LARISSA CARVALHO FERREIRA

SERRATIA GENOMICS: ASSEMBLY, ANNOTATION AND COMPARATIVE ANALYSES

LAVRAS-MG
2018

LARISSA CARVALHO FERREIRA

SERRATIA GENOMICS: ASSEMBLY, ANNOTATION AND COMPARATIVE ANALYSES

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia/Fitopatologia, área de concentração em Fitopatologia, para a obtenção do título de Mestre.

Prof. Dr. Jorge Teodoro de Souza
Orientador

LAVRAS-MG
2018

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pela própria autora.

Ferreira, Larissa Carvalho.
SERRATIA GENOMICS: ASSEMBLY, ANNOTATION AND COMPARATIVE ANALYSES / Larissa Carvalho Ferreira. - 2018.
83 p. : il.

Orientador: Jorge Teodoro de Souza.
.
Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Biological control of plant diseases. 2. Complete genome sequence. 3. Comparative genomics. I. de Souza, Jorge Teodoro. II. Título.

LARISSA CARVALHO FERREIRA

SERRATIA GENOMICS: ASSEMBLY, ANNOTATION AND COMPARATIVE ANALYSES

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agronomia/Fitopatologia, área de concentração em Fitopatologia Molecular, para a obtenção do título de Mestre.

APROVADA em 24 de Agosto de 2018.
Dr. Jorge Teodoro de Souza UFLA
Dr. Daniel P. Roberts USDA
Dr. Jude E. Maul USDA
Dr. Victor Satler Pylro UFLA

Prof. Dr. Jorge Teodoro de Souza
Orientador

LAVRAS-MG
2018

AGRADECIMENTOS

Agradeço primeiramente ao Professor, orientador e amigo Jorge. Durante esses anos de convivência você foi como um pai para mim. Obrigada por estar SEMPRE disposto a me ajudar. Obrigada por todas as aulas e momentos de ensinamentos. Obrigada me inspirar diariamente; quando eu crescer, quero ser uma cientista como você!

Ao demais professores do Departamento de Fitopatologia, por contribuírem em minha formação como Mestre. Agradeço especialmente ao professor Flávio pelos conselhos, mentoria e ensinamentos.

Agradeço à minha família e amigos, que entenderam minha ausência e me apoiaram sempre. Em especial quero agradecer ao meu pai, minha mãe e meu irmão: vocês são meu tudo!

Aos meus companheiros diários Carol, Júlio, Kize, Larissa, Luísa, Pablo, Paula, Rafa, Vanessa e Yasmim, agradeço pela convivência, pelas risadas e pela torcida. Agradeço também à Vanessa, Marcela e Norminha, pelo companheirismo dentro e fora das salas de aula.

Ao LGCM, agradeço pela ajuda durante e após minha estadia na UFMG. Vocês foram fundamentais na execução dessa dissertação.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

Por fim, agradeço a todos que contribuíram de alguma forma para o desenvolvimento desse trabalho. Muito obrigada!

ABSTRACT

Serratia is a genus of gram-negative bacteria, widespread in nature, important in agricultural, medical and industrial scenarios. Isolates from this *taxon* exhibits very diverse biological functions such as plant-associated (endophytes, plant growth-promoters, rhizobacteria, phytopathogens), insect-associated (endosymbionts, entomopathogens), fungus-associated (symbionts) and human pathogens. These different lifestyles are determined by the genetic information that each strain carries. The current DNA sequencing technologies provide us data to investigate this variation through genomic studies. Therefore, the aim of this work was to study the *Serratia* genus using genomic approaches. The biological control agent *Serratia marcescens* strain N4-5 was sequenced, the nucleotide sequences were assembled into the whole genome and annotated. The N4-5 genome comprises a singular chromosome of 5,074,473 bp, with 59.7% GC content and a naturally occurring plasmid. Both sequences were deposited in GenBank database under the accession numbers CP031315 and CP031316. From this newly sequenced genome, *in silico* comparisons of all other *Serratia* complete genomes available in GenBank were performed. Firstly, a taxonomical review of the *Serratia* genus was conducted based on multi-criteria, namely dDDH, ANI, 16S identity, phylogenetic trees of seven housekeeping genes individually and concatenated (MLSA), as well as phylogenomic tree with whole genomes. These analysis uncovered two misidentifications, supported a recent proposal of a novel *Serratia* species, confirmed the taxonomic placement of most strains and revealed that many *Serratia* genomes that are publicly deposited in GenBank are classified incorrectly. These organisms were correctly renamed and the two genomes erroneously identified as *Serratia* were excluded from the analyses. From these ascertained genomes, an updated pan-, core- and accessory-genomes were constructed. Analysis revealed an open pan-genome and 546 core genes. Descriptions of *Serratia* spp. genetic organization and presence of secretion systems, secondary metabolites biosynthetic gene clusters, chitinase genes and CRISPR arrays revealed no correlation between genome relatedness and these traits. Analysis of these genomic features evidenced that they are not related with the phenotypes/lifestyles exhibited by *Serratia* spp. strains. Beyond the new information provided on the plant-beneficial strain *S. marcescens* N4-5, altogether these results provide better understanding of *Serratia* at the genus level.

Keywords: Biological control of plant diseases. *Serratia marcescens.* Complete genome sequence. Genus-wide comparisons. Secondary metabolites. Secretion systems. CRISPR. Chitinases.

RESUMO

Serratia é um gênero de bactérias gram-negativas, bem difundido na natureza, importante em cenários agrícolas, médicos e industriais. Isolados desse *taxon* apresentam funções biológicas muito diversas, como por exemplo associados a plantas (endofíticos, promotores de crescimento, rhizobacterias, fitopatógenos), associados a insetos (endosimbiontes, entomopatogênicos), associados à fungos (simbiontes) e patógenos humanos. Esses diferentes estilos de vida são determinados pela informação genética que cada isolado carrega. As tecnologias de sequenciamento de DNA atuais proporcionam dados para investigar essa variação através de estudos genômicos. Assim, o objetivo desse trabalho foi estudar o gênero *Serratia* usando análises genômicas. O agente de controle biológico *Serratia marcescens* isolado N4-5 foi sequenciado, suas sequencias de nucleotídeos foram reunidas e o genoma completo foi obtido, e por sua vez, anotado. O genoma da N4-5 compreende um único cromossomo com 5,074,473 bp com 59% de conteúdo GC e um plasmídio com 11,089 bp. Ambas sequencias foram depositadas no banco de dados GenBank sob os números de acesso CP031315 and CP031316. A partir desse novo genoma sequenciado, foram feitas comparações *in silico* de todos os outros genomas completos de *Serratia* disponíveis no GenBank. Primeiramente, uma revisão taxonômica do gênero *Serratia* foi realizada baseada em multicritérios, sendo eles dDDH, ANI, identidade de 16S, árvores filogenéticas de sete genes housekeeping individuais e concatenados (MLSA), bem como, árvore filogenômica com genomas completos. Essas análises revelaram dois erros de identificação, confirmou a posição taxonômica da maioria dos isolados e revelou que muitos dos genomas de *Serratia* que estão publicamente depositadas no GenBank estão nomeadas incorretamente. Esses organismos foram corretamente renomeados e os dois genomas identificados erroneamente como *Serratia* foram excluídos das análises. A partir desse grupo de genomas corrigidos, foram construídos pan- e core-genoma atualizados. A análise revelou um pan-genoma aberto e 546 genes conservados. Descrições da organização genética e presença de sistemas de secreção, cluster de genes biossintéticos de metabólitos secundários, genes da quitinase e arranjos CRISPR em *Serratia* spp. revelaram a falta de correlação entre similaridade de genomas e esses atributos. Análises dessas características genômicas evidenciaram que elas não estão relacionadas com os fenótipos/estilos de vida exibidos pelos isolados de *Serratia* spp. Além das novas informações fornecidas sobre a cepa benéfica para plantas, *S. marcescens* N4-5, ao todo, esses resultados fornecem uma melhor compreensão de *Serratia* ao nível do gênero.

Palavras-chave: Controle biológico de doenças de plantas. *Serratia marcescens*. Sequência completa do genoma. Comparações em todo o gênero. Metabólitos secundários. Sistemas de secreção. CRISPR. Quitinases.

SUMMARY

FIRST PART

1 INTRODUCTION

Serratia are gram-negative bacteria widespread in the environment. These rod-shaped microorganisms can be found in water, soil, air, animals and plants. *Serratia* spp. isolates may be pathogenic to humans, insects and plants but also promote plant growth and act as biocontrol agents. Some strains are able to synthesize a pigment called prodigiosin. This molecule has potential to be used in medicinal and agricultural applications as it can destroy cancer cells and is also able to inhibit microorganisms. Moreover, *Serratia* spp. are able to synthesize and secrete other molecules with agricultural importance, such as chitinases, antibiotics, toxins, effectors, among others.

Plant-beneficial strains of the species *S. marcescens, S. plymuthica, S. fonticola* and *S. proteamaculans* have shown potential uses in experimental systems, but are not yet exploited in agriculture. Their role as biocontrol agents has been demonstrated through seed treatment, *in vitro* assays and application of cell suspensions, extracts and metabolites. Furthermore, their capacity to colonize plants provides benefits to its host by increasing the availability of nutrients through siderophores and nitrogen fixation, induction of defense mechanisms against pathogens and/or by direct antagonism towards pathogenic microorganisms. The ability to produce and secrete antimicrobial compounds as well as their ability to associate with plants is intrinsically related to their genetic characteristics. Interestingly, there is a large variation of phenotypes exhibited by different strains within the same *Serratia* species. These differences were explored in this study.

Due to the great advancements of sequencing technologies, numerous organisms are being sequenced and the amount of information is rising vertiginously. For instance, the amount of whole genome shotgun (WGS) sequences available in GenBank, one of the three primary databases, doubled from 2016 to 2018. The use and development of computational tools has been employed to explore these new genomic data. However, arranging the complete nucleotide sequence of the genome is only the beginning of genomics. Once the assembly is completely done, the generated sequence will be used for structural and functional genomics studies, as well as in comparative genomics among other organisms of interest. Therefore, the deposit of newly assembled genomes brings a responsibility with it once the assembly quality affects the results obtained in every research that uses this genome sequence. It also affects the constructions of updated pan- and core- genomes, frequently used to comprehend the diversity and evolution of the genus under study.

Many genetic characteristics in bacteria were acquired through horizontal gene transfer, which form syntenic blocks recognized as genomic islands. Therefore, studying bacterial genetic-related traits is a complex task and the employment of bioinformatics tools is an alternative way to answer some of these biological questions and generate new information. In spite of the wide applicability of these *in silico* tools, *in vivo* assays are not dispensable. For example, although plant growth promotion by *Serratia* is mediated by genetic information, the study of the genes involved in this interaction is only possible thanks to the employment of molecular techniques combined with *in silico* analyses. Therefore, these *in silico* studies along with *in vivo* assays provide a more complete and accurate way to understand selected characteristics of the organisms under study. In this study, genomics tools were used to understand some traits of the genus *Serratia*. In the first chapter, the genome of the biological control agent, *S. marcescens* strain N4-5, was sequenced, assembled and annotated. This new and complete genome sequence enabled to gain novel insights on the N4-5 strain and to uncover an assembly artefact in the N4-5 genome and in other 48 *Serratia* spp. complete genomes. In the second chapter, all the *Serratia* complete genome sequences available were gathered into a genus-wide comparative analysis. Several bioinformatics approaches were employed to study *Serratia* taxonomy, pan-genome, genetic organization of secretion systems, distribution of secondary metabolites biosynthetic gene clusters and the presence of other advantageous features, such as chitinases and CRISPR arrays.

2 LITERATURE REVIEW

2.1 Overview on *Serratia*

The gram-negative bacteria in the genus *Serratia* are members of the *Enterobacteriaceae* family (GRIMONT; GRIMONT, 1984). The genus *Serratia* is consistently different from *Escherichia*, *Shigella*, *Salmonella*, *Citrobacter*, *Klebsiella*, and *Enterobacter*, as shown by phylogenetic studies, physical properties, aminoacid sequence analyses and biochemical tests (GRIMONT et al., 1978a).

Currently 17 species of *Serratia* are recognized within the genus. These species are: *S. aquatilis* (KAMPFER; GLAESER, 2016), *S. entomophila* (GRIMONT et al., 1988), *S. ficaria* (GRIMONT et al., 1979), *S. fonticola* (GAVINI et al., 1979; GEIGER et al., 2010), *S. grimesii* (GRIMONT et al., 1982), *S. liquefaciens* (BASCOMB et al., 1971), *S. marcescens* (BIZIO, 1823; DE TONI; TREVISAN, 1889), *S. myotis* (GARCIA-

FRAILE et al., 2015), *S. nematodiphila* (ZHANG et al., 2009), *S. odorifera* (GRIMONT et al., 1978), *S. plymuthica* (BREED et al., 1948; LEHMANN; NEUMANN, 1896), *S. proteamaculans* (GRIMONT et al., 1978b; PAINE; STANSFIELD, 1919), *S. quinivorans* (ASHELFORD et al., 2002; GRIMONT et al., 1983), *S. rubidaea* (EWING et al., 1973; STAPP, 1940), *S. symbiotica* (MORAN et al., 2005; SABRI et al., 2011), *S. ureilytica* (BHADRA et al., 2005) and *S. vespertilionis* (GARCIA-FRAILE et al., 2015).

These rod-shaped bacteria are found in several ecological niches, such as: soil (GIRI et al., 2004; KAMENSKY et al., 2003; LAVANIA; NAUTIYAL, 2013), water (GAVINI et al., 1979; HENRIQUES et al., 2013; KÄMPFER; GLAESER, 2016), air (BENCINI et al., 2008), plants (LIM et al., 2015; AFZAL et al., 2017; ZAHEER et al., 2016), animals (ABEBE-AKELE et al., 2015; GARCÍA-FRAILE et al., 2015; GERC et al., 2012; FLYG; XANTHOPOULOS, 2017; VICENTE et al., 2016) and human beings (BONNIN et al., 2015; ROY et al., 2014). Naturally occurring strains of *Serratia* may be pathogenic or non-pathogenic to animals, plants and humans.

There are 523 *Serratia* spp. genomes sequenced available in the GenBank database, from which only 62 are complete sequences (BENSON et al., 2005). These *Serratia* genomes are represented by *S. marcescens* (407 genome assemblies); *S. plymuthica* (15), *S. fonticola* (10), *S. symbiotica* (6), *S. nematodiphila* (6), *S. liquefaciens* (8), *S. odorifera* (2), *S. proteamaculans* (2), *S. ficaria* (2), *S. rubidaea* (4), *S. grimesii* (3), *S. ureilytica* (2), *S. oryzae* (1) and *Serratia* sp. (55). *Serratia marcescens* accounts for 77.8% of all *Serratia* genomes sequenced. The species *S. aquatilis, S. entomophila, S. glossinae, S. grimesii, S. myotis, S. nematodiphila, S. odorifera, S. oryzae* and *S. quinovorans* do not have sequenced genomes or only draft genomes available, i.e. at the contigs or scaffolds level. The amount of whole genome sequences in *Serratia* taxa do not cover 100% of its species, however it is enough for comparative genomic studies such the ones presented by Abebe-Akele et al (2015), Manzano-Marín et al (2012), Li et al (2015) and the present study.

