
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 antimicrobial 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::CmR 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-encoded STAMPs.
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 rufl and rufl (Fig. S10B). The rufl cluster encodes several putative fimbrial elements and two exotoxins, whereas rufl 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 rufl 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 rufl, rufl and the A and B components of Spp, designated p149Δrufl, p149Δrufl and p149ΔSppAB, respectively. Twelve days post-challenge, larvae fed p149Δrufl or p149Δrufl 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., _Acrasisidius 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 in 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 (N₂) to ammonia (NH₃), 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) harbours 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 sub-region was homologous to _Enterobacter hormaechei_ subsp. _hormaechei_ isolate Eh1 plasmid p1 (GenBank accession CP034755.1) (99% nt id). Both _S. marcescens_ and _E. hormaechei_ subsp. _hormaechei_ 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 < 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 homologues 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 homologues are largely confined to their own cluster, which mirrors the *S. entomophila* 16s cluster, in line with the findings of Dimitriu and colleagues (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 homologues.

**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 back-bone 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).

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* Afp 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. coulioni*, originating
Evolution of virulence in mega-plasmids

Fig. 3. The Serratia proteamaculans based STAMPs appear less associated with specific S. proteamaculans isolates as opposed to Serratia entomophila pADAP homologue STAMPs. 16S gene-based maximum likelihood cladogram tree tangled with the STAMP core backbone gene summary cladogram tree. Isolates labelled 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 [KA93181], 6. S. proteamaculans sp. proteamaculans [AJ23435], 7. S. proteamaculans sp. proteamaculans [AJ23434]. An untransformed maximum likelihood backbone gene summary tree can be observed in Fig. S16.

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.

**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 (Paci 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 (Nadalín 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). Circulator 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 10 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/. Homologues 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 homologues 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 analysed 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 Fig. S16.

**Bioassay assessments**

Pathogenicity was determined based on maximum-dose oral-challenge bioassays using 12 healthy field-collected third instar *C. giveti* 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 and colleagues (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 *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⁻¹), spectinomycin (160 μg ml⁻¹), or kanamycin (50 μg ml⁻¹) and assessed for loss of resistance to the antibiotics (Table S3A). Additionally, these

---

© 2021 Society for Applied Microbiology and John Wiley & Sons Ltd., *Environmental Microbiology*, 23, 5289–5304
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 (CmR)- 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⁻¹ SOS activator Mitomycin C (MitC) (Janion, 2008), without antibiotics. Cultures were then passaged for 10 days using 1:100 dilutions. Dilutions of the final CmR cultures were plated on LB agar and 200 subsequent colonies were patched onto LB agar plates containing chloramphenicol (90 μg ml⁻¹) 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 analysed 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 analysed using FlowJo v10.6.1 (2019). Values presented in Table S3C are the mean values of three replicates.

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.

Data availability statement
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/amlampl/pADAP_project/.

References
Evolution of virulence in mega-plasmids


Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site:

**Fig. S1.** Geographical distribution of sequenced isolates.

**Fig. S2.** Comparative analysis of RepA machinery.

**Fig. S3.** Coverage plot of the six samples sequenced using both the PacBio and the Illumina platform.

**Fig. S4.** The proposed start of the plasmid backbone shared by pADAP orthologues.

**Fig. S5.** Comparison of the *Serratia entomophila* strain A1MO2 pADAP encoded replication and stability region to distant homologues.

**Fig. S6.** Comparison of the *Serratia entomophila* strain A1MO2 pADAP encoded conjugation region to distant homologues.
Fig. S7. STAMP backbone divergence.

Fig. S8. A depiction of several notable insertions into the plasmid backbone sequences.

Fig. S9. The traG, trbC and trbB region that differentiates the S. entomophila pADAP plasmid backbone along with eight other plasmids, from the 33 S. proteamaculans and S. liquefaciens backbones.

Fig. S10. Several novel genetic regions identified.

Fig. S11. STAMP meta-data.

Fig. S12. HGT hotspot appears to garner higher rates of mutation.

Fig. S13. A visualization of the absence/presence matrix produced by Roary.

Fig. S14. pADAP imparts potential growth benefits to host bacterium.

Fig. S15. Illustration showing the process with which the Colicin M and VapBC TAs, found on AfpX encoding STAMPs.

Fig. S16. Supporting Information.

Table S1. Supporting Information.

Table S2. Supporting Information.

Table S3. Supporting Information.

Table S4. Supporting Information.

Table S5. Supporting Information.

Table S6. Supporting Information.

Table S7. Supporting Information.

Table S8. Supporting Information.

Table S9. Supporting Information.

Table S10. Supporting Information.

Table S11. Supporting Information.

Table S12. Supporting Information.

Table S13. Supporting Information.

Table S14. Supporting Information.

Table S15. Supporting Information.