CRISPR-Cas systems in Serratia

Maria Scrascia[1], Roberta Roberto[1], Pietro Daddabbo[1], Yosra Ahmed[2], Francesco Porcelli[1], Marta Oliva[1], Carla Calia[1], Angelo Marzella[1], and Carlo Pazzani[1]

[1]University of Bari
[2]Plant Pathology Research Institute

February 22, 2024

Abstract

The CRISPR-Cas system of Prokaryotes is an adaptive immune defense mechanism to protect themselves from invading genetic elements (e.g. phages and plasmids). Studies that describe the genetic organization of these prokaryotic systems have mainly reported on the Enterobacteriaceae family (now reorganized within the order Enterobacteriales). For some genera, data on CRISPR-Cas systems remain poor, as in the case of Serratia (now part of the Yersiniaceae family) where data are limited to a few genomes of the species marcescens. This study describes the detection, in silico, of CRISPR loci in 146 Serratia complete genomes and 336 high-quality assemblies available for the species ficaria, fonticola, grimesii, inhibens, liquefaciens, marcescens, nematodiphila, odorifera, oryzae, plymuthica, proteomaculans, quinivorans, rubidaea, symbiotic, and ureilytica. Apart from subtypes I-E and I-F1, which had previously been identified in marcescens, we report that of I-C and the variants I-ES1, I-ES2 and I-F1S1. Analysis of the genomic contexts for CRISPR loci revealed mdtN-phnP as the region mostly shared (grimesii, inhibens, marcescens, nematodiphila, plymuthica, rubidaea, and Serratia sp.). Three new contexts detected in genomes of rubidaea and fonticola (puu genes-mnmA) and rubidaea (osmE-soxG and ampC-yebZ) were also found. Plasmid and/or phage origin of spacers was also established.

CRISPR-Cas systems in *Serratia*

Maria Scrascia[1]*, Roberta Roberto[2], Pietro D'Addabbo[1], Yosra Ahmed[3], Francesco Porcelli[2], Marta Oliva[1], Carla Calia[1], Angelo Marzella[1] and Carlo Pazzani[1]

[1] Department of Biology, University of Bari Aldo Moro, 70124 Bari, Italy;

[2] Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, University of Bari Aldo Moro, 70126 Bari, Italy;

[3] Plant Quarantine Pathogens Laboratory, Mycology Research & Disease Survey, Plant Pathology Research Institute, ARC, Giza, Egypt;

***** Correspondence: Dr. Maria Scrascia, Department of Biology, Via E. Orabona, 4 - 70124 BARI

maria.scrascia@uniba.it; +39-080-5442065

Keywords: bacteria immune system; subtype I-C; subtype I-E; subtype I-F1; CRISPR system; insect symbiont; RPW; *Rhynchophorus ferrugineus*

LIST OF ABBREVIATIONS

CRISPR-Cas: Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins

CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats

HQAs: High-Quality Assemblies

NCBI: National Center for Biotechnology Information database

RPW: Red Palm Weevil

CDRs: Consensus of Direct Repeats

CGs: Complete Genomes

ABSTRACT

The CRISPR-Cas system of Prokaryotes is an adaptative immune defense mechanism to protect themselves from invading genetic elements (e.g. phages and plasmids). Studies that describe the genetic organization of these prokaryotic systems have mainly reported on the *Enterobacteriaceae* family (now reorganized within the order *Enterobacteriales*). For some genera, data on CRISPR-Cas systems remain poor, as in the case of*Serratia* (now part of the *Yersiniaceae* family) where data are limited to a few genomes of the species *marcescens* . This study describes the detection, *in silico* , of CRISPR loci in 146 *Serratia* complete genomes and 336 high-quality assemblies available for the species *ficaria* , *fonticola* ,*grimesii* , *inhibens, liquefaciens* , *marcescens* ,*nematodiphila* , *odorifera* , *oryzae, plymuthica* ,*proteomaculans* , *quinivorans* , *rubidaea* ,*symbiotica,* and *ureilytica* . Apart from subtypes I-E and I-F1 which had previously been identified in *marcescens* , we report that of I-C and the variants I-ES1, I-ES2 and I-F1S1. Analysis of the genomic contexts for CRISPR loci revealed *mdtN -phnP* as the region mostly shared (*grimesii* , *inhibens* ,*marcescens* , *nematodiphila* , *plymuthica* ,*rubidaea,* and *Serratia* sp.). Three new contexts detected in genomes of *rubidaea* and *fonticola* (*puu*genes-*mnmA*) and *rubidaea* (*osmE -soxG* and*ampC -yebZ*) were also found. Plasmid and/or phage origin of spacers was also established.

INTRODUCTION

The prokaryotic system CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated proteins) is a defense mechanism for bacteria and archaea against the invasion of bacteriophages and selfish genetic elements such as plasmids. Since their discovery around 15 years ago [1-3], CRISPR-Cas systems have been the object of many studies and functions, other than adaptative immunity, as regulation of bacteria virulence and stress response have been reported [4,5]. Based on a census of complete genomes (CGs), it is now reckoned that these systems are distributed mainly in archaea (˜82,5%) and, to a lesser extent, bacteria (˜40%) [6]. CRISPR-Cas systems are composed of CRISPR arrays and adjacent CRISPR-associated (*cas*) genes. The former is composed of direct repeats interspaced by spacers; the latter coding for proteins involved in the immune response and DNA repair. This ever-expanding knowledge of the composition and architecture of*cas* gene clusters has led to an updated classification of CRISPR-Cas systems where two classes, six types, and various subtypes (some of which are further divided into different variants) are now reported [6,7]. Class 1 includes the types I, III, and IV, which are divided into seven subtypes I (I-A to I-G), six subtypes III (III-A to III-F), and three subtypes IV (IV-A to IV-C), respectively. Class 2 includes the types II, V, and VI; they are also divided into subtypes: three subtypes II (II-A to II-C), eleven subtypes V (V-A to V-K and V-U), and four subtypes VI (VI-A to VI-D), respectively. While Class 2 is found mainly in Bacteria, Class 1 is present both in Bacteria and Archaea. Studies on CRISPR-Cas systems have been performed on genomes of different bacteria families, with that of the *Enterobacteriaceae*being one of the most investigated [8-10]. This family was unique in the *Enterobacterales* order until 2016 when Adeolu and colleagues [11] reclassified the order by adding six new families (*Budviciaceae, Erwiniaceae, Hafniaceae, Morganellaceae, Pectobacteriaceae, Yersiniaceae*). Despite this reclassification, data on CRISPR-Cas systems remain mainly limited to genera of the*Enterobacteriaceae* family [12-15].

The genus *Serratia* , a Gram-negative rod, is now part of the family *Yersiniaceae* . *Serratia* species can be found in different environments (e.g. water, soil) and hosts (e.g. humans, insects, plants, vertebrates) where they may play different roles ranging from opportunistic pathogens to symbionts [16-18]. Among*Serratia* species, *marcescens* is undoubtedly the most studied mainly for its role played as a symbiont associated with

insects and nematodes [19] or as a human opportunistic pathogen (currently reported as one of the most important bacteria responsible for acquired hospital infections) [20]. A growing number of *marcescens*genomes have then been sequenced with a pangenome allele database available for different studies ranging from virulence, and antibiotic resistance to the identification of CRISPR systems [21]. Several studies, additionally to *marcescens* , have also been reported for other *Serratia* species that play different roles in human and insect pathogenesis[22]. Although the CRISPR systems represent a valuable substrate for diagnostic, epidemiologic, and evolutionary analyses [4], data on CRISPR-Cas systems in the genus are scarce and limited to the detection of subtypes I-E and I-F1 in genomes of the species *marcescens* [9,23-25].