2.2 Plant-beneficial properties of *Serratia* spp.

Non-pathogenic bacteria associated with plants play critical roles in plant development and in its interaction with other microorganisms (FINKEL et al., 2017). The way bacteria alter plant phenotypes and its microbiome can be through the biocontrol of pathogens (by antibiosis, competition, predation, plant growth promotion and/or induction of plant defense mechanisms), enhancing nutrient uptake (by phosphate

solubilization, nitrogen fixation, siderophores production, etc), providing tolerance to (a)biotic stresses, hormone production, among other mechanisms (BETTIOL, 1991; MENDES et al., 2013; ROMEIRO, 2007). Past research indicates that *Serratia* spp. have multiple ways of benefiting plants, which are highlighted below.

2.2.1 Plant colonization and growth promotion by *Serratia* spp.

Plant colonizing bacteria can be found in the rhizosphere (as free microorganisms, attracted by root exudates), or inside plant tissues as endophytes (LUGTENBERG; DEKKERS; BLOEMBERG, 2001). Endophytic bacteria are those that live symbiotically in the interior of the host plant without causing apparent damage (HALLMANN et al., 1997), or are not phytopathogenic in part of their life cycle (WILSON, 1995). Unlike fungi, bacteria enter plant tissues passively through wounds, hydathodes, nectaries, stomata, lenticels, emergence of lateral roots and radicles in germination (GAGNÉ et al., 1987; HALLMANN et al., 1997; HUANG, 1986; ROOS; HATTINGH, 1983; SCOTT et al., 1996). Interestingly, most of the endophytic bacteria come from the rhizospheric environment (COMPANT et al., 2005; HARDOIM et al., 2008). The rhizosphere is a highly competitive environment for microorganisms to settle and to utilize nutrients. Therefore, the organisms that are highly effective in colonizing plant tissues and obtaining nutrients, either phytopathogenic or potentially beneficial, will proliferate and perhaps have an outcome on plant growth and development (HAAS; KEEL, 2003). To colonize the plant tissues internally, bacterial endophytes are thought to possess specific genomic traits when compared to the bacteria from the rhizosphere, although definitive genes related with the endophytic behavior have not been identified yet (SANTOYO et al., 2016). In 2014, Ali et al utilized a bioinformatics approach to predict genes responsible for bacterial endophytic lifestyle. They identified 40 bacterial genes with the highest probability of being involved in endophytic colonization, however, this result still awaits to be tested and confirmed experimentally.

Initially, endophytic microorganisms were considered neutral with regard to their effects on host plants, however, their positive impact have been verified in a broad range of crops (RYAN et al., 2008). These microorganisms may contribute directly to plant growth by enhancing nutrient uptake and the production of phytohormones and siderophores (JUNG et al., 2015; KIM et al., 2011; SHISHIDO et al., 1999). Indirectly, they may also reduce microbial populations that are harmful to the plant, acting as agents of biological control through competition, antibiosis or systemic resistance induction

(RAMAMOORTHY et al., 2001; STURZ et al., 2000). Species of *Serratia* feature several of these mechanisms. For example, a *S. marcescens* isolated from citrus rhizosphere not only promoted citrus growth but also suppressed *Phytophthora parasitica* by more than 50% (QUEIROZ; MELO, 2006). *Serratia plymuthica* isolated from rapeseed roots manifested the ability to promote rapeseed growth, to control soil-borne fungal pathogens, such as *Verticillium dahlia* and *Rhizoctonia solani* and to enhance oilseed production (ALSTRÖM, 2001; NEUPANE et al., 2012a, 2012b). Cucumber seeds treated with *S. marcescens* strain 90-166 and *Pseudomonas putida* strain 89B-27 decreased the incidence of bacterial wilt disease (KLOEPPER et al., 1993), reduced the number of cucumber mosaic virus-infected plants (CMV) and delayed the development of symptoms in cucumber and tomato (RAUPACH et al., 1996). Also, the same strains of *S. marcescens* and *P. putida* induced systemic resistance in cucumber against bacterial angular leaf (LIU; KLOEPPER; TUZUN, 1995a), against fusarium wilt and fusiform rust in loblolly pine (LIU; KLOEPPER; TUZUN, 1995b; ENEBAK; CAREY, 2000). Similarly, *S. plymuthica* isolated from pumpkin anthosphere showed strong biocontrol activity against *Didymella bryoniae*, the causal agent of black rot in pumpkins, and also enhanced germination rate and controlled damping-off diseases when applied as seed treatment (FURNKRANZ et al., 2012; MULLER et al., 2013).

The endophytic *S. marcescens* strain AL2-16 enhances the growth of *Achyranthes aspera* L., a medicinal plant (DEVI et al., 2017). Likewise, *Serratia marcescens* has been described to be an important endophyte in rice (GYANESHWAR et al., 2001), root and stem of sweet corn and cotton (MCINROY; KLOEPPER, 1994). Later, endophytic colonization and *in planta* nitrogen fixation by a diazotrophic *Serratia* sp. in rice was demonstrated by Sandhiya et al (2005). Selvakumar et al (2008) reported the promotion of plant growth and its tolerance to cold by *S. marcescens* strain SRM isolated from flowers of summer squash (*Cucurbita pepo*). Lavania et al. (2006) first reported the growth promotion and biological control activity of phenolic compounds from *S. marcescens* NBRI1213 against *Phytophthora nicotianae*. Other *Serratia* species have been reported to enhance plant growth as well, namely *S. fonticola*, *S. proteamaculans* and *S. liquefaciens* (JUNG et al., 2017; TAGHAVI et al., 2009; ZHANG et al., 1996).

2.2.2 Secretion systems

According to Green and Mecsas (2016) protein secretion is an essential cell function in prokaryotes that transports “proteins from the cytoplasm into other

compartments of the cell, the environment, and/or other bacteria or eukaryotic cells". The proteins secreted by bacteria play a key role in cell detoxification and in intra- and inter-specific interactions (ABBY et al., 2016; RUHE et al., 2013; VIPREY et al., 1998). The process of secreting bacterial compounds occurs via cellular apparatuses called secretion systems (SSs) and, for each role cited above, a specific SS is required (BLEVES et al., 2010; CHANG et al., 2014; DALBEY; KUHN, 2012). Therefore, studying the secreted proteins as well as their secretion systems is essential to understand bacterial interactions with hosts and adjacent organisms (GERLACH; HENSEL, 2007).

There are several SSs types, however this review will cover only types III, IV and VI. The type III secretion system (T3SS) is involved in the transportation of bacterial cell proteins directly into the cytoplasm of the host cell through a complex secretory apparatus that crosses the inner and outer membrane of the bacterial cell (HECK, 1998). The T3SS apparatus features four structures: the cytosol platform, the export apparatus and the needle complex, which is the envelope-spanning basal body plus the extracellular needle (BURKINSHAW; STRYNADKA, 2014; GALAN et al., 2014). According to Hueck (1998), T3SS allows gram-negative bacteria to introduce toxic proteins into the cytoplasm of eukaryotic host cells. The bacterial type IV secretion system (T4SS) is involved in the conjugative and noncontact-dependent transfer of DNA (HAMILTON et al., 2005; WARD et al., 1988) and it mediates the transport of macromolecules, including effectors, through the cell lining of gram-negative and positive bacteria (BYNDLOSS et al., 2016; CASCALES; CHRISTIE, 2003; GROHMANN et al., 2003). The type VI secretion system (T6SS) is involved in pathogenesis and inter-bacterial competition as it delivers toxins into prokaryotic and eukaryotic cells (MOUGOUS et al., 2006; PUKATZKI et al., 2006). The first such "antibacterial" SS was described by Hood et al in 2010 in *Pseudomonas aeruginosa* and, shortly after, in other bacterial species, including *Vibrio cholerae*, *Burkholderia thailandensis* and *Serratia marcescens* (MACINTYRE et al., 2010; MURDOCH et al., 2011; SCHWARZ et al., 2010). T6SSs kill adjacent, non-immune bacterial cells by secreting toxins directly into the periplasm of the target cells by contact (ENGLISH et al., 2012; HOOD et al., 2010; RUSSELL et al., 2011). Antibacterial T6SSs secrete a broad spectrum of antibacterial substances and play a key role in survival and prospection of bacteria in an environment with multiple organisms (DURAND et al., 2014). In *Serratia marcescens*, the antibacterial activity of the small-secreted protein Ssp4 (and previously Ssp1 and 2) is dependent on T6SS (FRITSCH et al., 2013). Moreover, *S. marcescens* utilize T6SS to target bacterial competitors, such as

Enterobacter cloacae (MURDOCH et al., 2011). In short, secretion systems are essential for bacterial survival and evolution.

There are several public webservers and downloadable applications that identify components, clusters of components, or complete (eventually scattered) bacterial protein secretion system, such as MacSyFinder and T346HUNTER (ABBY et al., 2014; MARTINEZ-GARCIA et al., 2015), and also the secreted substances, such as SignalP and SecretomeP (BENDTSEN et al., 2004; BENDTSEN et al., 2005; NIELSEN, 2017).

2.2.3 Secondary metabolites

Bacterial secondary metabolites are non-essential compounds naturally produced that are useful for survival in nature (DEMAIN; FANG, 2000). Among these natural products, the antibiotics are key metabolites for microbe competition but also carry industrial and agricultural interests. *Serratia* spp. synthesize a number of antibiotics, such as prodigiosin ($C_{20}H_{25}N_3O$ - 2-methyl-3-amyl-6-methoxyprodigiosene), a linear tripyrrole red pigment bioactive secondary metabolite (RAPOPORT; HOLDEN, 1962). *Serratia marcescens* is the major producer of prodigiosin, however this pigment is also produced by other bacteria, such as *Alteromonas rubra* (GERBER; GAUTHIER, 1979), *Hahella chejuensi* (KIM et al., 2007), *Pseudoalteromonas denitrificans* (KAWAUCHI et al., 1997), *P. rubra* (FEHÉR et al., 2008), *Pseudomonas magnesiorubra* (GHANDI et al., 1976), *Pseudovibrio denitrificans* (GUZMAN et al., 2007), *Serratia plymuthica* (LEE et al., 2011), *Streptomyces coelicolor* (TSAO et al., 1985), *Vibrio psychroerythreus* (D'AOUST; GERBER, 1974), *V. gazogenes* (ALIHOSSEINI et al., 2008) and *Zooshikella rubidus* (LEE et al., 2011). Prodigiosin has recently received renewed attention due its antialgal, antibacterial, antifungal, antimalarial, antiprotozoal, anticancer, immunosuppressive, antiproliferative, nematicidal and UV-protective activities (BORIC et al., 2011; EL-BONDKLY et al., 2012; LEE et al., 2011; MONTANER; PEREZ-THOMAS, 2001; PARK et al., 2012; PATIL et al., 2011; RAHUL et al., 2014; SOTO-CERRATO et al., 2004; SURYAWANSHI et al., 2015). Moreover, its insecticidal activity, larvicidal and pupaecidal potential against *Aedes aegypti* and *Anopheles stephensi* was also reported (PATIL et al., 2011; WANG et al., 2012). The production of prodigiosin has been shown to be influenced by many factors, such as species and environmental factors, including inorganic phosphate availability, dissolved oxygen level, light, media composition, temperature, pH and incubation time (SOLÉ et al., 1997; SLATER et al., 2003; RYAZANTSEVA et al., 2012; WANG et al.,

2012; WILLIAMS et al., 1971; WITNEY et al., 1977). In *Serratia* sp., the prodigiosin biosynthesis gene cluster (pig cluster) is encoded by 14 genes and its genomic context as well as the biosynthesis pathway has been already described (CERDEÑO et al., 2001; HARRIS et al., 2004).

Besides prodigiosin, *Serratia* spp. synthesize other antibiotics such as pyrrolnitrin (KALBE et al., 1996; LEVENFORS et al., 2004), althiomycin (GERC et al., 2012), zeamine (HOUDT et al., 2005; HOUDT et al., 2014; MASSCHELEIN et al., 2013), carbapenem (PARKER et al., 1982; THOSON et al., 2000), among others.

2.2.4 Chitinolytic properties

Antagonism to fungal plant pathogens by microorganisms, specifically by the production of chitinases, plays a major role in biological control of diseases (CHERNIN et al., 1996; DAVISON, 1988). Chitin is the second most abundant renewable carbohydrate polymer in nature after cellulose and possibly the most abundant in marine environments (BANSODE; BAJEKAL, 2006). Chitinases hydrolyze chitin, an insoluble β-1,4-linked unbranched polymer of N-acetylglucosamine and a major fungal cell wall component, to oligomers, mainly dimers. Ordentlich et al (1988) showed the degradation of *Sclerotium rolfsii* hyphae by a chitinolytic filtrate obtained from a *S. marcescens* culture. This enzyme also exhibited antifungal activity against *Rhizoctonia solani*, *Bipolaris* sp., *Alternaria raphani*, *Alternaria brassicicola*, revealing a potential industrial application (ZAREI et al., 2011). Kobayashi et al (1995) pointed out that the growth suppression of *Magnaporthe poae* by *S. marcescens* 9M5 was through the secretion of chitinase. In addition to this, chitinase from *S. marcescens* JPP1 has antagonistic activity against aflatoxins (WANG; YAN; CAO, 2014).

Serratia marcescens is a great source of chitinase (LAMINE; LAMINE, 2012). This species was found to be the most active organism among 100 tested for the production of chitinase (MONREAL; REESE, 1969). Chitin-supplemented foliar application of *S. marcescens* GPS 5 improves control of late leaf spot disease of groundnut by activating defense-related enzymes (KISHORE; PANDE; PODILE, 2005). Likewise, application of *S. marcescens* effectively reduced mycelial growth of *Rhizoctonia solani*, the causal agent of the rice sheath blight (SOMEYA et al., 2005). Kalbe et al (1996) showed that *Serratia liquefaciens*, *S. plymuthica* and *S. rubidaea* produce antifungal compounds such as prodigiosin, pyrrolnitrin, chitinases and β-1-3-glucanases and also indirect by the secretion of siderophores.

2.3 From DNA sequencing to genomics

In 1975, Frederich Sanger described a new technology to determine DNA nucleotide sequences (SANGER; 1975; SANGER; COULSON; 1975). The Sanger method revolutionized molecular biology studies and initiated the genomic era. This period was characterized by countless technical advances, mainly due the automation of sequencing processes and the structuring of new equipment and sequencing techniques (HEATHER; CHAIN; 2016). A new technique, not based on the Sanger method, for sequencing DNA segments was presented by the company 454 Life Sciences (acquired by Roche later) in 2002. The 454 sequencing platform consists on the parallelization of sequencing, using nanotechnology and a pyrosequencing methodology (FAKHRAI-RAD et al., 2002; RONAGHI, 2001). The main advantage was the high amount of sequences generated without the need for cloning the DNA fragments prior to sequencing (RONAGHI et al., 1996; 1998; ROTHBERG; LEAMON; 2008). Soon thereafter, new technologies emerged, called ultra-high performance parallel mass sequencing systems, or ultra-high throughput sequencing (JU et al., 2006; WOLD; MEYERS, 2008). Since 2006, Illumina Inc. has made available a new technology, capable of generating about 100 million reads-per-run segments. Although Roche sequencing technology could produce longer reading segments, the Illumina platform allows deeper coverage and greater accuracy at the same cost (LUO et al., 2012). Other technologies, such as Applied Biosystems SOLiD were made available in the period, and newer technologies, such as IonTorrent and PacBio, began to reach large-scale sequencing capabilities (HEATHER; CHAIN; 2016; QUAIL et al., 2012). As a result of these advances, the number of sequenced genomes increases exponentially. To illustrate, the number of bases available in GenBank doubles approximately every 18 months (NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION - NCBI; 2018).

Bacterial taxonomic studies have been greatly impacted by the increasing number of sequenced genomes as well as by the parallel development of computational tools to assess relatedness between whole genome sequences. For instance, Auch et al (2010) developed a digital method for genome-genome comparisons - as an alternative to the wet-lab DNA-DNA hybridization (DDH) method - and it is known as *in silico* DDH (dDDH). Likewise, the similarity between genomes can be calculated using the Average Nucleotide and Amino acid Identity (ANI/AAI) formulas (KONSTANTINIDIS; TIEDJE; 2005a; 2005b). These *in silico* genome-based methods for species delineation

are equally or more precise than other molecular approaches such as 16S gene sequencing analysis, and less tedious than *in vivo* methods (ARAHAL, 2014; GALPERIN; KOLKER; 2006; KONSTANTINIDIS; TIEDJE; 2005a; 2005b; ROSSELLÓ-MÓRA; AMANN, 2015). Consequently, taxonomic reviews, reclassification proposals and misidentifications are being uncovered by analyses based on these whole genome metrics (COUTINHO et al., 2016; FERREIRA et al., 2017; SAFNI et al., 2014). Besides taxonomy, functional genomics and comparative genome analysis have provided great findings into understanding the genetic information comprised in organisms (IGUCHI et al., 2014; MANZANO-MARÍN et al., 2012; MITTER et al., 2013; WANG et al., 2017). All these studies mentioned above are feasible if whole genome sequences are available and assembled (EKBLOM; WOLF, 2014). The union of millions of WGS sequencing reads into a linear nucleotide genomic sequence is realized through thorough genome assemblies' methods, reviewed next.

2.3.1 Bacterial genome assembly and annotation

As stated previously, the development of new technologies capable of performing large-scale DNA sequencing have boosted the field of genomics, which is another way of studying microorganisms (GOODWIN et al., 2016). The employment of bioinformatics along with the "OMICS" sciences revolutionized biological research. However, even though these technologies yield millions of sequencing fragments, the assembly of these reads into full-length chromosomes is a challenge (BRADNAM et al., 2013; POP, 2009; SALZBERG et al., 2012).

Genome assembly consists in a set of procedures that aim to arrange a large number of DNA sequences (reads) in a linear form, in order to represent the genome of the studied organism (FLICEK; BIRNEY, 2009; NAGARAJAN; POP, 2013; POP, 2009). These procedures convert millions of reads into contigs and then into scaffolds. According to Yandell and Ence (2012) "contigs are contiguous consensus sequences that are derived from collections of overlapping reads and scaffolds are ordered and orientated sets of contigs". In prokaryotes, the genome assembly can be done in two different ways: one is the reference-guided genome assembly as it uses existing information from a genome that is already fully assembled; the other is the *de novo* or "*ab initio*" assembly, which is performed from scratch without any references and thus reduces bias in the genome (POP, 2009). The *de novo* approach is only possible due to the development of assemblers and increasingly accurate assembly algorithms (MEDVEDEV et al., 2007;

MYERS, 1995). Also, the *de novo* approach can be employed to reconstruct regions of the sequenced genome that are significantly different or inexistent in the reference, e.g. large insertions (POP, 2009). The current *de novo* assemblers available use the following algorithms: overlay-layout-consensus (OLC), *de Brujin* graph, string graph, greedy and hybrid algorithms (KHAN et al., 2018; MILLER et al., 2010). These algorithms have distinct performances on different data sets. For example, the OLC algorithm is most suitable for short sequences of small genomes whereas for large data sets of short reads the *de Bruijn* graph is more appropriate (LI et al., 2012; ZHANG et al., 2011). To choose the appropriate assembler, in addition to the sequencing platform used and read length, the type of reads generated should also be considered, i.e., if they are paired-end, mated-pair, single, long and/or short reads (BAKER; 2012; KHAN et al., 2018). For instance, assemblers based on *de Brujin* graph are the best options for paired-end and single-end data sets from prokaryotic genomes, in terms of memory use, time and accuracy (KHAN et al., 2018). Choosing the right assembler and converting reads into contigs are very early steps within the genome assembly pipeline (BAKER; 2012). By the end of this process, complete genome sequences (one scaffold per chromosome) or draft genomes (more than one scaffold per chromosome) are obtained and ready to be annotated.

Genome annotation is the process of identifying and labelling all the genomic features in a nucleotide genome sequence (RICHARDSON; WATSON, 2013). Functional and structural annotation include the prediction of coding sequences (CDS) and their putative products (YANDELL; ENCE, 2012). It can be done manually and/or by automated annotation tools such as Prokka (SEEMANN; 2014), RAST (AZIZ et al., 2008), KAAS (MORIYA et al., 2007), among others. Manual annotation is time consuming and a tedious process whereas automatic pipelines are simpler and faster, at the same time, they can produce/propagate poor annotations and errors and manual curation is desirable (RICHARDSON; WATSON, 2013). Therefore, high-quality annotations rely on the adoption of manual curation by the scientist and also on the genome assembly quality. As a result, each assembled and annotated genome will affect every study that includes this genome thus genome assembly and annotation should be rigorously done.

REFERENCES

ABBY, S. S.; CURY, J.; GUGLIELMINI, J.; NÉRON, B.; TOUCHON, M.; ROCHA, E. P. C. **Identification of protein secretion systems in bacterial genomes**. Nature Publishing Group, n. March, p. 1–14, (http://dx.doi.org/10.1038/srep23080), 2016.

ABBY, S. S.; NÉRON, B.; MÉNAGER, H.; TOUCHON, M.; ROCHA, E. **MacSyFinder : A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems.** PLOS ONE, v. 9, n. 10, 2014.

ABEBE-AKELE, F.; TISA, L. S.; COOPER, V. S.; HATCHER, P. J.; ABEBE, E.; THOMAS, W. K. **Genome sequence and comparative analysis of a putative entomopathogenic Serratia isolated from *Caenorhabditis briggsae*.** BMC Genomics, v. 16, n. 1, p. 1, 2015.

AFZAL, I.; IQRAR, I.; SHINWARI, Z. K.; YASMIN, A. **Plant growth-promoting potential of endophytic bacteria isolated from roots of wild *Dodonaea viscosa* L.** Plant Growth Regulation, p. 1-10, 2017.

ALI, S.; DUAN, J.; CHARLES, T. C.; GLICK, B. R. **A bioinformatics approach to the determination of genes involved in endophytic behavior in *Burkholderia* spp.** J. Theor. Biol., v. 343, p. 193–198, 2014.

ALIHOSSEINI, F.; JU, K.-S.; LANGO, J.; HAMMOCK, B. D.; SUN, G. **Antibacterial colorants: characterization of prodiginines and their applications on textile materials**. Biotechnology Progress, v. 24, n. 3, p. 742–747, 2008.

ALSTRÖM, S. **Characteristics of bacteria from oilseed rape in relation to their biocontrol activity against *Verticillium dahliae***. J Phytopathol., v. 149, p. 57- 64, 2001.

ARAHAL, D. **Whole-genome analyses: average nucleotide identity**. Methods Microbiol. V. 41, p. 103–122, 2014.

ASHELFORD, K. E.; FRY, J. C.; BAILEY, M. J.; DAY, M. J. **Characterization of *Serratia* isolates from soil, ecological implications and transfer of *Serratia proteamaculans* subsp. *quinovora* Grimont et al. 1982 to *Serratia quinovorans* corrig., sp. nov.** Int.J. Syst. Evol. Microbiol., v. 52, p. 2281-2289, 2002.

AUCH, A. F.; VON JAN, M.; KLENK, H-P.; GOKER, M. **Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison.** Standards in genomic sciences, v. 2 p. 117-134, (DOI:10.4056/sigs.531120), 2010.

AZIZ, R. K., et al. **The RAST Server: rapid annotations using subsystems technology**. BMC Genomics, v. 9, n. 75, 2008.

BAKER, M. **De novo genome assembly: what every biologist should know.** Nature Methods, v. 9, p.333-337, 2012.

BANSODE, V. B.; BAJEKAL, S. S. **Characterization of chitinases from microorganisms isolated from Lonar lake.** Indian Journal of Biotechnology, v. 5, n. 3, p. 357–363, 2006.

BASCOMB, S.; LAPAGE, S. P.; CURTIS, M. A.; WILCOX, W. R. **Numerical classification of the tribe *Klebsielleae*.** J. Gen. Microbiol., v. 66, p. 279-295, 1971.

BENCINI, M. A.; YZERMANA, E. P. F.; BRUINA, J. P.; DEN BOER, J. W. **Airborne dispersion of *Serratia marcescens* as a model for spread of *Legionella* from a whirlpool**. The Royal Institute of Public Health, p. 962–964, 2008.

BENDTSEN, J. D.; JENSEN, L. J.; BLOM, N. VON HEIJNE, G.; BRUNAK, S. **Feature based prediction of non-classical and leaderless protein secretion.** Protein Eng. Des. Sel., v. 17, p. 349-356, 2004.

BENDTSEN, J. D.; KIEMER, L.; FAUSBØLL, A.; BRUNAK, S. **Non-classical protein secretion in bacteria.** BMC Microbiology, v. 5, n. 58, 2005.

BENSON, D. A.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; WHEELER, D. L. **GenBank.** Nucleic Acids Research, v. 33, p. D34–D38, 2005.

BETTIOL, W. **Controle biológico de doenças de plantas.** Jaguariúna: EMBRAPA-CNPDA., 338p., 1991.

BHADRA, B.; ROY, P.; CHAKRABORTY, R. ***Serratia ureilytica* sp. nov., a novel urea-utilizing species.** Int. J. Syst. Evol. Microbiol., v. 55, p. 2155-2158, 2005.

BIZIO, B. **Lettera di Bartolomeo Bizio al chiarissimo canonico Angelo Bellani sopra il fenomeno della polenta porporina.** Biblioteca Italiana o sia Giornale di Letteratura Scienze e Arti (Anno VIII), v. 30, p. 275-295, 1823.

BLEVES, S. et al. **Protein secretion systems in *Pseudomonas aeruginosa*: A wealth of pathogenic weapons.** Int. J. Med. Microbiol., v. 300, n. 534, 2010.

BONNIN, R. A.; GIRLICH, D.; IMANCI, D.; DORTET, L.; NAAS, T**. Draft genome sequence of the *Serratia rubidaea* CIP 103234T reference strain, a human-opportunistic pathogen.** Genome Announcements, v. 3, n. 6, p. e01340-15, 2015.

BORIC, M.; DANEVCIC, T.; STOPAR, D. **Prodigiosin from *Vibrio* sp. DSM 14379; a new UV-protective pigment.** Microbiol. Ecol., v. 62, p. 528-536, 2011.

BRADNAM, K. R.; FASS, J. N.; ALEXANDROV, A.; BARANAY, P.; BECHNER, M.; BIROL, I.; BOISVERT, S.; CHAPMAN, J. A.; CHAPUIS, G.; CHIKHI, R.; et al. **Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species**. Gigascience, v.2, p. 1–31, 2013.

BREED, R. S.; MURRAY, E. G. D.; HITCHENS, A. P. (Editors). **Bergey's Manual of Determinative Bactgeriology**, 6th ed. Williams and Wilkins Co., Baltimore., p. 1-1529, 1948.

BURKINSHAW, B.J.; STRYNADKA, N.C. **Assembly and structure of the T3SS.** Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, v. 1843, p. 1649–1663, 2014.

BYNDLOSS, M. X.; RIVERA-CHÁVEZ, F.; TSOLIS, R. M.; BAUMLR, A. J. **How bacterial pathogens use type III and type IV secretion systems to facilitate their transmission.** Current Opinion in Microbiology, v. 35, p. 1–7, 2016.

CASCALES, E.; CHRISTIE, P. J. **The versatile bacterial type IV secretion systems.** Nat. Rev. Microbiol., v. 1, p. 137–49, 2003.

CHANG, J. H.; DESVEAUX, D.; CREASON, A. L. **The ABCs and 123s of bacterial secretion systems in plant pathogenesis.** Annu Rev Phytopathol., v. 52, n. 317, 2014.

CHERNIN, L.; BRANDIS, A.; ISMAILOV, Z.; CHET, I. **Pyrrolnitrin production by an *Enterobacter agglomerans* strain with a broad spectrum of antagonistic activity towards fungal and bacterial phytopathogens.** Curr. Microbiol., v. 32, p. 208–212, 1996.

COMPANT, S.; REITER, B.; SESSITSCH, A.; NOWAK, J.; CLÉMENT, C.; BARKA, E.A. **Endophytic colonization of *Vitis vinifera* L. by plant growth-promoting bacterium *Burkholderia* spp. strain PsJN.** Appl. Environ. Microbiol., v. 71, p. 1685–1693, 2005.

COUTINHO, F. H.; DUTILH, B. E.; THOMPSON, C. C.; THOMPSON, F. L. **Proposal of ffteen new species of *Parasynechococcus* based on genomic, physiological and ecological features.** Arch Microbiol., v. 198, p. 973-986, 2016.

D'AOUST, J. Y.; GERBER, N. N. **Isolation and purification of prodigiosin from *Vibrio psychroerythrus*.** J. Bacteriol., v. 118, p. 756-757, 1974.

DALBEY, R. E.; KUHN, A. **Protein traffic in Gram-negative bacteria–how exported and secreted proteins find their way.** FEMS Microbiol. Rev., v. 36, n. 1023, 2012.

DAVISON, J. **Plant beneficial bacteria.** Bio/Technology, v. 6, p. 282–286, 1988.

DE TONI, J. B.; TREVISAN, V. **Schizomycetaceae Naeg**. In: Saccardo (Editor), Sylloge fungorum omnium hujusque cognitorum, v. 8, p. 923-1087, 1889.

DEMAIN, A. L.; FANG, A. **The natural functions of secondary metabolites.** In: Advances in Biochemical Engineering/Biotechnology: History of Modern Biotechnology I, v. 69, p. 2–39, 2000.

DEVI, K. A.; PANDEY, P.; SHARMA, G. D. **Plant Growth-Promoting Endophyte *Serratia marcescens* AL2-16 Enhances the Growth of *Achyranthes aspera* L., a Medicinal Plant**. HAYATI J Biosci, p. 1-8, 2017.