In this study, 146 *Serratia* CGs and 336 High-Quality Assemblies (HQAs) were available for the species *ficaria* , *fonticola* , *grimesii* ,*inhibens* , *liquefaciens* , *marcescens* ,*nematodiphila* , *odorifera* , *oryzae, plymuthica* ,*proteomaculans* , *quinivorans* , *rubidaea* ,*symbiotica,* and *ureilytica* were explored for the presence and type of *cas* gene clusters and/or CRISPR arrays. Apart from subtypes I-E and I-F1, the results presented in this study show the presence (first detected in *Serratia*) of subtype I-C and that of the variants I-ES1, I-ES2, and I-F1S1. Moreover, this study extends the previously reported *mdtN -phnP* CRISPR-genomic context, identified in *marcescens* , to the species *grimesii* ,*inhibens* , *nematodiphila* , *plymuthica* , and*rubidaea* , reporting three new possible shared contexts. One,*puu* genes-*mnmA* , was detected in genomes of *rubidaea*and *fonticola* , and two (*osmE -soxG* and*ampC -yebZ*) in genomes of *rubidaea* . Spacers' content was also assessed to establish the plasmid and/or phage origin of the matched protospacers. The discovery of CRISPR-Cas systems has allowed the development of new technology tools in the bioengineering field [26]. A clear example is represented by gene editing strategies based on CRISPR/Cas9 technique successfully used in agriculture, nutrition, and human health [27]. The development of new CRISPR-based applications also relies on the continuous update of CRISPR systems data and knowledge. Our study, in providing more comprehensive data on CRISPR in *Serratia* , has undoubtedly contributed to an expanded knowledge of these systems.

MATERIALS AND METHODS

Genomes analyzed

One hundred and forty-six *Serratia* CGs were considered in this study. The set of genomes encompasses the 15 *S. marcescens* CGs we previously analyzed [25] and those of the genus *Serratia*available at the CRISPR-Cas++ database (https://crisprcas.i2bc.paris-saclay.fr/MainDb/StrainList) up to 12/12/2020 [28,29] (Table S1). Among genome sequences available at the assembly level of scaffolds or contigs available at the National Center for Biotechnology Information database (NCBI) (https://www.ncbi.nlm.nih.gov/assembly) up to 12/12/2020, we selected the HQAs (N50>50kb). Species attribution and strain details (name, place, date of isolation) were recovered (when available) from GenBank or related articles. *Serratia* strains FGI94 (NC_020064), FS14 (NZ_CP005927), SCBI (NZ_CP003424), YD25 (NZ_CP016948), and DSM21420 (GCA_-000738675) were reclassified as reported by Sandner-Miranda *et al* ., 2018 [30]. We also included sequences with the accessions MK507743, MK507744, MK507745, and MK507746 referring to contigs (N50 ranging from 228817 and 291462) harboring CRISPR loci in genome assemblies (unpublished) of 4 *S. marcescens* strains reported as secondary symbionts in the Red Palm Weevil (RPW) *Rhynchophorus ferrugineus* (Olivier, 1790) (Coleoptera: Curculionidae) [25,31] (Supporting Table S1), an alien invasive pest now threatening South America [32].

Detection of CRISPR-Cas *loci.*

Details about the detection of the *cas* gene cluster and/or CRISPR array(s) for CGs were retrieved from the CRISPR-Cas++ database. CRISPR array(s) recorded by CRISPR-Cas++ were assigned to levels 1 to 4 based on criteria required to select the minimal structure of putative CRISPR as reported by Pourcel *et al* . [28]. Level 1 is the lowest level of confidence. Levels 2 to 4 were assigned based on the conservation of repeats (which must be high in a real CRISPR array) and on the similarity of spacers (it must be low). Level 4 CRISPRs were defined as the most reliable ones. Levels 1 to 3 may correspond to false CRISPRs. In our study, only CRISPR array(s) recorded with level 4, were considered. CRISPR arrays without a complete

set of *cas* genes in the host genome were defined as "orphans". Genomes harboring *cas* gene clusters were then submitted to the CRISPRone analysis suite (http://omics.informatics.indiana.edu/CRISPRone/) [33] to graphically visualize the architecture of each cluster. The same suite was used to search and visualize *cas* gene clusters in the HQAs. A subtype of *cas* gene clusters was assigned according to the recent classification update for CRISPR-Cas systems [6].

In silico **analyses of consensus of direct repeats.**

A consensus of Direct Repeats (CDRs) from CRISPR arrays was clustered by BLAST similarity. Some CDRs were manually trimmed when just a few terminal nucleotides were the only difference from the other members of the same cluster. The CDRs were used as input for CRISPRBank (http://crispr.otago.ac.nz/CRISPRBank/index.html) and CRISPR-Cas++ to assign, based on identity with known CDRs [28,29,34], a specific CDR type to CRISPR arrays. CRISPR arrays whose CDR type was consistent with the subtype of the *cas* gene set harbored in the same genome were defined as "canonical". While those not consistent with the subtype of the *cas* gene set harbored in the same genome were defined as "alien". A schematic diagram of an alien, canonical and orphan array is shown in Figure 1. CDRs and the number of repeats of the CRISPR arrays in the HQAs of *Serratia* sp strains DD3, Ag1, and Ag2 were recovered from the CRISPRone output. Spacers' analysis for duplications (spacers of Ag1, Ag2, and DD3 included) was performed through the CRISPRCasdb spacer database at the CRISPRCas++ site (https://crisprcas.i2bc.paris-saclay.fr/MainDbQry/Index). Phagic and/or plasmidic origin of matching protospacers were searched at the CRISPRTarget site (http://crispr.otago.ac.nz/CRISPRTarget/crispr_analysis.html) [34].

Genomic contexts of CRISPR positive genomes.

Analysis of CRISPR positive CGs and HQAs was performed to better characterize the genomic context surrounding the *cas* gene set(s) and/or CRISPR array(s). HQAs with at least 4kb flanking the *cas* gene set(s) were considered. These regions were annotated by Prokka (https://github.com/tseemann/prokka) [35]. Synteny was established by either the Mauve algorithm (http://darlinglab.org/mauve/mauve.html) [36] or visual inspection of annotated proteins.

Phylogenetic analyses.

The evolutionary relationship of *Serratia* strains found positive for *cas* genes set(s) was established and graphically depicted by the Cas3 sequence tree. All the protein sequences were aligned by the MUSCLE algorithm (https://www.ebi.ac.uk/Tools/msa/muscle/) [37,38]. The 16S rRNA gene tree was also drawn for comparison. Dendrograms were generated by the Neighbour-Joining clustering method and average distance trees with JalView (https://www.jalview.org/) [39]. For the 16S rRNA gene tree, the multiple sequence alignment was obtained by retrieving from 1 to 7 full gene sequences (CGs) or truncated 16S rRNA gene sequences (HQAs). A phylogenetic tree was obtained by multiple alignments of all retrieved 16S rRNA genes; an abbreviated tree was constructed by using one sequence from each genome.

RESULTS

CRISPR positive genomes.

A collection of 146 *Serratia* CGs was explored for the presence of *cas* gene cluster and/or CRISPR array(s). Most of the genomes (134) were reported as known species: *ficaria* (1),*fonticola* (7), *grimesii* (1), *inhibens* (1),*liquefaciens* (7), *marcescens* (87), *nematodiphila*(1), *plymuthica* (11), *proteomaculans* (2),*quinivorans* (2), *rubidaea* (8), *symbiotica* (4),*ureilytica* (2). The remaining 12 genomes were of unidentified species and, from here on, they will be referred to as *Serratia*sp. (Supporting Table S1). *cas* gene cluster and/or CRISPR array(s) were detected in 35 CGs (24%) of which 17 harbored a single *cas* gene cluster associated with one or more arrays, while 18 harbored orphan array(s). Some CGs records were assigned to the same genome being characterized by the same *cas* gene set subtype and identical numbers of both CRISPR arrays and spacers (Table 1). All detected *cas* gene clusters were of Class 1. Nine were canonical and distributed as follows: 2 subtypes I-C (*rubidaea*) (Figure 2A), 1 I-E (*plymuthica*), and 6 I-F1 (1 *fonticola* , 3*marcescens* , 1 *inhibens,* and 1 *rubidaea*) (Figures 2B and 2C). The remaining 8 clusters were found atypical and assigned, in this

study, to I-ES1 (3 *marcescens* and 1 *plymuthica*) and I-F1S1 (1 *marcescens* , 2 *rubidaea,* and 1 *Serratia*sp.) as variants of subtypes I-E and I-F1, respectively.

The variant I-ES1 had the *cas3-cas8e* genes spaced by ˜600 nt while the variant I-F1S1 had the*cas3* - *cas8f1* genes separated from each other by ˜400 nt (Figures 1B and 1C). Since the I-ES1 and I-F1S1*cas* gene clusters have never been reported in *Serratia* , their presence was further explored among 336 *Serratia* HQAs. The assemblies were distributed as follows: *ficaria* (1),*fonticola* (6), *grimesii* (2), *liquefaciens* (3),*marcescens* (295), *nematodiphila* (2), *odorifera* (2),*oryzae* (1), *plymuthica* (4), *proteomaculas* (1),*rubidaea* (2), *symbiotica* (1), *ureilytica* (1) and*Serratia* sp (15) (Supporting Table S1). Of the 336 analyzed genomes, 46 (13.7%) were positive for the presence of *cas* gene clusters. Twenty-six were subtype I-F1 (21 *marcescens* , 1*fonticola,* and 4 *Serratia* sp.) (Figure 1C), 2 subtype I-C (*rubidaea*) (Figure 1A), and 3 subtype I-E (*marcescens*) (Figure 2B) (Supporting Table S2). The variant I-ES1 was detected in 2 genomes of *marcescens* , the I-F1S1 in 8 genomes of*marcescens,* and 1 of *grimesii* . In 3 genomes of*Serratia* sp. (strains Ag1, Ag2, and DD3) an additional variant of the subtype I-E, here named I-ES2, was detected (Figure 2B). The variant I-ES2 was characterized by the translocation of *cas6e* between*cas7* and *cas11* , and the presence (upstream of *cas3*) of a gene harboring the WYL domain and encoding for a potential functional partner of the CARF (CRISPR–Cas Associated Rossmann Fold) superfamily proteins [6]. Proteins containing the WYL domain (name standing for the three conserved amino acids tryptophan, tyrosine, and leucine, respectively) have been reported for subtypes I-D and VI-D [40,41]. The distribution of CRISPR-positive genomes, over the total analyzed among *Serratia* species, is shown in Figure 3. Coexistence in the same genome of different sets of *cas* genes was also detected: subtypes I-E and I-F1 were found in the single HQA of*oryzae* , while subtypes I-ES2 and I-F1were detected in 2 HQAs of*Serratia* sp (strains Ag1 and Ag2) (Supporting Table S2).

CDRs and spacers.

The 35 CRISPR-positive CGs harbored 78 CRISPR arrays of which 48 were canonical. The latter were distributed as follows: *fonticola* (4),*inhibens* (1), *marcescens* (19), *plymuthica* (5),*rubidaea* (15), and *Serratia* sp (4). Twenty-three arrays were orphan and detected in genomes of *marcescens* (8),*plymuthica* (4), *symbiotica* (1), *nematodiphila* (1),*rubidaea* (5), and *Serratia* sp (4) (Table 1 and Figure 1). Alien arrays (8) were only detected in the species *rubidaea* . For a comprehensive analysis, arrays in the 3 HQAs Ag1, Ag2, and DD3 were included (Supporting Table S2). All disclosed CRISPR arrays were assigned, by comparative sequence analyses, to CDR types I-C, I-E, or I-F (Table 1). The association between CDR types and *cas* gene sets (canonical and variant) is reported in Table 2. Based on their nucleotide identity, the CDRs identified for subtype I-E and variants I-ES1 and I-ES2 could be arranged into two clusters named CDR-I and CDR-II. CDR-I was composed of 6 CDRs (identity from 83 to 96%) and linked to the *cas* gene sets I-E and I-ES1. CDR-II was composed of 2 CDRs (identity of about 96%) and linked to the *cas* gene set I-ES2. When the CDRs of the two clusters were compared to each other, the nucleotide identity dropped to 55-62%.

The architecture of the *cas* gene set I-ES2 has previously been reported as I-E* for *Klebsiella* and I-E variant for *Vibrio cholerae* [14,42]. We then compared the CDRs sequences I-E* and I-E variant with those of CDR-II and the identity was found between 82 to 96%. This association has further been confirmed by results obtained from the analysis of the *cas* gene clusters identified in 99 genomes retrieved from CRISPRBank and by searching for the presence of CDRs I-ES2. Results showed that 95 of these genomes had a *cas*gene architecture identical to that of I-ES2. The remaining 4 genomes harbored a truncated set of *cas* genes. The overall of these data linked specifically CDR-II to the *cas* gene set I-ES2.

A total of 1391 spacers were identified. Identical arrays were shared by*rubidaea* strains FDAARGOS_926 and NCTC12971. Likewise, different sets of identical arrays were shared by *plymuthica* strains AS9, AS12, and AS13; *marcescens* strains KS10 and EL1;*marcescens* strains CAV1761 and CAV1492 (Supporting Table S3). These findings confirmed multiple records of the same genome for each group of strains and the total number of spacers was estimated at 1290 of which 1219 were unique and 330 matched protospacers with the following origin: 131 phage, 132 plasmid, and 67 phage/plasmid (Supporting Table S3).

Phylogenetic trees.

The phylogenetic tree generated by multiple alignments of the amino acid sequences of Cas3 showed a clusterization of the subtypes I-C, I-E, and I-F1 into 3 distinct branches (Figure 4). The variants I-ES1 and I-F1S1 were randomly distributed among the I-E and I-F1 respectively, while the variant I-ES2 appears to group within a sub-lineage of I-E. Within the I-C, I-E, and I-F1 branches, strains belong to the same group of species. A phylogenetic tree based on multiple alignments of the 16S rRNA gene sequences was generated for comparison (Figure 5 and Supporting Figure S1). The 16S rRNA gene trees showed, as expected, nesting of the strains belonging to the same species. The phylogenetic distribution of *Serratia* species in the Cas3 tree may suggest a possible independent intra-species evolutionary pathway. However, being that the number of available CRISPR-positive genomes is too low for most*Serratia* species such a hypothesis needs to be validated by future studies. The position of strains TEL in the cluster*marcescens* and JUb9 in the cluster *rubidaea* shown in the Cas3 phylogenetic tree was confirmed by the 16S rRNA gene tree, which might suggest a species assignment for these strains.

CRISPR genomic contexts.

The 35 CRISPR-positive CGs and 28 of the 46 CRISPR-positive HQAs were analyzed to identify the possible shared genomic context(s). Eight different genomic contexts, named from A to H, were identified. Contexts A to D (Figure 6) were shared by different genomes, while those from E to H were identified in single genomes. The genomic context A (*mdtN -phnP*) has previously been described in *S .marcescens* strains isolated as a secondary symbiont of RPW and in other *marcescens* CGs available in the NCBI database [25] becoming the most commonly shared in this study being identified in 55 genomes distributed as follows: 35 *marcescens* , 1 *grimesii* , 1 *inhibens* , 1 *nematodiphila* , 6 *plymuthica,* 6*rubidaea,* and 5 *Serratia* sp. Contexts B (*puu*genes-*mnmA*), C (*osmE -soxG*), and D (*ampC -yebZ*) were shared by 11, 4, and 6 genomes respectively; context B by genomes of species *fonticola* (2),*rubidaea* (7) and *Serratia* sp. (2); C and D only by*rubidaea* genomes. For context D, assignment to *rubidaea*was assumed for the strain JUb9 (see above). The contexts E (*nrdG -bglH*) and F (*sucD -vasK*) were both identified in the single genome of *S . oryzae* strain J11-6; while G (*gntR -cda*) and H (*gutQ -queA*) in genomes of the *Serratia* sp. Ag1 and *S . symbiotica*CWBI-2.3, respectively (Table 3). Distribution of the genomic contexts by subtypes of *cas* gene sets and/or CDR types is reported in the supporting Table S4. Genomes of species *rubidaea* were characterized by the presence of multiple CRISPR contexts (A, B, C, D) with the context C associated with the *cas* gene set of subtype I-C.