DURAND, E.; CAMBILLAU, C.; CASCALES, E.; JOURNET, L. **VgrG, Tae, Tle, and beyond: the versatile arsenal of Type VI secretion effectors.** Trends Microbiol., v. 22, p. 498–507, 2014.

EKBLOM, R.; WOLF, J. B. W. **A field guide to whole-genome sequencing, assembly and annotation.** Evolutionary Applications, v. 7, n. 9, p. 1026–1042, 2014.

EL-BONDKLY, A. A. M.; EL-GENDY, M. M. M.; BASSYOUNI, R. H. **Overproduction and biological activity of prodigiosin-like pigments from**

recombinant fusant of endophytic marine *Streptomyces* species. Antonie van Leeuwenhoek, v. 102, p. 719-734, 2012.

ENEBAK, S. A.; CAREY, W. A. **Evidence for Induced Systemic Protection to Fusiform Rust in Loblolly Pine by Plant Growth-Promoting Rhizobacteria.** v. 84, n. 3, p. 306–308, 2000.

ENGLISH, G.; TRUNK, K.; RAO, V. A.; SRIKANNATHASAN, V.; HUNTER, W. N.; COULTHURST, S. J. **New secreted toxins and immunity proteins encoded within the Type VI secretion system gene cluster of *Serratia marcescens*.** Mol. Microbiol., v. 86, p. 921–936, 2012.

EWING, W. H.; DAVIS, B.R.; FIFE, M.A.; LESSEL, E.F. **Biochemical characterization of *Serratia liquefaciens* (Grimes and Hennerty) Bascomb et al. (formerly *Enterobacter liquefaciens*) and *Serratia rubidaea* (Stapp) comb. nov. and designation of type and neotype strains.** Int. J. Syst. Bacteriol., v. 23, p. 217-225, 1973.

FAKHRAI-RAD, H.; POURMAND, N.; RONAGHI, M. **Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms.** Hum. Mutat. v. 19, p. 479–485, 2002.

FEHÉR, D.; BARLOW, R. S.; LORENZO, P. S.; HEMSCHEIDT, T. K. **A 2-Substituted Prodiginine, 2-(p-Hydroxybenzyl)prodigiosin, from *Pseudoalteromonas rubra*.** Journal of Natural Products, v. 71, n. 11, p. 1970-1972, 2008.

FERREIRA, L. C.; VIANA, M. V. C.; ROBERTS, D. P.; MAUL, J.; AZEVEDO, V. A. C.; SOUZA, J. T. DE. **Phylogenomics and Inconsistences in the Taxonomy of Serratia marcescens: Data from Whole Genomes.** 50º Brazilian Congress of Plant Pathology, 2017.

FINKEL, O. M.; CASTRILLO, G.; PAREDES, S. H.; GONZÁLEZ I. S.; DANGL, J. L. **Understanding and exploiting plant beneficial microbes.** Current Opinion in Plant Biology, v. 38, p. 155-163, 2017.

FLICEK, P.; BIRNEY E. **Sense from sequence reads: methods for alignment and assembly.** Nature Methods., v. 6, 2009.

FLYG, C.; XANTHOPOULOS, K. G. **Insect pathogenic properties of *Serratia marcescens*. Passive and active resistance to insect immunity studied with protease-deficient and phage-resistant mutantes**. Journal of General Microbiology, n. 1983, p. 453–464, 2017.

FRITSCH, M. J.; TRUNK, K.; DINIZ, J. A.; GUO, M.; TROST, M.; COULTHURST, S. J. **Proteomic identification of novel secreted antibacterial toxins of the *Serratia marcescens* type vi secretion system.** Molecular & Cellular Proteomics, v. 12, n. 10, p. 2735–2749, 2013.

FURNKRANZ, M.; ADAM, E.; MÜLLER, H.; GRUBE, M.; HUSS, H.; WINKLER, J.; BERG, G. **Promotion of growth, health and stress tolerance of Styrian oil pumpkins by bacterial endophytes.** Eur. J. Plant Pathol., v. 134, p. 509–519, 2012.

GAGNÉ, S.; RICHARD, C.; ROUSEAU, H.; ANTOUN, H. **Xylem-residing bacteria in alfalfa roots.** Can. J. Microbiol., v. 33, p. 996–1000, 1987.

GALAN, J. E.; LARA-TEJERO, M.; MARLOVITS, T. C.; WAGNER, S. **Bacterial type III secretion systems: Specialized nanomachines for protein delivery into target cells.** Annu. Rev. Microbiol., v. 68, p. 415–438, 2014.

GALPERIN, M. Y.; KOLKER, E. **New metrics for comparative genomics.** Curr Opin Biotechnol. v. 17, n. 5, p. 440-447, 2006.

GARCÍA-FRAILE, P.; CHUDÍČKOVÁ, M.; BENADA, O.; PIKULA, J.; KOLAŘÍK, M. ***Serratia myotis*** **sp. nov. and** ***Serratia vespertilionis*** **sp. nov., isolated from bats hibernating in caves.** International Journal of Systematic and Evolutionary Microbiology, v. 65, n. 1, p. 90- 94, 2015.

GAVINI, F.; FERRAGUT, C.; IZARD, D.; TRINEL, P. A.; LECLERC, H.; LEFEBVRE, B.; MOSSEL, D. A. A. ***Serratia fonticola,*** **a new species from water.** International Journal of Systematic and Evolutionary Microbiology, v. 29, n. 2, p. 92-101, 1979.

GEIGER, A.; FARDEAU, M.-L.; FALSNE, E.; OLLIVIER, B.; CUNY, G. **Serratia glossinae sp. nov., isolated from the midgut of the tsetse fly Glossina palpalis gambiensis.** Int. J. Syst. Evol. Microbiol., v. 60, p. 1261-1265, 2010.

GERBER, N. N.; GAUTHIER, M. J. **New Prodigiosin-Like Pigment from** ***Alteromonas rubra*****.** Applied and Environmental Microbiology, p. 1176-1179, 1979.

GERC, A. J.; SONG, L.; CHALLIS, G. L.; STANLEY-WALL1, N. R.; COULTHURST, S. J. **The Insect Pathogen Serratia marcescens Db10 Uses a Hybrid Non-Ribosomal Peptide Synthetase-Polyketide Synthase to Produce the Antibiotic Althiomycin.** PLOS ONE, v. 7, n. 9, p. 1–13, 2012.

GERLACH, R. G.; HENSEL, M. **Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens.** Int. J. Med. Microbiol., v. 297, p. 401–415, 2007.

GHANDI, N. M.; PATELL, J. R.; GHANDI, J.; DE SOUZA, N. J.; KOHL, H. **Prodigiosin metabolites of a marine** ***Pseudomonas*** **species.** Mar. Biol., v. 34, p. 223-227, 1976.

GIRI, A. V.; ANANDKUMAR, N.; MUTHUKUMARAN, G.; PENNATHUR, G. **A novel medium for the enhanced cell growth and production of prodigiosin from Serratia marcescens isolated from soil.** BMC microbiology, v. 4, n. 1, p. 1, 2004.

GOODWIN, S. MCPHERSON, J. D.; MCCOMBIE, W. R. **Coming of age: ten years of next-generation sequencing technologies.** Nature Reviews Genetics, v. 17, p. 333–351, 2016.

GREEN, E. R.; MECSAS, J. **Bacterial Secretion Systems – An overview.** Microbiology Spectrum, v. 4, n. 1, 2016.

GRIMONT, et al. ***Serratia proteamaculans*** **subsp.** ***quinovora*****: validation list no. 10.** Int. J. Syst. Bacteriol., v. 33, p. 438-440, 1983.

GRIMONT, P. A. D.; GRIMONT, F. **Genus VIII. Sermtia.** In: Krieg NR, Holt JG (eds) Bergey's Manual of systematic bacteriology, Baltimore, Williams and Wilkins, v. 1, p. 477-484, 1984.

GRIMONT, P. A. D.; GRIMONT, F.; IRINO, K**. Biochemical characterization of Serratia liquefaciens sensu stricto, Serratia proteamaculans, and *Serratia grimesii* sp. nov.** Curr. Microbiol., v. 7, p.69-74, 1982.

GRIMONT, P. A. D.; GRIMONT, F.; RICHARD, C.; DAVIS, B. R., STEIGERWALT, A. G.; BRENNER, D. J. **Deoxyribonucleic acid relatedness between Serratia plymuthica and other *Serratia* species, with a description of *Serratia odorifera* sp. nov. (Type strain: ICPB 3995).** Int. J. Syst. Bacteriol., v. 28, p. 453-463, 1978a.

GRIMONT, P. A. D.; GRIMONT, F.; STARR, M. P. ***Serratia ficaria* sp. nov., a bacterial species associated with Smyrna figs and the fig wasp *Blastophaga psenes*.** Curr. Microbiol., v. 2, p. 277-282, 1979.

GRIMONT, P. A. D.; GRIMONT, F.; STARR, M. P. ***Serratia proteamaculans* (Paine and Stansfield) comb. nov., a senior heterotypic synonym of *Serratia liquefaciens* (Grimes and Hennerty) Bascomb et al.** Int. J. Syst. Bacteriol., v. 28, p. 503-510, 1978b.

GRIMONT, P. A. D.; JACKSON, T. A.; AGERON, E.; NOONAN, M. J. ***Serratia entomophila* sp. nov., associated with amber disease in the New Zealand grass grub Costelytra zealandica.** Int. J. Syst. Bacteriol., v. 38, p. 1-6, 1988.

GROHMANN, E.; MUTH, G.; ESPINOSA, M. **Conjugative plasmid transfer in Gram-positive bacteria.** Microb. Mol. Biol. Rev., v. 67, p. 277–301, 2003.

GUZMAN, A. A. S.; PREDICALA, R. Z.; BERNARDO, E. B.; NEILAN B. A.; ELARDO, S. P.; MANGALINDAN, G. C.; TASDEMIR. D.; IRELAND, C. M.; BARRAQUIO, W. L.; CONCEPCION, G. P. ***Pseudovibrio denitrificans* strain Z143-1, a heptylprodigiosin-producing bacterium isolated from a Philippine tunicate.** FEMS Microbiol Lett., v. 277, n. 2, p. 188-96, 2007.

GYANESHWAR, P.; JAMES, E. K.; MATHAN, N.; REDDY, P. M.; REINHOLD-HUREK, B.; LADHA, J. K**. Endophytic Colonization of Rice by a Diazotrophic Strain of *Serratia marcescens*.** Journal of Bacteriology, v. 183, n. 8, , p. 2634–2645, 2001.

HAAS, D.; KEEL, C. **Regulation of antibiotic production in root-colonizing *Pseudomonas* spp: and relevance for biological control of plant disease.** Ann. Rev. Phytopathol., v. 41, p. 117–153, 2003.

HALLMANN, J.; QUADT-HALLMANN, A.; MAHAFFEE, W. F.; KLOEPPER, J. W. **Bacterial endophytes in agricultural crops.** Canadian Journal of Microbiology, v. 43, p. 895-914, 1997.

HARDOIM, P. R.; VAN OVERBEEK, L. S.; ELSAS, J. D. V. **Properties of bacterial endophytes and their proposed role in plant growth.** Trends Microbiol., v. 16, p. 463–471, 2008.

HARRIS, A. K. P.; et al. **The *Serratia* gene cluster encoding biosynthesis of the red antibiotic, prodigiosin, shows species- and strain-dependent genome context variation.** Microbiology, v. 150, n. 11, p. 3547–3560, 2004.

HEATHER, J. M; CHAIN, B. **The sequence of sequencers : The history of sequencing DNA.** Genomics, v. 107, n. 1, p. 1–8, 2016.

HENRIQUES, I.; RAMOS, R. T. J.; BARAÚNA, R. A.; DE SÁ, P. H.; ALMEIDA, D. M.; CARNEIRO, A. R.; EGAS, C. **Draft genome sequence of *Serratia fonticola* UTAD54, a carbapenem-resistant strain isolated from drinking water.** Genome announcements, v. 1, n. 6, p. e00970-13, 2013.

HOOD, R. D.; et al. **A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria.** Cell Host Microbe., v. 7, p. 25–37, (doi:10.1016/j.chom.2009. 12.007), 2010.

HOUDT, R. V.; IZQUIERDO, J. A.; AERTSEN, A.; MASSCHELEIN, J.; LAVIGNE, R.; TAGHAVI, S.; MICHIELS, C.; VAN DER LELIE, D. **Genome sequence of *Serratia plymuthica* RVH1, isolated from a raw vegetable processing line.** Genome Announc., 2014.

HOUDT, R. V.; MOONS, P.; JANSEN, A.; VANOIRBEEK, K.; MICHIELS, C. W. **Genotypic and phenotypic characterization of a biofilm-forming *Serratia plymuthica* isolate from a raw vegetable processing line.** FEMS Microbiol Lett., v. 246, p. 265–272, 2005.

HUANG, J. **Ultrastructure of bac- terial penetration in plants.** Annu. Rev. Phytopathol., v. 24, p. 141–157, 1986.

HUECK, C. J. **Type III protein secretion systems in bacterial pathogens of animals and plants.** Microbiol. Mol. Biol. Rev., v. 62, n. 2, p. 379-433, 1998.

IGUCHI, A.; NAGAYA, Y.; PRADEL, E.; OOKA, T.; et al**. Genome evolution and plasticity of *Serratia marcescens*, an important multidrug-resistant nosocomial pathogen.** Genome Biology and Evolution, v. 6, n. 8, p. 2096–2110, 2014.

JU, J. et al. **Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators**. Proc. Natl Acad. Sci., v.103, p. 19635–19640, 2006.

JUNG, B. K.; KHAN, A. R.; HONG, S.; PARK, G.; PARK, Y.; PARK, C. E.; JEON, H.; LEE, S.; SHIN, J. **Genomic and phenotypic analyses of *Serratia fonticola* strain GS2 : a rhizobacterium isolated from sesame rhizosphere that promotes plant growth and produces N -acyl homoserine lactone.** Journal of Biotechnology, v. 241, p. 158–162, 2017.

JUNG, B. K.; KIM, S.D.; KHAN, A. R.; LIM, J. H.; AN, C. H.; KIM, Y. H.; SONG, J. H.; HONG, S. J.; SHIN, J. H. **Rhizobacterial communities and red pepper (*Capsicum annum*) yield under different cropping systems.** Int. J. Agric. Biol., v. 17, p. 734–740, 2015.

KALBE, C.; MARTEN, P.; BERG, G. **Strains of the genus Serratia as beneficial rhizobacteria of oilseed rape with antifungal properties.** Microbiol. Res., v. 151, p. 433-439, 1996.

KAMENSKY, M.; OVADIS, M.; CHET, I.; CHERNIN, L. **Soil-borne strain IC14 of *Serratia plymuthica* with multiple mechanisms of antifungal activity provides biocontrol of *Botrytis cinerea* and *Sclerotinia sclerotiorum* diseases.** Soil Biology and Biochemistry, v. 35, n. 2, p. 323-331, 2003.

KAMPFER, P.; GLAESER, S. P. ***Serratia aquatilis* sp. nov., isolated from drinking water systems.** Int. J. Syst. Evol. Microbiol., v. 66, p. 407-413, 2016.