DISCUSSION

Bacteria of the genus *Serratia* are ubiquitous and have been isolated from soil, water, plant roots, insects, and the gastrointestinal tract of animals [16-18]. This broad range of environments exposes *Serratia* strains to exogenous genetic elements such as plasmids, phages, and chromosomal fragments of other bacteria. Some of them may represent a life threat (e.g. phages) or a metabolic burden (e.g. plasmids). To overcome this, defense mechanisms such as CRISPR-Cas have been developed during bacterial evolution. Studying the presence/absence of CRISPR-Cas systems and their features in different genera of families is a relatively new scientific approach of investigation to gain data on the evolution of these systems and their role played during bacteria lifetime [43]. The average percentage of CRISPR distribution among Bacteria is the outcome of processes and/or factors that play different ecological roles within a genus/species. Among these processes/factors noteworthy is the balance between protection provided by CRISPR systems and their possible deleterious effects (e.g. self-targeting spacers), the role played by exogenous genetic elements (e.g. plasmids, phages, etc.) in bacteria evolution, and the horizontal transfer of CRISPR systems.

Data on CRISPR loci in *Serratia* are limited to CGs of *S .marcescens* strains [9,23-25]. In the present study, along with the species *marcescens* , we extended data on CRISPR loci to 14 additional *Serratia* species. CRISPRs were detected in 24% of the CGs and about 14% of the HQAs analyzed. The percentage of detection is lower than that reported for Bacteria (about 40%) [6]. However, whether the lower percentage of detection in *Serratia* reflects a distinguishing feature of the genus (particularly for the most representative analyzed *marcescens* species where the percentage was 12.6%) or a misrepresentative distribution of the

available genomes in databases, remains to be established.

Most of the loci identified in this study were located within the genomic context *mdtN -phnP* previously reported in the species *marcescens* and now further extended to those of*grimesii* , *inhibens* , *nematodiphila* ,*plymuthica* , and *rubidaea* . Three new possible contexts were also identified: one (*puu* genes-*mnmA*) shared by genomes of*rubidaea* and *fonticola* ; and two (*osmE -soxG*and *ampC -yebZ*) detected in those of *rubidaea* . The context *osmE – soxG* might be closely linked to the*cas* gene set of subtype I-C (Supporting Table S4). Due to the low number of CRISPR-positive genomes of *rubidaea* and*fonticola* and genomes positive for the *cas* gene set I-C, further analyses are required to confirm this hypothesis.

A previous comprehensive study on the distribution of CRISPR-Cas systems in genomes of the *Enterobacteriaceae* family (now reorganized within the *Enterobacteriales* order) showed the predominant presence of subtype I-E and the rare coexistence of subtypes I-E and I-F1 in the same genome [9]. Our data show the prevalence of subtype I-F1 (39,5%), followed by subtypes I-E (about 5%) and I-C (about 5%). Detection of subtype I-C is, to the best of our knowledge, the first report in *Serratia* . The prevalence of the subtype I-F1 in our subset of CRISPR-positive genomes is consistent with both the new reorganized *Enterobacteriales* order [11] and data produced by Medina-Aparicio *et al* . [9]. Indeed, in the aforementioned study subtype I-F1 was found prevalent in genera *Yersinia* ,*Rahnella,* and *Serratia* which are now part of the new *Yersiniaceae* family. On the other hand, the subtype I-E remains predominant within the *Enterobacteriaceae* family. Moreover, the finding of two distinct *cas* -gene sets (I-E/I-F1 or I-ES2/I-F1) in only 3 *Serratia* genomes, confirms that the coexistence of these subtypes is not frequent.

Six different *cas* -gene set architectures were identified of which those reported as I-ES1 (characterized by a 0.6kb*cas3 /cas8e* intergenic sequence), I-ES2 (characterized by the *cas6e* translocation between *cas7* and *cas11*), and I-F1S1 (characterized by 0.4kb *cas3 /cas8f1* intergenic sequence) are, to the best of our knowledge, the first ever detected in*Serratia* . Similar or identical architectures of I-ES1, I-ES2, and I-F1S1 have been reported for other bacteria genera: a similar architecture to I-ES1 has been described in *Escherichia coli*(IGLB fragment) where the *cas3 /cas8e* intergenic sequence was ˜0,4kb [44,45]; an identical architecture of I-ES2 has already been detected in *Klebsiella* (I-E*) and *Vibrio* (I-E variant) strains [14,42]; a similar architecture to I-F1S1 was reported in *V. cholerae* (I-FV1), where the*cas3 /cas8f1* intergenic sequence was ˜0.1kb [42].

This study also supplies data on the presence/number of CRISPR arrays and their CDRs sequences in *Serratia* . Apart from canonical arrays (61.5% of the total disclosed arrays), orphans (29.4%) and aliens (10.2%) arrays were also detected (Table 1 and Figure 1). Orphan arrays might represent remnants of previous complete CRISPR-Cas systems [33]. The presence of alien arrays found only in *rubidaea* CGs is, as far as we know, the first report in bacteria CRISPR-positive genomes. Its detection might be explained as traces of ancient complete CRISPR-Cas systems I-E/I-F1 or I-C/I-E/I-F1 coexistent within the same genome (Table 1). Alternatively, the aliens might result from single horizontal gene transfer events. Further analyses could unveil their genetic origin and the entity of their distribution among CRISPR-positive bacteria genomes. Detection of more alien arrays might unveil that the presence of multiple subtypes in a genome is more frequent than has been reported so far. Furthermore, CDRs specifically associated with the *cas* gene set variant I-ES2 were also first described (Table 2).

Finally, the phylogenetic tree generated by multiple alignments of the Cas3 sequences showed a potential sub-lineage (variant I-ES2) within the I-E branch and thus might represent and/or anticipate a distinct clonal expansion of an I-E sub-population (Figure 4).

Knowledge of CRISPR-*Cas* systems is constantly expanding due to studies on newly available genomic sequences or genomic sequences not yet explored. The CRISPR-Cas systems classification is thus continuously updating also in the light of their possible applications. Indeed, the CRISPR-Cas technology has undoubtedly revolutionized systems of genome editing with a wide range of potential industrial and biomedical applications. Other, more recent genome-editing tools are based on methods that make use of the Cas9 protein [46]. However, the expression of foreign proteins with DNA-binding and editing activity appears toxic for

many bacteria. Harness of endogenous CRISPR systems is a recent and promising new line of approach for bacteria genome editing [47,48].

Our study has contributed to expanding knowledge on the variability and distribution of CRISPR systems in the *Serratia* genus. Data here presented might be exploitable for native CRISPR effectors of this genus that includes species (e.g. *marcescens*) relevant in environmental and clinical fields. Moreover, detection of the same subtype of*cas* -gene sets in different *Serratia* species and other genera highlights the open question on the molecular mechanism(s) yet to be identified that have been allowed intra- and inter-species spread.

ACKNOWLEDGMENTS

We would like to thank Julian Laurence for his writing assistance.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interest.

DATA AVAILABILITY STATEMENT

The authors declare that all data supporting the findings of this study are available within the article and its Supporting Information files.

ORCID

Scrascia Maria https://orcid.org/0000-0003-3351-5273

REFERENCES

[1] Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology (Reading). 2005;151:2551-2561.

[2] Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. Journal of molecular evolution. 2005;60:174-182.

[3] Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biology direct. 2006;1:7.

[4] Louwen R, Staals RH, Endtz HP, van Baarlen P, van der Oost J. The role of CRISPR-Cas systems in virulence of pathogenic bacteria. Microbiol Mol Biol Rev. 2014;78:74-88.

[5] Faure G, Makarova KS, Koonin EV. CRISPR-Cas: Complex Functional Networks and Multiple Roles beyond Adaptive Immunity. J Mol Biol. 2019;431:3-20.

[6] Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol. 2020;18:67-83.

[7] Koonin EV, Makarova KS. Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. Genome Biol Evol. 2017;9:2812-2825.

[8] Shariat N, Dudley EG. CRISPRs: molecular signatures used for pathogen subtyping. Appl Environ Microbiol. 2014;80:430-439.

[9] Medina-Aparicio L, Davila S, Rebollar-Flores JE, Calva E, Hernandez-Lucas I. The CRISPR-Cas system in Enterobacteriaceae. Pathogens and disease. 2018;76:1-15.

[10] Xue C, Sashital DG. Mechanisms of Type I-E and I-F CRISPR-Cas Systems in Enterobacteriaceae. EcoSal Plus. 2019;8:1-38.

[11] Adeolu M, Alnajar S, Naushad S, R SG. 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. Int J Syst Evol Microbiol. 2016;66:5575-5599.

[12] Diez-Villasenor C, Almendros C, Garcia-Martinez J, Mojica FJ. Diversity of CRISPR loci in Escherichia coli. Microbiology. 2010;156:1351-1361.

[13] Shariat N, Timme RE, Pettengill JB, Barrangou R, Dudley EG. Characterization and evolution of Salmonella CRISPR-Cas systems. Microbiology (Reading). 2015;161:374-386.

[14] Shen J, Lv L, Wang X, Xiu Z, Chen G. Comparative analysis of CRISPR-Cas systems in Klebsiella genomes. J Basic Microbiol. 2017;57:325-336.

[15] Wang P, Zhang B, Duan G, Wang Y, Hong L, Wang L, et al. Bioinformatics analyses of Shigella CRISPR structure and spacer classification. World J Microbiol Biotechnol. 2016;32:38.

[16] Gupta V, Sharma S, Pal K, Goyal P, Agarwal D, Chander J. Serratia no longer an opportunistic uncommon pathogen - case series & review of literature. Infectious disorders drug targets. 2021;21:e300821191666

[17] Cristina ML, Sartini M, Spagnolo AM. Serratia marcescens Infections in Neonatal Intensive Care Units (NICUs). Int J Environ Res Public Health. 2019;16:1-10.

[18] Lo WS, Huang YY, Kuo CH. Winding paths to simplicity: genome evolution in facultative insect symbionts. FEMS Microbiol Rev. 2016;40:855-874.

[19] Chen S, Blom J, Walker ED. Genomic, Physiologic, and Symbiotic Characterization of Serratia marcescens Strains Isolated from the Mosquito Anopheles stephensi. Front Microbiol. 2017;8:1483.

[20] Ferreira RL, Rezende GS, Damas MSF, Oliveira-Silva M, Pitondo-Silva A, Brito MCA, et al. Characterization of KPC-Producing Serratia marcescens in an Intensive Care Unit of a Brazilian Tertiary Hospital. Front Microbiol. 2020;11:956.

[21] Abreo E, Altier N. Pangenome of Serratia marcescens strains from nosocomial and environmental origins reveals different populations and the links between them. Scientific reports. 2019;9:46.

[22] Petersen LM, Tisa LS. Friend or foe? A review of the mechanisms that drive Serratia towards diverse lifestyles. Can J Microbiol. 2013;59:627-640.

[23] Vicente CS, Nascimento FX, Barbosa P, Ke HM, Tsai IJ, Hirao T, et al. Evidence for an Opportunistic and Endophytic Lifestyle of the Bursaphelenchus xylophilus-Associated Bacteria Serratia marcescens PWN146 Isolated from Wilting Pinus pinaster. Microb Ecol. 2016;72:669-681.

[24] Srinivasan VB, Rajamohan G. Genome analysis of urease positive Serratia marcescens, co-producing SRT-2 and AAC(6')-Ic with multidrug efflux pumps for antimicrobial resistance. Genomics. 2019;111:653-660.

[25] Scrascia M, D'Addabbo P, Roberto R, Porcelli F, Oliva M, Calia C, et al. Characterization of CRISPR-Cas Systems in Serratia marcescens Isolated from Rhynchophorus ferrugineus (Olivier, 1790) (Coleoptera: Curculionidae). Microorganisms. 2019;7:1-9.

[26] Dong H, Cui Y, Zhang D. CRISPR/Cas Technologies and Their Applications in Escherichia coli. Frontiers in bioengineering and biotechnology. 2021;9:762676.

[27] Nidhi S, Anand U, Oleksak P, Tripathi P, Lal JA, Thomas G, et al. Novel CRISPR-Cas Systems: An Updated Review of the Current Achievements, Applications, and Future Research Perspectives. International journal of molecular sciences. 2021;22:1-42.

[28] Pourcel C, Touchon M, Villeriot N, Vernadet JP, Couvin D, Toffano-Nioche C, et al. CRISPRCasdb a successor of CRISPRdb containing CRISPR arrays and cas genes from complete genome sequences, and tools to download and query lists of repeats and spacers. Nucleic Acids Res. 2020;48:D535-D544.

[29] Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Neron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246-W251.

[30] Sandner-Miranda L, Vinuesa P, Cravioto A, Morales-Espinosa R. The Genomic Basis of Intrinsic and Acquired Antibiotic Resistance in the Genus Serratia. Front Microbiol. 2018;9:828.

[31] Scrascia M, Pazzani C, Valentini F, Oliva M, Russo V, D'Addabbo P, et al. Identification of pigmented Serratia marcescens symbiotically associated with Rhynchophorus ferrugineus Olivier (Coleoptera: Curculionidae). MicrobiologyOpen. 2016;5:883-890.

[32] Dalbon VA, Acevedo JPM, Ribeiro Junior KAL, Ribeiro TFL, da Silva JM, Fonseca HG, et al. Perspectives for Synergic Blends of Attractive Sources in South American Palm Weevil Mass Trapping: Waiting for the Red Palm Weevil Brazil Invasion. Insects. 2021;12:1-16.

[33] Zhang Q, Ye Y. Not all predicted CRISPR-Cas systems are equal: isolated cas genes and classes of CRISPR like elements. BMC Bioinformatics. 2017;18:92.

[34] Biswas A, Staals RH, Morales SE, Fineran PC, Brown CM. CRISPRDetect: A flexible algorithm to define CRISPR arrays. BMC Genomics. 2016;17:356.

[35] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-2069.

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

[37] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.

[38] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792-1797.

[39] Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189-1191.

[40] Makarova KS, Anantharaman V, Grishin NV, Koonin EV, Aravind L. CARF and WYL domains: ligand-binding regulators of prokaryotic defense systems. Frontiers in genetics. 2014;5:102.

[41] Makarova KS, Gao L, Zhang F, Koonin EV. Unexpected connections between type VI-B CRISPR-Cas systems, bacterial natural competence, ubiquitin signaling network and DNA modification through a distinct family of membrane proteins. FEMS Microbiol Lett. 2019;366:fnz088.

[42] McDonald ND, Regmi A, Morreale DP, Borowski JD, Boyd EF. CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species. BMC Genomics. 2019;20:105.