KAWAUCHI, K; SHIBUTANI, K; YAGISAWA, H; KAMATA,H.; NAKATSUJI, S.; ANZAI, H.; YOKOYAMA, Y.; IKEGAMI, Y.; MORIYAMA, Y.; HIRATA, H. A **possible immunosuppressant, cycloprodigiosin hydrochloride, obtained from *Pseudoalteromonas denitrificans*.** Biochemical and biophysical research communications, v. 237, p. 543–547, 1997.

KHAN, A. R.; PERVEZ, M. T.; BABAR, M. E.; NAVEED, V.; SHOAIB, M. A **Comprehensive Study of De Novo Genome Assemblers: Current Challenges and Future Prospective.** Evolutionary Bioinformatics, v. 14, p. 1-8, 2018.

KIM, D.; LEE, J. S, PARK, Y. K.; et al. **Biosynthesis of antibiotic prodiginines in the marine bacterium *Hahella chejuensis* KCTC 2396.** J Appl Microbiol., v. 102, p. 937-944, 2007.

KIM, Y. C.; LEVEAU, J.; MCSPADDEN GARDENER, B. B.; PIERSON, E. A, PIERSON III, L. S.; RYU, C. M. **The multifactorial basis for plant health promotion by plant-associated bacteria.** Appl Environ Microbiol v. 77, p. 1548-55, 2011.

KISHORE, G K; PANDE, S; PODILE, A R. **Chitin-supplemented foliar application of *Serratia marcescens* gps 5 improves control of late leaf spot disease of groundnut by activating defence-related enzymes.** J. Phytopathology v. 173, p. 169–173, 2005.

KLOEPPER, J. W.; TUZUN, S.; LIU, L.; WEI, G. **Plant growth-promo- ting rhizobacteria as inducers of systemic disease resistance.** In: Lumsden, R.D., Waughn, J.L. (Eds.), Pest Management: Biolo- gically Based Technologies. American Chemical Society Books, Washington, DC, p. 156-165, 1993.

KOBAYASHI, D. Y.; GUGLIELMONI, M.; CLARKE, B. B. **Isolation of the chitinolytic bacteria *Xanthomonas maltophilia* and *Serratia marcescens* as biological control agents for summer patch disease of turfgrass.** Soil Biol. Biochem., v. 27, n.. 1 I, p. 1479-1487, 1995.

KONSTANTINIDIS, K. T.; TIEDJE, J. M. **Genomic insights that advance the species definition for prokaryotes.** PNAS, v. 102, n. 7, p. 2567-2572, 2005b.

KONSTANTINIDIS, K. T.; TIEDJE, J. M. **Towards a Genome-Based Taxonomy for Prokaryotes.** Journal of Bacteriology, v. 187, n. 18, p. 6258-6264, 2005a.

LAMINE, B. M.; LAMINE, B. M. **Optimisation of the Chitinase Production by *Serratia Marcescens* DSM 30121T and Biological Control of Locusts.** Journal of Biotechnology & Biomaterials, v. 2, n. 3, p. 3–7, 2012.

LAVANIA, M.; CHAUHAN, P. S.; CHAUHAN, S. V. S.; SINGH, H. B.; NAUTIYAL, C. S.. **Induction of Plant Defense Enzymes and Phenolics by Treatment With Plant Growth – Promoting Rhizobacteria *Serratia marcescens* NBRI1213.** Current Microbiology, v. 52, p. 363–368, 2006.

LAVANIA, M.; NAUTIYAL, C. S. **Solubilization of tricalcium phosphate by temperature and salt tolerant *Serratia marcescens* NBRI1213 isolated from alkaline soils.** African Journal of Microbiology Research, v. 7, n. 34, p. 4403-4413, 2013.

LEE, J. S.; KIM, Y. S.; PARK, S.; et al. **Exceptional production of both prodigiosin and cycloprodigiosin as major metabolic constituents by a novel marine bacterium, *Zooshikella rubidus* S1-1.** Appl Environ Microbiol., v. 77, p. 4967-4973, 2011.

LEHMAN, K. B.; NEUMANN, R. **Atlas und Grundriss der Bakeriologie und Lehrbuch der speziellen bakteriologischen Diagnositk**. 1st Ed., J.F. Lehmann, Munchen, 1896.

LEVENFORS, J. J.; HEDMAN, R.; THANING, C.; GERHARDSON, B.; WELCH, C. J. **Broad-spectrum antifungal metabolites produced by the soil bacterium *Serratia plymuthica* A 15.** Soil Biol Biochem., v. 36, p. 677–685, 2004.

LI, P.; KWOK, A. H. Y.; JIANG, J.; RAN, T.; XU, D.; WANG, W.; LEUNG, F. C. **Comparative Genome analyses of *Serratia marcescens* FS14 reveals Its high antagonistic potential.** PLoS One, v. 10, n. 4, 2015.

LI, R.; ZHU, H.; RUAN, J.; QIAN, W.; FANG, X.; et al. ***De novo* assembly of human genomes with massively parallel short read sequencing.** Genome Res, v. 20, p. 265–272, 2010.

LI, Z.; CHEN, Y.; MU, D. **Comparison of the two major classes of assembly algorithms: overlap–layout–consensus and de-Bruijn-graph.** Brief Func Genom., v. 11, p. 25–37, 2012.

LIM, Y. L.; YONG, D.; EE, R.; KRISHNAN, T.; TEE, K. K.; YIN, W. F.; CHAN, K. G. **Complete genome sequence of *Serratia fonticola* DSM 4576 T, a potential plant growth promoting bacterium.** Journal of Biotechnology, v. 214, p. 43-44, 2015.

LIU, L.; KLOEPPER, J. W.; TUZUN, S. **Induction of systemic resistance in cucumber against bacterial angular leaf spot by plant growth-promoting rhizobacteria.** Phytopatholohy, v. 85, p. 843-847, 1995a.

LIU, L.; KLOEPPER, J. W.; TUZUN, S. **Induction of systemic resistance in cucumber against Fusarium wilt by plant growth-promoting rhizobacteria.** Phytopatholohy, v. 85, p. 695-698, 1995b.

LUGTENBERG, B. J. J.; DEKKERS, L.; BLOEMBERG, G. V. **Molecular determinants of rhizosphere colonization by *Pseudomonas***. Annu. Rev. Phytopathol., v. 39, p. 461–90, 2001.

LUO, C.; TSEMENTZI, D.; KYRPIDES, N.; READ, T.; KONSTANTINIDIS, K.T. **Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample**. PLoS One, v. 7, 2012.

MACINTYRE, D. L.; MIYATA, S. T.; KITAOKA, M.; PUKATZKI1, S. **The *Vibrio cholerae* type VI secretion system displays antimicrobial properties.** Proc. Natl. Acad. Sci. U.S.A., v. 107, p. 19520–19524, 2010.

MANZANO-MARÍN, A.; LAMELAS, A.; MOYA, A.; LATORRE, A. **Comparative Genomics of *Serratia* spp.: Two Paths towards Endosymbiotic Life.** PLoS ONE, v. 7, n. 10, 2012.

MARTINS, A. M. et al. **Sequenciamento de DNA, montagem de novo do genoma e desenvolvimento de marcadores microssatélites, indels e SNPs para uso em análise genética de *Brachiaria ruziziensis*.** Brasília: UnB, p. 1–198, 2013.

MASSCHELEIN, J.; MATTHEUS, W.; GAO, L. J.; MOONS, P.; HOUDT, R. V.; UYTTERHOEVEN, B.; LAMBERIGTS, C.; LESCRINIER, E.; ROZENSKI, J.; HERDEWIJN, P.; AERTSEN, A.; MICHIELS, C.; LAVIGNE, R. **A PKS/NRPS/FAS hybrid gene cluster from *Serratia plymuthica* RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics.** PLoS One, v. 8, 2013.

MCINROY, J. A.; KLOEPPER, J. W. **Studies on indigenous endophytic bacteria of sweet com and cotton.** In: O'Gara, F. a., Dowling, D. N., Boeston, B.: Molecular Ecology of Rhizosphere Microorganisms. VCH, Weinheim, p. 19-28, 1994.

MEDVEDEV, P.; GEORGIOU, K.; MYERS, E. W.; et al. **'Computability and Equivalence of Models for Sequence Assembly'.** Workshop on Algorithms in Bioinformatics, Philadelphia, PA: Springer; 2007.

MILLER, J. R.; KOREN, S.; SUTTON, G. **Assembly algorithms for next-generation sequencing data.** Genomics, v. 95, p. 315–327, 2010.

MITTER, B.; PETRIC, A.; SHIN, M. W.; CHAIN, P. S. G.; HAUBERG-LOTTE, L.; REINHOLD-HUREK, B.; NOWAK, J.; SESSITSCH, A. **Comparative genome analysis of *Burkholderia phytofirmans* PsJN reveals a wide spectrum of endophytic lifestyles based on interaction strategies with host plants**. Frontiers in Plant Science, v. 4, p. 1–15, 2013.

MONREAL, J.; REESE, E. T. **The chitinase of *Serratia marcescens*.** Can. J. Microbiol., v. 15, p. 689-696, 1969.

MONTANER, B.; PEREZ-THOMAS, R. **Prodigiosin induced apoptosis in human colon cancer cells.** Life Sci., v. 68, p. 2025-2036, 2001.

MORAN, N. A.; RUSSELL, J. A., KOGA, R.; FUKATSU, T. **Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects.** Appl. Environ. Microbiol., v. 71, p. 3302-3310, 2005.

MURDOCH, S. L.; TRUNK, K.; ENGLISH, G.; FRITSCH, M. J.; POURKARIMI, E.; COULTHURST, S. J. **The opportunistic pathogen *Serratia marcescens* utilizes type VI secretion to target bacterial competitors**. J. Bacteriol., v. 193, p. 6057–6069, 2011.

MYERS, E. W. **Towards Simplifying and Accurately Formulating Fragment Assembly**. J. Comp. Bio., v. 2, p. 275–90, 1995.

NAGARAJAN, N.; POP, M. **Sequence assembly demystified.** Nature Reviews Genetics., v. 14, p. 157–167, 2013.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. GenBank and WGS Statistics. Disponível em: < https://www.ncbi.nlm.nih.gov/genbank/statistics/>

NEUPANE, S.; FINLAY, R. D; KYRPIDES, N. C.; et al. **Complete genome sequence of the plant-associated *Serratia plymuthica* strain AS13.** Standards in Genomic Sciences, v. 7, p. 22–30, 2012b.

NEUPANE, S.; HOGBERG, N.; ALSTROM, S.; et al. **Complete genome sequence of the rapeseed plant-growth promoting *Serratia plymuthica* strain AS9.** Standards in Genomic Sciences, v. 6, p. 54–62, 2012a.

NIELSEN, H. **Predicting Secretory Proteins with SignalP.** In Kihara, D (ed):Protein Function Prediction (Methods in Molecular Biology vol. 1611) pp. 59-73, Springer 2017.

ORDENTLICH, A.; ELAD, Y.; CHET, I. **The role of the chitinase of *Serratia marcescens* in Biocontrol of *Sclerotium rolfsii*.** Phytopatology, v. 78, p. 84-88, 1988.

PAINE, S. G.; STANFIELD, H. **Studies in bacteriosis. III. a bacterial leaf-spot disease of *Protea cynaroides* exhibiting a host reaction of possibly bacteriolytic nature.** Ann. Appl. Biol., v. 6, p. 27-39, 1919.

PARK, H.; LEE S. G.; KIM, T. K.; et al. **Selection of extraction solvent and temperature effect on stability of the algicidal agente prodigiosin.** Biotechnol Bioprocess Eng., v. 17, p. 1232-1237, 2012.

PARKER, W. L.; RATHNUM, M. L.; WELLS, J. S.; TREJO, W. H.; PRINCIPE, P. A.; SYKES, R. B. **SQ 27,860, a simple carbapenemproduced by species of *Serratia* and *Erwinia*.** J Antibiot., v. 35, p. 653-660, 1982.

PATIL, C. D.; PATIL, S. V.; SALUNKE, B. K.; et al. **Prodigiosin produced by *Serratia marcescens* NMCC46 as a mosquito larvicidal agent against *Aedes aegypti* and *Anopheles stephensi*.** Parasitol Res., v. 109, p. 1179, (DOI:10.1007/s00436-011-2365-9), 2011.

POP, M. **Genome assembly reborn: recent computational challenges**. Briefings in Bioinformatics, v. 10, n. 4, p. 354–366, 2009.

QUAIL, M.A.; SMITH, M.; COUPLAND, P.; OTTO, T. D.; HARRIS, S. R.; CONNOR, T. R.; BERTONI, A.; SWERDLOW, H. P.; GU, Y. **A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and IlluminaMiSeq sequencers**. BMC Genomics, v. 13, 2012.

QUEIROZ, B. P. V. DE; MELO, I. S. DE. **Antagonism of *Serratia marcescens* towards *Phytophthora parasitica* and its effects in promoting the growth of citrus.** Brazilian Journal of Microbiology, v. 37, p. 448-450, 2006.

RAHUL, S.; CHANDRASHEKHAR, P.; HEMANT, B.; CHANDRAKANT, N.; LAXMIKANT,S.; SATISH, P. **Nematicidal activity of microbial pigment from *Serratia marcescens*.** Natural Product Research, v. 28, n. 17, 2014.

RAMAMOORTHY, V.; VISWANATHAN, R.; RAGUCHANDER, T; PRAKASAM, V.; SMAIYAPPAN, R. **Induction of systemic resistance by plant growth promoting rhizobacteria in crop plants against pests and diseases.** Crop. Prot., v. 20, p; 1-11, 2001.

RAPOPORT, H.; HOLDEN, H. G. **The synthesis of prodigiosin**. J Am Chem Soc., v. 84: p. 635- 42, 1962.

RAUPACH, G. S.; LIU, L.; MURPHY, J. F.; TUZUN, S.; KLOEPPER, J. W. **Induced systemic resistance in cucumber and tomato against cu- cumber mosaic cucumovirus using plant growth promoting rhizobateria (PGPR).** Plant Dis., v. 80, p. 891-894, 1996.

RICHARDSON, E. J.; WATSON, M. **The automatic annotation of bacterial genomes.** Briefings in Bioinformatics, v. 14, n. 1, p. 1–12. 2013.

ROMEIRO, R. S. **Controle biológico de doenças de plantas – Fundamentos**. Viçosa - MG: UFV, 296 p., 2007.

RONAGHI, M. **Pyrosequencing sheds light on DNA sequencing**. *Genome Res.* v. 11, p. 3–11, 2001.

RONAGHI, M.; KARAMOHAMED, S.; PETTERSSON, B.; UHLEN, M.; NYREN, P. **Real-time DNA sequencing using detection of pyrophosphate release.** Anal. Biochem, v. 242, p. 84–89, 1996.

RONAGHI, M.; UHLEN, M.; NYREN, P. **A sequencing method based on real-time pyrophosphate.** Science, v. 281, p. 363–365, 1998.