[43] Butiuc-Keul A, Farkas A, Carpa R, Iordache D. CRISPR-Cas System: The Powerful Modulator of Accessory Genomes in Prokaryotes. Microbial physiology. 2022;32:2-17.

[44] Westra ER, Pul U, Heidrich N, Jore MM, Lundgren M, Stratmann T, et al. H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. Mol Microbiol. 2010;77:1380-1393.

[45] Pul U, Wurm R, Arslan Z, Geissen R, Hofmann N, Wagner R. Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. Mol Microbiol. 2010;75:1495-1512.

[46] Arroyo-Olarte RD, Bravo Rodriguez R, Morales-Rios E. Genome Editing in Bacteria: CRISPR-Cas and Beyond. Microorganisms. 2021;9:1-25.

[47] Strecker J, Ladha A, Gardner Z, Schmid-Burgk JL, Makarova KS, Koonin EV, et al. RNA-guided DNA insertion with CRISPR-associated transposases. Science. 2019;365:48-53.

[48] Klompe SE, Vo PLH, Halpin-Healy TS, Sternberg SH. Transposon-encoded CRISPR-Cas systems direct RNA-guided DNA integration. Nature. 2019;571:219-225.

Table 1. *Cas* genes clusters and CRISPR array(s) in complete genomes

Subtype of *cas* cluster	**CRISPR array(s) CDR type**	**CRISPR array(s) Category**	**CRISPR array(s) #Arrays (#spacers)**	***Serratia* species**	**Strain**	**Source**	**Place of isolation**	**Year of isolation**	**Ac As bly**
I-C	I-C	canonical	1 (14)	*rubidaea*	FDAARGOS_926[a]	n/a	n/a	n/a	NZ CP
	I-E	alien	1 (7)						
	I-F	alien	2 (2, 5)						
I-C	I-C	canonical	1 (14)	*rubidaea*	NCTC12971[a]	n/a	n/a	n/a	LR
	I-E	alien	1 (7)						
	I-F	alien	2 (2, 5)						
I-E	I-E	canonical	2 (43, 30)	*plymuthica*	NCTC8900	n/a	n/a	n/a	LR
I-ES1	I-E	canonical	4 (6, 8, 27, 44)	*marcescens*	E28	Hospital Ensuite	Australia	2012	CP
I-ES1	I-E	canonical	3 (7, 10, 22)	*marcescens*	SER00094	clinical	USA	2017	CP
I-ES1	I-E	canonical	3 (11, 39, 69)	*marcescens*	MSB1_9C-sc-2280320	n/a	n/a	n/a	LR
I-ES1	I-E	canonical	2 (35, 47)	*plymuthica*	NCTC8015	Canal water	n/a	n/a	LR
I-F1	I-F	canonical	2 (25, 27)	*marcescens*	12TM	pharyngeal secretions	Romania	2014	CM
I-F1	I-F	canonical	2 (8, 17)	*marcescens*	N4-5	soil	USA	1995	CP
I-F1	I-F	canonical	2 (6, 45)	*marcescens*	PWN146	*Bursaphelenchus xylophilus*	Portugal	2010	LT
I-F1	I-F	canonical	3 (11, 13, 42)	*fonticola*	DSM 4576	water	n/a	1979	NZ CP
I-F1	I-F	canonical	2 (15, 24)	*inhibens*	PRI-2c	maize rhizosphere soil	Netherlands	2004	NZ CP
I-F1	I-F	canonical	6 (1, 3, 7, 7, 14, 14)	*rubidaea*	FDAARGOS_880	n/a	n/a	n/a	CP
I-F1S1	I-F	canonical	3 (5, 10, 29)	*marcescens*	FZSF02	soil	China	2014	CP
I-F1S1	I-E	alien	1 (9)	*rubidaea*	FGI94	*Atta colombica*	Panama	2009	NC 020
	I-F	canonical	3 (6, 15, 16)						CP

I-F1S1	I-F	canonical	4 (3, 6, 7, 8)	*rubidaea*	NCTC10036	finger	n/a	n/a	LR
	I-E	alien	1 (3)						
I-F1S1	I-F	canonical	4 (2, 2, 7, 7, 10)	*Serratia* sp.	JUb9	compost	France	2019	CP
n/a	I-F	orphan	1 (21)	*marcescens*	SCQ1	blood from silkworm	China	2009	CP
n/a	I-F	orphan	1 (3)	*marcescens*	AR_-0130	n/a	n/a	n/a	CP
n/a	I-F	orphan	1 (6)	*plymuthica*	AS9[b]	plant	Sweden	n/a	NC 015 CP
n/a	I-F	orphan	1 (6)	*plymuthica*	AS12[b]	plant	Sweden	1998	NC 015 CP
n/a	I-F	orphan	1 (6)	*plymuthica*	AS13[b]	plant	Sweden	n/a	NC 017 CP
n/a	I-F	orphan	1 (3)	*marcescens*	B3R3	*Zea mays*	China	2011	NZ CP
n/a	I-F	orphan	2 (1, 2)	*Serratia* sp.	MYb239	compost	Germany	n/a	CP
n/a	I-F	orphan	1 (3)	*Serratia* sp.	SSNIH1	n/a	USA	2015	CP
n/a	I-F	orphan	1 (3)	*nematodiphila*	DH-S01	n/a	n/a	n/a	CP
n/a	I-F	orphan	2 (4, 6)	*rubidaea*	NCTC9419	n/a	n/a	n/a	LR
n/a	I-F	orphan	2 (6, 2)	*rubidaea*	NCTC10848	n/a	n/a	n/a	LS
n/a	I-E	orphan	1 (3)						
n/a	I-E	orphan	1 (26)	*marcescens*	KS10[c]	marine	USA	2006	CP
n/a	I-E	orphan	1 (26)	*marcescens*	EL1[c]	marine	USA	2002	CP
n/a	I-E	orphan	2 (3, 32)	*marcescens*	CAV1761[d]	Peri-rectal	Virginia	2014	CP
n/a	I-E	orphan	2 (3, 32)	*marcescens*	CAV1492[d]	clinical	USA	2011-12	NZ CP
n/a	I-E	orphan	1 (2)	*Serratia* sp.	KUDC3025	rhizospheric soil	South Korea	2017	CP
n/a	I-F	orphan	1 (2)	*plymuthica*	V4	milk pro-cessing plant	Portugal	2006	CP
n/a	I-C	orphan	1 (8)	*symbiotica*	CWBI-2.3	*Aphis fabae* (type strain of S. symbiotica)	Belgium	2009	CP

[a] Possible multiple records of the same genome

[b] Possible multiple records of the same genome

[c] Possible multiple records of the same genome

Table 2. Association between CDRs and *cas* gene set(s).

CDR sequence (5'-3')	# nt	Record in CRISPRBank and CRISPR-Cas++	Cl
GT*CGTGCCT*CATGC*AGGCACG*TGGATTGAAAC	32	I-C	I-C
GT*CGTGCCT*CACGT*AGGCACG*TGGATTGAAA	31	I-C	I-C
CGGTTCA*TCCCCGC*TGGC*GCGGGGA*ATAG^{+}	29	I-E	I-E
CGGTTTA*TCCCCGC*TCTC*GCGGGGA*ACAC^{+}	29	I-E	I-E
CGGTTTA*TCCCCGC*TGAC*GCGGGGA*ACAC^{+}	29	I-E	I-E
CGGTTTA*TCCCCGC*TGGC*GCGGGGA*ACAC^{+}	29	I-E	I-E
CGGTTTA*TCCCCGC*TCGC*GCGGGGA*ACAC^{+}	29	I-E	I-E
CGGTTTA*TCCCCGC*TAGC*GCGGGGA*ACAC^{+}	29	I-E	I-E
GAAACA*CCCCCAC*GTGC*GTGGGG*AAGAC*$^{++}$	28	I-E	I-E
GAAACA*CCCCCAC*GTGC*GTGGGG*AAGGC**$^{++}$	28	I-E	I-E
GTGCA*CTGCC*GTACA*GGCAG*CTTAGAAA	28	I-F	I-F
GTTCA*CTGCC*GCATA*GGCAG*CTTAGAAA	28	I-F	I-F
GTTCA*CTGCC*GTGCA*GGCAG*CTTAGAAA	28	I-F	I-F
GTTCA*CTGCC*GTATA*GGCAG*CTTAGAAA	28	I-F	I-F
GTTC*GCTGCC*GTGCA*GGCAGC*TTAGAAA	28	I-F	I-F
GTTCA*CTGCC*GTACA*GGCAG*CTTAGAAA	28	I-F	I-F

The palindrome identified in each CDR is underlined.