ROOS, I. M. M.; HATTINGH, M. J. **Scanning electron microscopy *Pseudomonas syringae* pv. *morsprunorum* on sweet cherry leaves.** Phytopathol Z., v. 108, p. 18–25, 1983.

ROSSELLÓ-MÓRA, R.; AMANN, R. **Past and future species definitions for Bacteria and Archaea.** Systematic and Applied Microbiology, v. 38, p. 209-216, 2015.

ROTHBERG, J. M.; LEAMON, J. H. **The development and impact of 454 sequencing.** Nature Biotechnology, v. 26, p. 1117–1124, 2008.

ROY, P.; AHMED, N. H.; GROVER, R. K. **Non-pigmented strain of *Serratia marcescens*: An unusual pathogen causing pulmonary infection in a patient with malignancy.** Journal of Clinical and Diagnostic Research: JCDR, v. 8, n. 6, 2014.

RUSSELL, A. B.; HOOD, R. D.; BUI, N. K.; LEROUX, M.; VOLLMER, W.; MOUGOUS, J. D. **Type VI secretion delivers bacteriolytic effectors to target cells.** Nature, v. 475, p. 343–347, (DOI:10.1038/nature10244), 2011

RYAN, R. P.; GERMAINE, K.; FRANKS, A.; RYAN, D. J.; DOWLING, D. N. **Bacterial endophytes: recent developments and applications.** FEMS Microbiol. Lett., v. 278, p. 1-9, 2008.

RYAZANTSEVA, I. N.; SAAKOV, V. S.; ANDREYEVA, I. N.; et al. **Response of pigmented *Serratia marcescens* to the illumination.** J. Photoch. Photobio. B., v. 106, p. 18-23, 2012.

SABRI, A.; LEROY, P.; HAUBRUGE, E.; HANCE, T.; FRERE, I.; DESTAIN, J.; THONART, P. **Isolation, pure culture and characterization of *Serratia symbiotica* sp. nov., the R-type of secondary endosymbiont of the black bean aphid Aphis fabae.** Int. J. Syst. Evol. Microbiol., v. 61, p. 2081-2088, 2011.

SAFNI, I.; CLEENWERCK, I.; DE VOS, P.; FEGAN, M.; SLY, L.; KAPPLER, U. **Polyphasic taxonomic revision of the *Ralstonia solanacearum* species complex: proposal to emend the descriptions of *Ralstonia solanacearum* and *Ralstonia syzygii* and reclassify current *R. syzygii* strains as *Ralstonia syzygii* subsp. *syzygii* subsp. nov., *R. solanacearum* phylotype IV strains as *Ralstonia syzygii* subsp. *indonesiensis* subsp. nov., banana blood disease bacterium strains as *Ralstonia syzygii* subsp. *celebesensis* subsp. nov. and *R. solanacearum* phylotype I and III strains as *Ralstonia pseudosolanacearum* sp. nov.** Int J Syst Evol Microbiol., v. 64, p. 3087-3103, 2014.

SALZBERG, S. L.; PHILLIPPY, A. M.; ZIMIN, A.; PUIU, D.; MAGOC, T.; KOREN S.; TREANGEN, T. J.; SCHATZ, M. C.; DELCHER, A. L.; ROBERTS, M.; et al. **GAGE: a critical evaluation of genome assemblies and assembly algorithms.** Genome Res., v. 22, n. 3, p. 557–567, 2012.

SANDHIYA, G. S.; SUGITHA, T. C. K.; BALACHANDAR, D.; KUMAR, K. **Endophytic colonization and in planta nitrogen fixation by a diazotrophic *Serratia* sp. in rice.** Indian Journal of Experimental Biology, v. 43, p. 802-807, 2005.

SANGER, F. **Nucleotide sequences in DNA.** Proc. R. Soc. Lond. B Biol. Sci. v. 191, p. 317–333, 1975.

SANGER, F.; COULSON, A. R. **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.** J. Mol. Biol., v. 94, p. 441–448, 1975.

SANTOYO, G.; MORENO-HAGELSIEB, G.; OROZCO-MOSQUEDAC, M. DEL C.; GLICK, B. R. **Plant growth-promoting bacterial endophytes.** Microbiological Research, v. 183, p. 92–99, 2016.

SCHWARZ, S.; WEST, T. E.; BOYER, F.; CHIANG, W. C.; CARL, M. A.; HOOD, R. D.; ROHMER, L.; TOLKER-NIELSEN, T.; SKERRETT, S. J.; MOUGOUS J. D. ***Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions.** PLoS Pathog., v6, p. e1001068, 2010.

SCOTT, R. I.; CHARD, J. M.; HOCART, M. J.; LENNARD, J. H., GRAHAM, D. C. **Penetration of potato tuber lenticels by bacteria in relation to biological control of blackleg disease.** Potato Res., v. 39, p. 333–334, 1996.

SEEMANN, T. **Prokka: rapid prokaryotic genome annotation.** Bioinformatics, v. 30, n. 14, p. 2068–2069, 2014.

SELVAKUMAR, G.; Mohan, M.; Kundu, S.; Gupta, A. D.; Joshi, P.; Nazim, S.; Gupta, H.S. **Cold tolerance and plant growth promotion potential of *Serratia marcescens* strain SRM (MTCC 8708) isolated from flowers of summer squash (*Cucurbita pepo*).** Letters in Applied Microbiology, v. 46, n. Mtcc 8708, p. 171–175, 2008.

SHISHIDO, M.; BREUIL, C.; CHANWAY, C. P. **Endophytic colonization of spruce by growth-promoting rhizobacteria.** FEMS Microbiol Ecol v. 29, p. 191-196, 1999.

SLATER, H.; CROW, M. A.; EVERSON, L.; et al. **Phosphate availability regulates biosynthesis of two antibiotics, prodigiosin and carbapenem, in *Serratia* via both quorum-sensing-dependent and-independent pathways.** Mol Microbiol v. 47, p. 303-320, 2003.

SOLÉ, M.; FRANCIA, A.; RIUS, N.; et al. **The role of pH in the glucose effect on prodigiosin production by non-proliferating cells of *Serratia marcescens*.** Lett Appl Microbiol v. 25, p. 81-84, 1997.

SOMEYA, N.; NAKAJIMA, M.; WATANABE, K.; HIBI, T.; AKUTSU, K. **Potential of *Serratia marcescens* strain B2 for biological control of rice sheath blight.** Biocontrol Science and Technology, v. 15, n. 1, p. 105–109, 2005.

SOTO-CERRATO, V.; LLAGOSTERA, E., MONTANER, B.; et al. **Mitochondria-mediated apoptosis operating irrespective of multidrug resistance in breast cancer cells by the anticâncer agent prodigiosin.** Biochem Pharmacol v. 68, p. 1345-1352, 2004.

STAPP, C. **Bacterium rubidaeum nov. spec.** Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. II, v. 102, p. 252-260, 1940.

STURZ, A. V.; CHRISTIE, B. R.; MATHESON, B. G.; ARSENAULT W. J.; BUCHANAN N. A. **Endophytic bacterial communities in the periderm of potato tubers and their potential to improve resistance to soil-borne plant pathogens**. Plant Pathol, v. l48, p. 360-9, 2000.

SURYAWANSHI, R. K.; PATIL, C. D.; BORASE, H. P.; NARKHEDE, C. P. SALUNKE, B. K.; PATIL, S. V. **Mosquito larvicidal and pupaecidal potential of prodigiosin from *Serratia marcescens* and understanding its mechanism of action.** Pesticide Biochemistry and Physiology, v. 123, p. 49–55, 2015.

TAGHAVI, S.; GARAFOLA, C.; MONCHY, S.; NEWMAN, L.; HOFFMAN, A.; WEYENS, N.; BARAC, T.; VANGRONSVELD, J.; VAN DER LELIE, D. **Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees.** Appl Environ Microbiol; v. 75, p. 748-757, 2009.

TSAO, S. W.; RUDD, B. A.; HE, X. G.; CHANG, C. J.; FLOSS, H. G. **Identification of a red pigment from Streptomyces coelicolor A3(2) as a mixture of prodigiosin derivatives**. J. Antibiot. (Tokyo), v. 38, p. 128–131, 1985.

VICENTE, C. S.; NASCIMENTO, F. X.; IKUYO, Y.; COCK, P. J.; MOTA, M.; HASEGAWA, K. **The genome and genetics of a high oxidative stress tolerant *Serratia***

sp. LCN16 isolated from the plant parasitic nematode *Bursaphelenchus xylophilus*. BMC Genomics, v. 17, n. 1, 1 p. , 2016.

WANG, K.; YAN, P. S.; CAO, L. X. **Chitinase from a novel strain of *Serratia marcescens* JPP1 for biocontrol of aflatoxin: Molecular characterization and production optimization using response surface methodology.** BioMed Research International, v. 2014, p.1-8, 2014.

WANG, L.; WANG, J.; JING, C. **Comparative genomic analysis reveals organization, function and evolution of ars genes in *Pantoea* spp.** Frontiers in Microbiology, v. 8, p. 1–12, 2017.

WANG, S. L.; WANG, C. Y.; YEN, Y. H.; et al. **Enhanced production of insecticidal prodigiosin from *Serratia marcescens* TKU011 in media containing squid pen.** Process Biochem v. 47, p. 1684-1690, 2012.

WARD, J. E.; AKIYOSHI, D. E.; REGIER, D.; DATTA, A.; GORDON, M. P.; NESTER, E. W. **Characterization of the virB operon from an *Agrobacterium tumefaciens* Ti plasmid**. J. Biol. Chem., v. 263, p. 5804–5814, 1988.

WILLIAMS, R. P.; GOTT, C. L.; QADRI, S. M. H. **Induction of pigmentation in nonproliferating cells of *Serratia marcescens* by addition of single amino acids.** J Bacteriol, v. 106, p. 444-448, 1971.

WILSON, D. **Endophyte: the evolution of a term, and clarification of its use and definition.** OIKOS, p. 274-276, 1995.

WITNEY, F. R.; FAILIA, M. L.; WEINBERG E. D**. Phosphate inhibition of secondary metabolism in *Serratia marcescens*.** Appl Environ Microbiol v. 33, p. 1042-1046, 1977.

WOLD, B.; MYERS, R.M. **Sequence census methods for functional genomics**. Nat Methods, v. 5, n. 1, p. 19-21, 2008.

YANDELL, M.; ENCE, D. **A beginner's guide to eukaryotic genome annotation.** Nature Reviews Genetics, v. 13, p. 329–342, 2012.

ZAHEER, A.; MIRZA, B. S.; MCLEAN, J. E.; YASMIN, S.; SHAH, T. M.; MALIK, K. A.; MIRZA, M. S. **Association of plant growth-promoting *Serratia* spp. with the root nodules of chickpea.** Research in Microbiology. v. 6, n. 167, p. 510–520, 2016.

ZAREI, M.; AMINZADEH, S.; ZOLGHARNEIN, H.; SAFAHIEH, A.; DALIRI, M.; NOGHABI1, K. A.; GHOROGHI, A.; MOTALLEBI, A. **Characterization of a chitinase with antifungal activity from a native *Serratia* marcescens B4A.** Brazilian Journal of Microbiology, v. 42, n. 3, p. 1017–1029, 2011.

ZHANG, C.-X.; YANG, S.-Y.; XU, M.-X.; SUN, J.; LIU, H.; LIU, J.-R.; LIU, H.; KAN, F.; SUN, J.; LAI, R.; ZHANG, K.-Y. ***Serratia nematodiphila* sp. nov., associated symbiotically with the entomopathogenic nematode *Heterorhabditidoides chongmingensis* (*Rhabditida: Rhabditidae*).** Int. J. Syst. Evol. Microbiol., v. 59, p. 1603-1608, 2009.

ZHANG, F.; DASHTI, N.; HYNES, R. K.; SMITH, D. L. **Plant growth promoting rhizobacteria and soybean [*Glycine max* (L.) Merr.] nodulation and nitrogen fixation at suboptimal root zone temperature.** Annals of Botany, v. 77, n. 5, p. 453-459, 1996.

ZHANG, W.; CHEN, J.; YANG, Y.; TANG, Y.; SHANG, J.; SHEN, B. **A Practical Comparison of De Novo Genome Assembly Software Tools for Next-Generation Sequencing Technologies**. PLoS ONE, v. 6, n. 3, p. e17915, 2011.

SECOND PART – ARTICLES

ARTICLE 1 - Complete genome sequence of *Serratia marcescens* strain N4-5, a biocontrol agent of soil-borne plant pathogens

Submitted to Standards in Genomic Sciences

Extended Genome Reports

Title

Complete genome sequence of *Serratia marcescens* strain N4-5, a biocontrol agent of soil-borne plant pathogens

Authors

Larissa Carvalho Ferreira[1], Jude E. Maul[2], Marcus Vinicius Canário Viana[3], Thiago Jesus de Sousa[3], Vasco Ariston de Carvalho Azevedo[3], Daniel P. Roberts[2], Jorge Teodoro de Souza[1]

Institutional Affiliation

[1]Plant Pathology Department, Federal University of Lavras, Lavras, Brazil

[2]USDA-Agricultural Research Service, Sustainable Agricultural Systems Laboratory, Beltsville, MD, USA

[3]Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

Corresponding author

Jorge Teodoro de Souza[1]

Abstract

Serratia marcescens are gram-negative bacteria found in several environmental niches, including plant rhizosphere and hospitals. Here we present the genome of *Serratia marcescens* strain N4-5 (=NRRL B-65519) isolated from New Jersey (USA) soil. N4-5 is a non-pathogen powerful biological control agent as its use as seed treatment is a successful method to control seed and seedling disease of cucurbits caused by the soil-borne plant pathogen *Pythium ultimum*. This strain represents the third strain of this species with plant-beneficial properties with a complete genome sequence. The genome size of *S. marcescens* N4-5 is 5,074,473 bp (664-fold coverage) and contains 4,840 protein coding genes, 21 RNA genes, and an average G+C content 59.7 %. We present the genome and discuss the presence of genes of interest for direct and indirect biocontrol acitivity against plant pathogens, such as production of prodiogiosin, chitinases and siderophores. Our genome assembly also uncovered an artifact present in other 48 *Serratia* spp. complete genomes deposited in public databases. This newly assembled genome artifact-free will

improve the quality of new genome assemblies and enhance the understanding of plant-beneficial bacteria.

Keywords

Enterobacteriaceae, NGS, gram-negative, non-sporulating, agriculture, plant-associated, biological control, plant growth-promoting rhizobacteria

Introduction

Soil-borne plant pathogens cause diseases resulting in major reductions in crop yields [1]. These diseases are typically controlled in conventional crop production systems with strategies that include chemical pesticides [2,3]. Biologically based methods, such as the use of microbial biological control agents, are being developed to control these soil-borne pathogens due to problems associated with the availability and effectiveness of chemical pesticides and concerns regarding the impact of these chemical pesticides on the environment and human health [4-6]. Biological control agents are increasingly being developed and applied in organic crop production systems where the use of chemical pesticides is prohibited and disease control strategies are limited. These microbial biological control agents control disease via several mechanisms including predation where the biological control agent produces an assortment of enzymes such as chitinase, protease, and glucanase that degrade pathogen cell wall and other cellular components [7-11]. Biological control agents can also produce antibiotics and other inhibitory molecules that kill or slow growth of the pathogen, and can compete with the pathogen for resources such as nutrients and space. Finally, certain biological control agents have been shown to associate with plants and induce defense responses that protect the plant from diseases [12,13].

The enteric bacterium *Serratia marcescens* is ubiquitous in the environment and has been detected in association with plants [14-16], animals [17-19] including humans in hospital settings [20,21], soil [22-25], water [26-28] and air [29]. Live cells and cell-free extracts of *S. marcescens* strains isolated from the environment have been shown to be effective in controlling certain soil-borne plant pathogens [30-37]. Here we report the characteristics of *S. marcescens* N4-5 (=NRRL B-65519), isolated from soil by baiting with chitin. We present the genome of strain N4-5 and provide insights into the mechanisms by which this strain associates with plants and controls diseases.

Organism Information

Classification and features

Strain N4-5 was obtained from New Jersey (USA) soil samples by baiting with chitin and was shown to be active as a biological control agent of summer patch disease in Kentucky bluegrass caused by *Magnaporthe poae* [30]. Strain N4-5 was later shown to be particularly effective in controlling the oomycete soil-borne plant pathogen *Pythium ultimum* when applied as a seed treatment [34]. Subsequent characterization showed that N4-5 was a gram-negative, motile, rod-shaped bacterium able to produce a red-colored pigment in culture (Fig. 1a), which was later identified as prodigiosin. This compound has anti-biotic properties and is partially responsible for activity of strain N4-5 against plant pathogens.

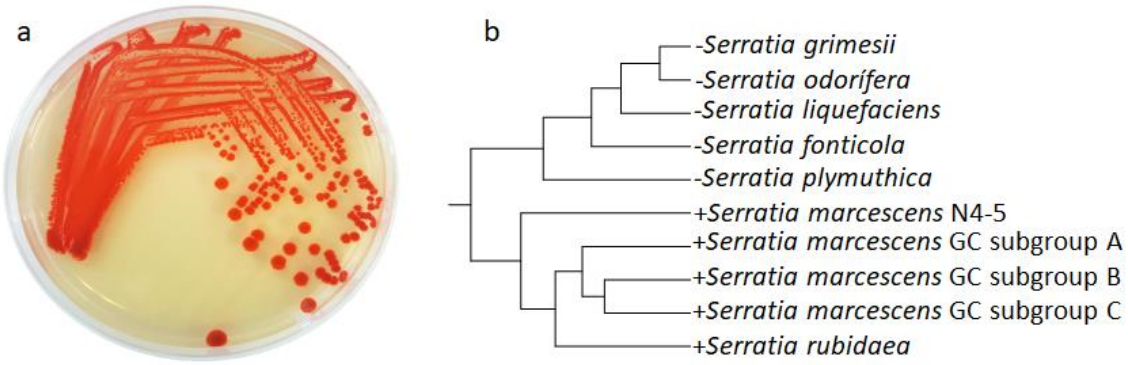

Fig. 1 Chemotaxonomic information of the strain N4-5. **(a)** *S. marcescens* N4-5 grown in solid media exhibiting red-colored colonies due to prodigiosin biosynthesis, a common feature in *Serratia* taxa. **(b)** Hierarchical clustering dendrogram derived from Phospholipid Fatty Acid (PLFA) profiles of the *S. marcescens* N4-5 and 9 other *Serratia* taxa. Profiles are based on fatty acids 12:0,10:0 3OH, 12:0 2OH, 12:1 3OH,12:0 3OH,14:0, 14:0 2OH, 16:0, 17:0 cyclo, 17:0, 16:0 3OH, 18:0, 19:0 cyclo w8c.

Fatty acid methyl ester (FAME) analysis (Fig. 1b) indicated that strain N4-5 was *Serratia marcescens*. *S. marcescens* strains are known as major prodigiosin producers, although not all strains of this species produce this compound. Research focus shifted to the use of cell-free natural product extracts of this strain due to concerns regarding the use of live cells of *S. marcescens* in agricultural applications [34]. Certain strains of *S. marcescens* are considered opportunistic human pathogens and can be problematic in hospital settings because of multidrug resistance [22,41-44]. The plant-beneficial properties of strain N4-5 and its natural product extracts have been reported in several publications [34-37]. Besides prodigiosin, it produces multiple surfactants including serrawettin W1, and chitinase and protease [34]. Other phenotypes of strain N4-5 are listed in Table 1.

Phylogenetic trees constructed with sequences of the 16S gene (Fig. 2) and with whole genome sequences (Additional file 1: Fig. S1) placed strain N4-5 within the *S. marcescens* clade. The consensus sequence obtained from the seven copies of the 16S rRNA gene of strain N4-5 is 99.87% identical to the 16S sequences of the type strain *S. marcescens* DSM 30121. The identity

of these seven copies of the 16S rRNA gene found in the genome of strain N4-5 vary from 99.7 to 100%. Further analyses, including Average Nucleotide/Amino Acid Identity (ANI/AAI) and digital DNA-DNA Hybridization (dDDH) calculated using JspeciesWS [38], Kostas Lab [39] and GGDC [40], confirmed the classification of strain N4-5 as *S. marcescens*. The values for ANI, AAI and dDDH were above the cutoff for species delineation, 95, 95 and 70% respectively, when compared with other *Serratia* genomes (Additional file 1: Table S1).

Table 1. Classification and general features of *Serratia marcescens* strain N4-5 [82]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS [83]
		Phylum *Proteobacteria*	TAS [84]
		Class *Gammaproteobacteria*	TAS [85,86]
		Order *Enterobacteriales*	TAS [87]
		Family *Enterobacteriaceae*	TAS [88-91]
		Genus *Serratia*	TAS [88,30,34]
		Species *Serratia marcescens* strain N4-5	TAS [88,83,84]
	Gram stain	Negative	TAS [30]
	Cell shape	Rod	NAS
	Motility	Motile	IDA
	Sporulation	Non-spore forming	NAS
	Temperature range	5-40°C	NAS
	Optimum temperature	37°C	NAS
	pH range; Optimum	5–9	NAS
	Carbon source	D glucose, D fructose, N-acetyl glucosamine, aspartate, citrate, gluconate, L-malate, mannitol, ribose	NAS
MIGS-6	Habitat	Soil	TAS [30]
MIGS-6.3	Salinity	Up to 10% NaCl (w/v)	IDA
MIGS-22	Oxygen requirement	Facultative anaerobic	NAS
MIGS-15	Biotic relationship	Free-living	TAS [30]
MIGS-14	Pathogenicity	Non-pathogen	NAS
MIGS-4	Geographic location	USA/New Jersey	NAS
MIGS-5	Sample collection	1996	TAS [30]
MIGS-4.1	Latitude	Not Reported	NAS
MIGS-4.2	Longitude	Not Reported	NAS
MIGS-4.4	Altitude	Not Reported	NAS

[a] Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [92]

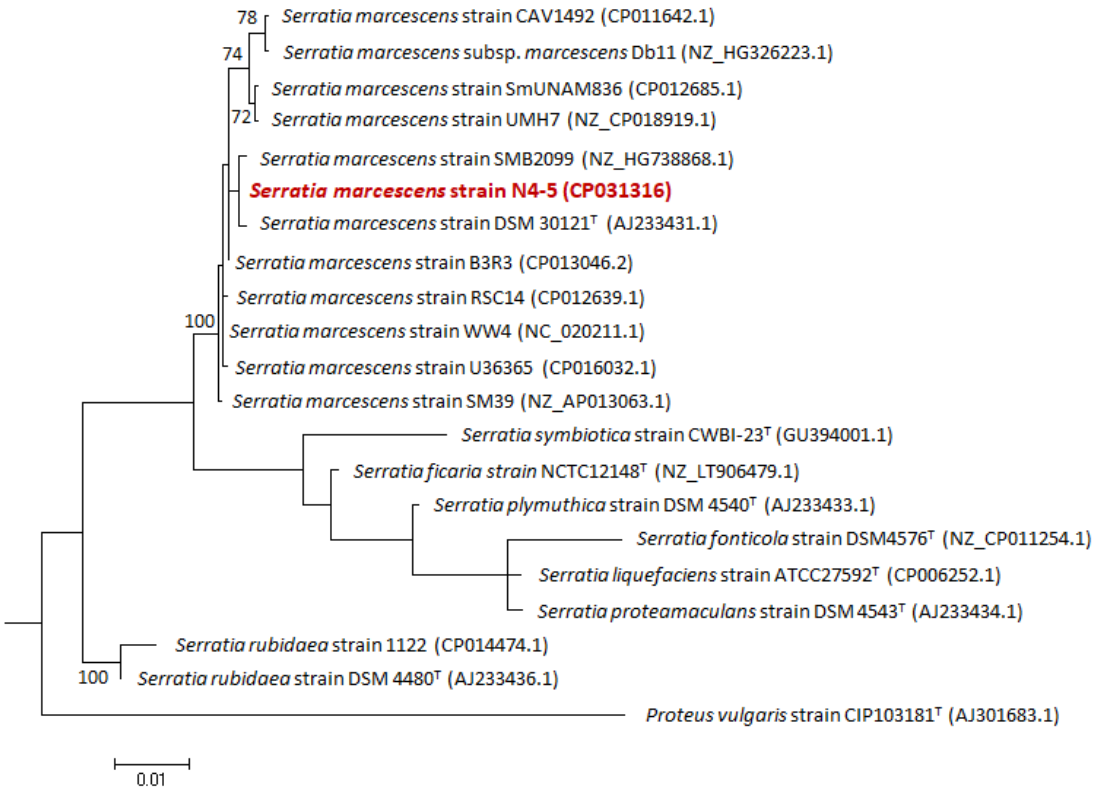

Fig. 2 Phylogenetic tree indicating the current taxonomic placement of strain N4-5. The phylogenetic tree was constructed in MEGA 6.06 [93] based on complete nucleotide sequences of the 16S gene (1479 bp) aligned in MAFFT [94]. The tree construction method was Maximum Likelihood with the Kimura 2-parameter model. Bootstrap values were calculated with 1,000 resamplings and values higher than 70% are shown on the appropriate branching points. The scale indicates the number of substitutions per site.

Chemotaxonomic information

The major components of the *S. marcescens* N4-5 fatty acid profile were $C_{16:0}$ (22.4%), $C_{17:0\ cyclo}$ (12.13%), $C_{10:0\ 3OH}$ (12.07%) and $C_{12:0\ 3OH}$ (5.15%) fatty acids. Minor fatty acid components were identified at less than 5%, most of which being common among previously identified species within the genera *Serratia*: $C_{12:0}$, $C_{12:0\ 2OH}$, $C_{14:0}$, $C_{14:0}$, $C_{18:0}$ and $C_{19:0\ cyclo\ w8c}$. Some fatty acids isolated from N4-5 co-occurred in just a few other species. For example $C_{10:0\ 3OH}$ was only shared with *C. plymuthica* and *C. rubidaea*, whereas $C_{12:1\ 3OH}$, $C_{12:0\ 3OH}$, $C_{14:0\ 2OH}$, and $C_{17:0}$ were only identified in other *S. marcescens*-GC subgroups.

Genome sequencing information

Genome project history

Strain N4-5 was isolated from soil and has been studied since 1996 [30]. Due to its effectiveness against multiple plant pathogens, including *M. poae*, *P. ultimum* and *Rhizoctonia solani* and its antimicrobial properties [30,33-37], strain N4-5 was selected for whole genome

sequencing in 2017. Among all the *S. marcescens* complete genomes available, there are only two others from strains that are also beneficial to plants. The addition of the N4-5 complete genome sequence into public databases will allow comparative genome analysis to better understand the mechanisms by which *S. marcescens* associates with plants and controls plant diseases, as well as the variety of lifestyles presented by *S. marcescens* strains. The N4-5 whole genome nucleotide sequence was deposited with GenBank under accession number CP031316 (Table 2).

Table 2. Project information.

MIGS ID	Property	Term
MIGS 31	Finishing quality	Finished
MIGS-28	Libraries used	Four Illumina pair end libraries
MIGS 29	Sequencing platforms	Illumina NextSeq-500
MIGS 31.2	Fold coverage	664
MIGS 30	Assemblers	SPAdes 3.10.0, IDBA 1.1.1, Velvet
MIGS 32	Gene calling method	RASTtk
	Locus Tag	Smn45
	Genbank ID	CP031316
	GenBank Date of Release	August 05, 2018
	GOLD ID	Ga0268725
	BIOPROJECT	PRJNA477367
MIGS 13	Source Material Identifier	*Serratia marcescens N4-5*
	Project relevance	Biocontrol, Agricultural, Environmental

Growth conditions and genomic DNA preparation

For DNA isolation, *S. marcescens* N4-5 was grown in batch Luria-Bertani broth (LB) culture for 16 h at 22°C and 150 rpm orbital agitation. Genomic DNA was extracted from cells with the QIAGEN Blood & Tissue genomic DNA isolation kit, using the manufacturer's protocol (Qiagen Inc., Gaithersburg, MD).

Genome sequencing and assembly

The sequence data was generated on four lanes of an Illimina NextSeq-500 using the run kit Illumina NextSeq® 500/550 High Output Kit v2. The indexed library was constructed using Nextera® XT Index Kit v2 Set A. The sequencing resulted in 22,789,104 reads, with length varying from 32 to 151 bp, which comprised a total of 3,369,822,757 bases and represented 664-fold genome coverage. The quality was checked with the program FastQC v0.11.5 [45]. The genome was assembled using the four paired reads libraries employing the assembly service "auto" available in PATRIC (Pathosystems Resource Integration Center) [46]. This strategy runs BayesHammer [47] on short reads, followed by three assembly strategies that include Velvet [48], IDBA 1.1.1 [49] and SPAdes 3.10.0 [50]. Based on each assembly score provided by the QUAST (Quality Assessment Tool for Genome Assemblies) algorithm [51], the SPAdes assembly was chosen to move on to the subsequent steps. The 1,634 contigs generated were united into 19

scaffolds using the CONTIGuator web server [52] with the *S. marcescens* strain B3R3 (accession number CP013046.2) as the reference genome. The dnaA gene was determined as the beginning of the chromosome using an in-house script. Finally, gaps were closed using the tools FGAP [53], NCBI's BLASTn [54] and CLC Genomics Workbench 7 (Qiagen Inc.), respectively. The plasmid was assembled using plasmidSPAdes [55].

Genome annotation

The N4-5 genome was annotated using the RASTtk (Rapid Annotation Using Subsystem Technology) [56] annotation service in PATRIC. Manual curation was conducted through Artemis 16.0.0 software [57] and insertion/deletion (indel) errors were checked in CLC Genomics Workbench 7. Genes with potential frameshifts were compared to other complete genes with BLASTn against the NR database at NCBI. Translated protein sequences were determined with BLASTP against the UniProt database [58]. Ribosomal RNAs were verified using the web-tool RNAmmer 1.2 [59] and tRNAs were verified with tRNAscan-SE 2.0 [60]. Signal peptides, transmembrane domains, protein families, COG categorization and type III, IV and VI secretion systems (T3SS, T4SS, T6SS) were predicted using the web-based version of SignalP [61], TMHMM [62], Pfam [63], BASys [64] and T346HUNTER [65], respectively. The SecretomeP 2.0 Server was used to predict non-classical (i.e. non-signal peptide triggered) protein secretion [66,67]. Genomic islands (GIs) were identified using IslandViewer 3 [68] and were manually investigated.