* CDR associated with the 20DRs array in Ag1 strain, the 3DRs array in Ag2 strain, and the DD3 arrays (Supporting Table S2).

** CDR associated with the 5DRs arrays in Ag1 and Ag2 strains (Supporting Table S2).

$^{+}$ CDR-I group

$^{++}$ CDR-II group

Table 3. Genomic contexts

Genomic

context

Chromosomal region

Species (#genomes)

Strain(s)

A

mdtN - phnP

marcescens (35)

E28; S5; S8; B3R3; PWN146; CAV1492; 12TM; 2880STDY5682818; 2880STDY5682863; AH0650_Sm1; AR_0130; CAV1761; EGD-HP20; EL1; FZSF02; KS10; MC459; 2880STDY5682911; 2880STDY5683032; 2880STDY5682819; 2880STDY5682934; 2880STDY5682957; 2880STDY5682995; 454_SMAR; 420_SMAR; 395_SMAR; 370_SMAR; 1145_SMAR; MSB1_9C-sc-2280320; N4-5; SER00094; SCQ1; SM03; MGH136; at10508;

grimesii (1)

NBRC 13537

inhibens (1)

PRI-2c

nematodiphila (1)

DH-S01

plymuthica (6)

AS9; AS12; AS13; NCTC8015; NCTC8900; V4

unknown (5)

TEL; SSNIH1; KUDC3025; MYb239; JUb9

rubidaea (6)

FGI94; NCTC10848; FDAARGOS_880; NCTC10036; NCTC12971; FDAARGOS_926

B

puu genes-*mnmA*

fonticola (2)

DSM 4576; 5l

rubidaea (7)

NCTC10848; FDAARGOS_880; NCTC9419; NCTC10036; NCTC12971; FDAARGOS_926; FGI94

unknown (2)

JUb9; MYb239

C

osmE - *sox*G

rubidaea (4)

NBRC 103169; CFSAN059619; NCTC12971; FDAARGOS_926

D

ampC - *yebZ*

rubidaea (5)

FDAARGOS_926; NCTC12971; NCTC10036; NCTC9419; FDAARGOS_880;

unknown (1)

JUb9

E

nrdG - *bglH*

oryzae (1)

J11-6

F

sucD - v*asK*

G

gntR - *cda*

unknown (1)

Ag1

H

gutQ - *queA*

symbiotica (1)

CWBI-2.3

FIGURE LEGENDS

Figure 1.

Schematic diagram of the three categories of arrays described in the study.

DRs and spacers are depicted with diamonds and rectangles respectively.*cas* genes are shown as arrows pointing in the direction of transcription. The yellow color highlights the consistency between the DR type and the cas subtype; while the blue color indicates inconsistency.

Figure 2. Architectures of canonical and variant *cas* gene sets.

Genes are shown as arrows pointing in the direction of transcription. Grey shadows highlight the distinguishing features of the variants I-ES1, I-ES2, and I-F1S1. Species in which the architectures were detected are reported on the right side and the number of genomes is reported in brackets. Truncated *cas* gene sets (due to end of contigs) were not shown. (A) Genetic organization of the canonical*cas* gene set I-C. (B) Genetic organization of *cas* gene sets for the canonical I-E and the variants I-ES1 and I-ES2. The WYL domain is highlighted as a red arrow. (C) Genetic organization of*cas* gene sets for the canonical I-F1 and the variant I-F1S1.

Figure 3. Distribution of CRISPR-positive genomes.

Solid boxes represent the total number (top of boxes) of genomes analyzed per species. Dashed boxes show the number (top of boxes) of genomes for which CRISPR-Cas system(s) or CRISPR array(s) were detected.

Figure 4. Cas3 phylogenetic tree.

Species are shown in different colors. In brackets, the accession number of the *cas3* nucleotide sequence is reported.

Figure 5. 16S rRNA gene phylogenetic tree.

Species are shown in different colors. In brackets, the accession number of the 16S rRNA gene nucleotide sequence is reported.

Figure 6. Schematic diagram of the shared genomic contexts A to D .

Capital letters on the left indicate the type of genomic context. The pink dashed box represents the genomic region harboring *cas* set and/or CRISPR array(s). Black thick lines depict flanking regions. Genes are shown as arrow boxes pointing in the direction of transcription.

Supporting Information

Supporting Table S1: list of *Serratia* genome assemblies.

Supporting Table S2: *cas* gene set positive contigs/scaffolds.

Supporting File S3: spacers' analyses.

Supporting Table S4: distribution of genomic contexts.

Supporting Figure S1: 16S rRNA gene phylogenetic tree.

Figure 1. Schematic diagram of the three categories of arrays described

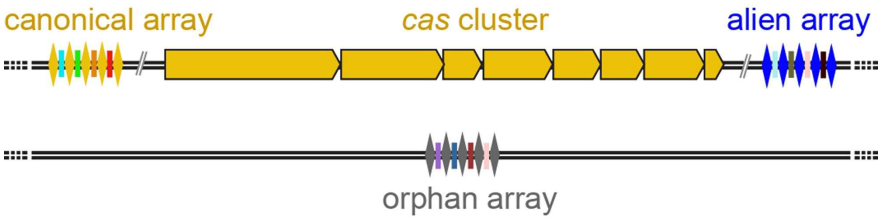

Figure 2. Architectures of canonical and variant cas gene sets.

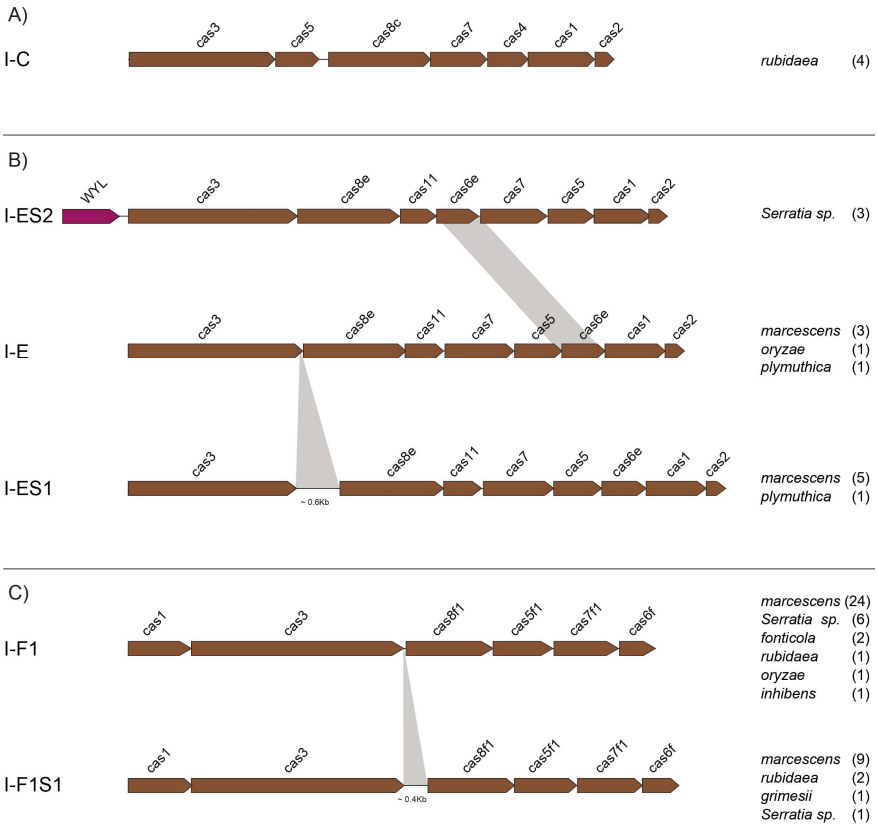

Figure 3. Distribution of CRISPR-positive genomes.

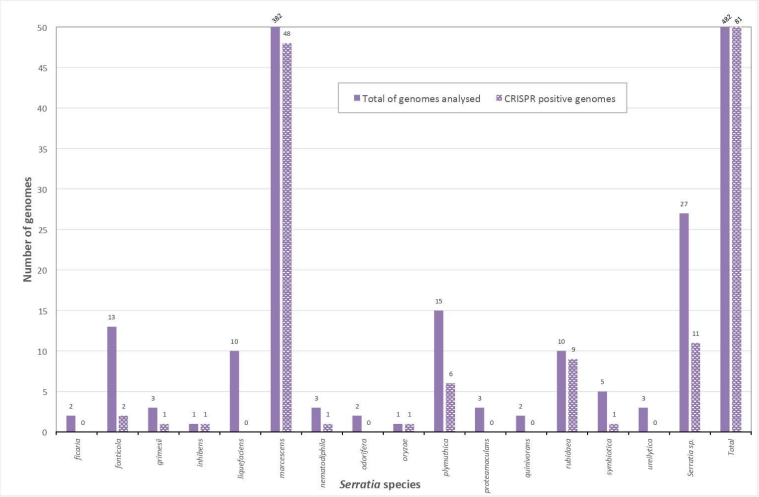

Figure 4. Cas3 phylogenetic tree.

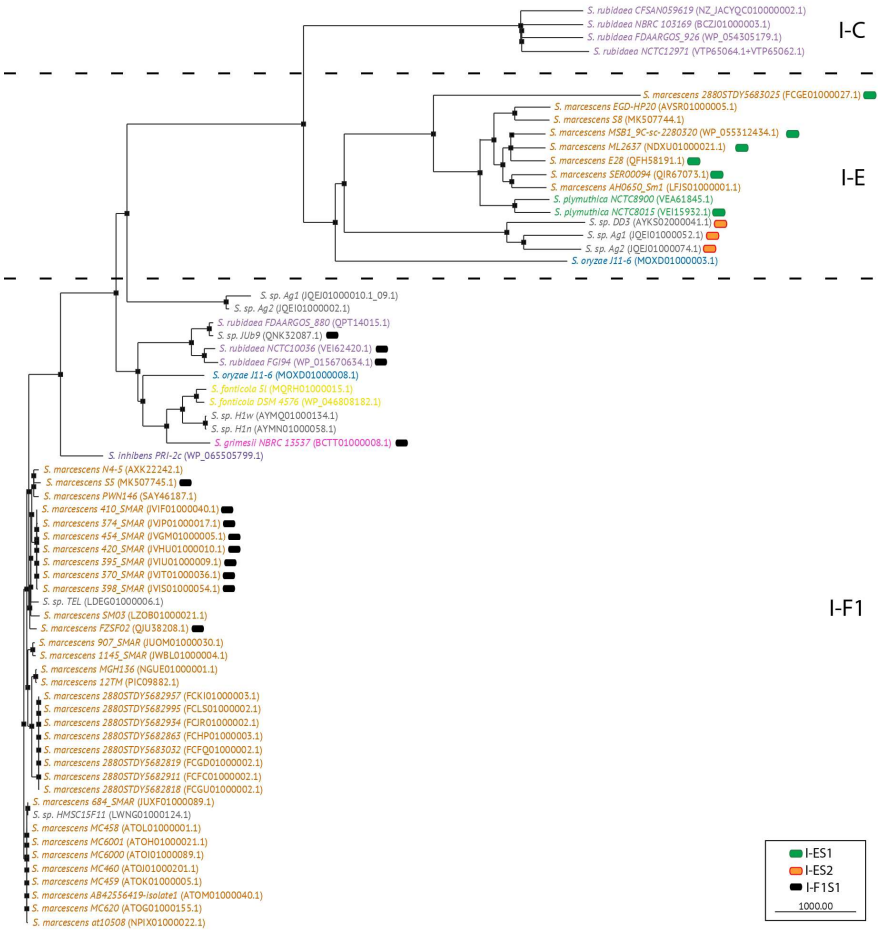

Figure 5. 16S rDNA phylogenetic tree.

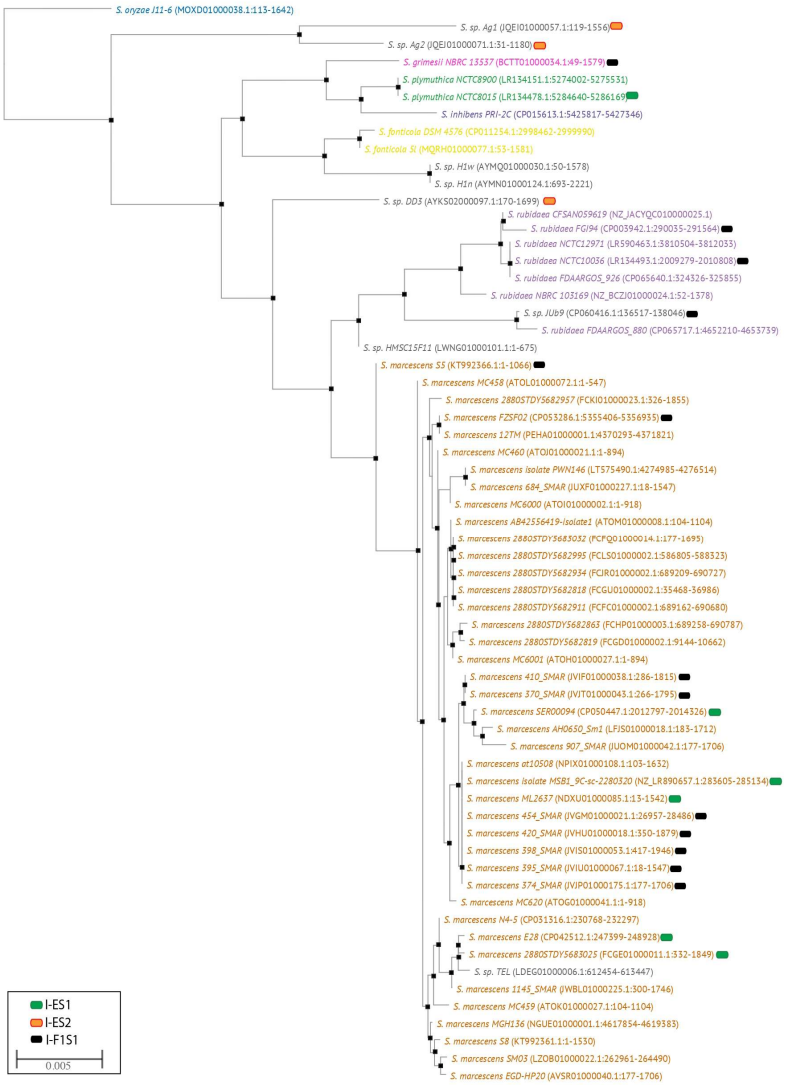

Figure 6. Schematic diagram of the shared genomic contexts A to D.