Genome Properties

The *S. marcescens* N4-5 genome comprised a single chromosome of 5,074,473 bp, with 59.7% GC content and a naturally occurring plasmid (Table 3 and Fig. 3). The chromosome had 4,884 protein-coding genes, of which 4,020 genes were functionally assigned while the remaining genes were annotated as hypothetical proteins. The pseudogenes represented 0.9% of the total number of genes. There were 82 tRNA genes and 7 copies of the ribosomal RNA operon distributed throughout the genome, which accounted for 21 rRNA genes (Table 4). The N4-5 genomic nucleotide sequence contained 2,747 transcription units and 992 operons. The gene classification into COG functional categories is presented in Table 5.

Table 3. Summary of genome: one chromosome and one plasmid

Label	Size (bp)	Topology	INSDC identifier	RefSeq ID
Chromosome	5,074,473	Circular	GenBank	CP031316
Plasmid	11,089	Circular	GenBank	CP031315

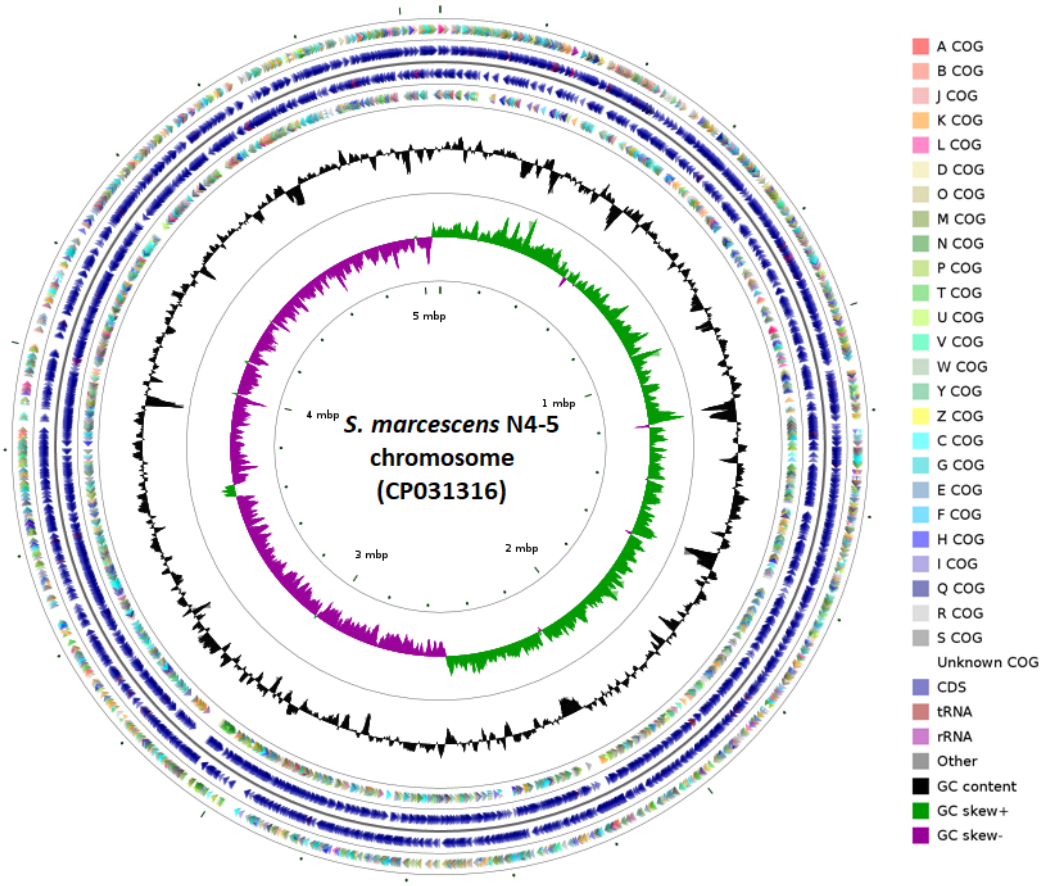

Fig. 3 Graphical circular map of *Serratia marcescens* strain N4-5 chromosome. From outer circle to the center: CDS on forward strand (colored according to COG categories), all CDS and RNA genes on forward strand, all CDS and RNA genes on reverse strand, CDS on reverse strand (colored according to COG categories), GC content and GC skew. The map was generated using GCView Comparison Tool [95].

Table 4. Genome statistics.

Attribute	Value	% of Total
Genome size (bp)	5,074,473	100.00
DNA coding (bp)	4,426,935	87.23
DNA G+C (bp)	3,029,510	59.70
DNA scaffolds	1	100.00
Total genes	4,884	100.00
Protein coding genes	4,840	99.09
RNA genes	103	2.10
Pseudo genes	44	0.90
Genes in internal clusters	NA	-
Genes with function prediction	4,020	82.31
Genes assigned to COGs	3,604	73.79
Genes with Pfam domains	4,290	87.84
Genes with signal peptides	468	9.58
Genes with transmembrane helices	1,181	9.23
CRISPR repeats	2	-

The circular plasmid comprised 11,089 bp and had 43.5% GC content. The size of the plasmid was confirmed by digestion with restriction enzymes followed by electrophoresis. *In silico* restriction and BLAST analysis showed that the plasmid was composed of segments repeated 2.5 times. The plasmid sequence encoded six unique CDS that were repeated 2.5 times totaling 13 CDS. From the six unique CDS, four were annotated as hypothetical proteins.

Table 5. Number of genes associated with general COG functional categories.

Code	Value	%age	Description
J	176	3.60	Translation, ribosomal structure and biogenesis
A	1	0.02	RNA processing and modification
K	261	5.34	Transcription
L	138	2.83	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	36	0.74	Cell cycle control, cell division, chromosome partitioning
V	49	1.00	Defense mechanisms
T	115	2.35	Signal transduction mechanisms
M	233	4.77	Cell wall/membrane biogenesis
N	78	1.60	Cell motility
U	34	0.70	Intracellular trafficking and secretion
O	138	2.83	Posttranslational modification, protein turnover, chaperones
C	212	4.34	Energy production and conversion
G	267	5.47	Carbohydrate transport and metabolism
E	401	8.21	Amino acid transport and metabolism
F	95	1.94	Nucleotide transport and metabolism
H	139	2.85	Coenzyme transport and metabolism
I	108	2.21	Lipid transport and metabolism
P	239	4.89	Inorganic ion transport and metabolism
Q	82	1.68	Secondary metabolites biosynthesis, transport and catabolism
R	393	8.05	General function prediction only
S	408	8.35	Function unknown
-	1280	26.21	Not in COGs

The total is based on the total number of protein coding genes in the genome.

Insights from the genome sequence

Strain N4-5 contained 468 proteins with a signal peptide, 1,181 proteins with transmembrane domains TMHMM (Table 4) and 451 proteins were predicted to be secreted through non-classical pathways. Searches for secretion systems revealed the presence of a flagellar type III secretion system (T3SS) and one copy of the type VI secretion system (T6SS) with 14 core components. Interestingly, the T6SS was found to be in a genomic island, which includes parts of the genome that were acquired by horizontal gene transfer (HGT) [69]. Traits obtained by HGT may be beneficial to the

host under certain conditions. For example, *S. marcescens* Db10 utilised the T6SS to target bacterial competitors [70]. Two CRISPR arrays were detected by the CRISPRfinder Server [71] and these arrays comprise the prokaryote immune system [72]. Therefore, both features may have contributed to the survival and evolution of the *S. marcescens* N4-5 genome and its success as a biological control agent of plant pathogens. Additionally, strain N4-5 is a known producer of the broad-spectrum anti-microbial prodigiosin, which contributes to its biological control activity [34-36]. In accordance, the genome of N4-5 harboured the 14 canonical genes for prodigiosin biosynthesis (pig cluster, *pigA-N*) described by Cerdeño [73] and Harris [74]. The pig cluster is predicted to be transcribed in two operons, one containing *pigA-C* and the other containing *pigD-N* genes. As seen in other bacteria [74], the N4-5 pig cluster was flanked by *copA* and cueR homologues; however, differently from the other studied strains, N4-5 has a putative membrane protein (41 amino acids) annotated between *pigA* and *cueR*. Additionally, strain N4-5 contained chitinase genes (*chiA* and *chiB*) in the genome. These enzymes hydrolyze chitin, an essential fungal cell wall component [75]. Due to their potential uses in agriculture and industry, chitinases have been intensively studied in several bacterial taxa, including *S. marcescens* [76-78]. Several multidrug resistance genes were found during functional annotation of the N4-5 genome. These data indicated that strain N4-5 contains machinery for microbial competition, which enhances the potential of this strain as a biological control agent. Strain N4-5 also had plant-beneficial traits such as coding genes for siderophores. The siderophore enterobactin gene cluster contained *entA*, *entB*, *entC*, *entE*, *entF* and *entH* but the vibriobactin genes were absent. Additionally, N4-5 carried 16 tonB-dependent transporter genes, which are cellular receptors of siderophores. The production of siderophore complexes by bacteria contributes to plant growth as they sequester iron from the environment and make it available for plant uptake [79,80]. The ability to utilize carbon sources provides a fitness advantage during microbial competition. The N4-5 genome had 267 genes responsible for carbohydrate transport and metabolism, comparable with *Pseudomonas alcaliphila* JAB1, a degrader of organic pollutants that had 196 genes with this functionality [81]. The surfactant Serrawettin W1 was coded by one NRPS (non-ribosomal peptide synthase) gene with 3,936 bp. Serrawetin W1 has antimicrobial, antitumor and zoosporicidal activities and has potential uses in agriculture, medicine and industry [82]. Altogether, these genome features support N4-5 as a plant-beneficial strain. On the other hand, N4-5 genome lacks genes for the

synthesis of the antibiotic pyrrolnitrin (*prnA-D*) and nitrogen fixation (such as *nif*, *fix* and *nod*).

Extended insights: Artefact

Two extra copies of the rRNA 5S gene were found in the fifth ribosomal cluster of strain N4-5 after the initial assembly. Extra copies were also found in other *Serratia* genomes deposited in public databases (Additional file 1: Table S2/Fig. 4). To verify whether it was an artefact or a natural feature in strain N4-5, PCR amplification and standard sequencing with the Sanger method of the rRNA 5S gene region were performed. Primers MetA1F (5'- ACC GCA GGT AAC TCA TCA GG -3') and 23S1R (5'- GAC GTT GAT AGG CTG GGT GT- 3') were anchored on the regions flanking the 5S duplication as visualized in the program Artemis. The sequence obtained with these primers was mapped to the assembly and unequivocally showed that these extra copies of the 5S rRNA gene are assembly artefacts. The genome sequence was corrected accordingly and therefore, strain N4-5 has regular ribosomal operons, i.e. one copy of the 5S, 16S and 23S genes per cluster (Fig. 4). The extra copies of the 5S gene found in the complete genomes of the other 48 *Serratia* spp. deposited in public databases (Additional file 1: Table S2/Fig. 4) are almost certainly artefacts generated during assembly and propagated with the deposit of new genomes. Strain N4-5 is the only *S. marcescens* genome in GenBank without this artefact. Therefore, the deposit of this genome will contribute to improved future assemblies and to prevent the propagation of this artefact when strain N4-5 is used as reference.

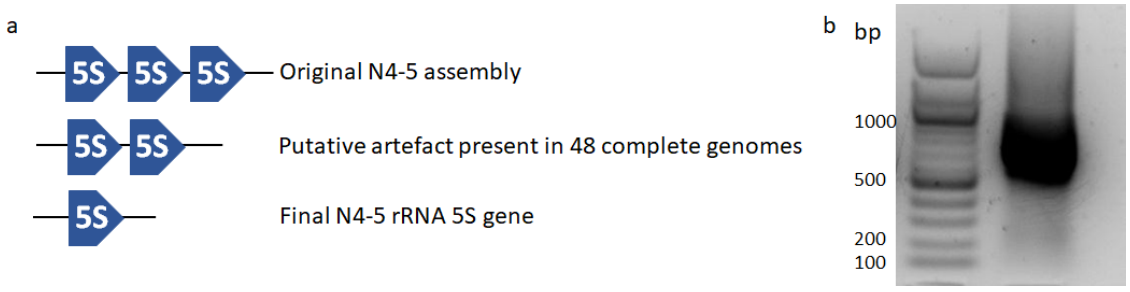

Fig. 4 Identification of putative assembly artefacts in complete genomes of *Serratia* spp. deposited in public databases. **(a)** Schematic representation of the artefact in strain N4-5 and in other *Serratia* genomes and **(b)** the electrophoretic gel image representing the expected band for a single 5S gene of approximately 800bp generated with primeirs MetA1F (5'- ACC GCA GGT AAC TCA TCA GG -3') and 23S1R (5'- GAC GTT GAT AGG CTG GGT GT- 3').

Conclusions

We obtained the *Serratia marcescens* strain N4-5 nucleotide genome sequence in its full extent. It comprised a naturally occurring plasmid and one circular chromosome with size and G+C content typical of *Serratia* genomes. This strain was sequenced based on its ability to produce antimicrobial compounds and suppress soil-borne phytopathogens. The plant beneficial properties of strain N4-5 included genes for the biosynthesis of prodigiosin, serrawetin W1, chitinases and siderophores. Manual curation

combined with *in vitro* experimentation allowed us to detect and to resolve an assembly artefact detected in the genome of N4-5. Further analysis and comparative genomics among functionally related strains will enhance our understanding of the plant beneficial properties shown by *Serratia* and other bacterial biocontrol agents.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Performed genome sequencing, assembly, annotation and analysis: LCF, JEM, MVCV and TJS. Performed microbiological experiments: DPR, JEM. Contributed with resources: VACA, DPR. Wrote the paper: LCF, JEM, JTS and DPR. JTS and DPR conceived the study and supervised the project. All authors read and approved the final manuscript.

Acknowledgements

LCF and JTS - CNPq

References

1. Raaijmakers JM, Paulitz TC, Steinberg C, Alabouvete C, Moenne-Loccoz Y. et al. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. Plant Soil 2009;321:341-361. https://doi.org/10.1007/s11104-008-9568-6.
2. Zitter TA, Hopkins DL, Thomas CE. Compendium of cucurbit diseases. St. Paul: APS, 1996;87.
3. Chauhan MS, Yadav JPS, Gangopadhyay S. Chemical control of soilborne fungal pathogen complex of seedling cotton. Tropical Pest Management 2008;34:159-161
4. Geiger F, et al. Persistent negative effects of pesticides on biodiversity and biological control potential on European farmland. Bacis and Applied Ecology 2010;11:97-105.
5. Cox C, Surgan M. Unidentified Inert Ingredients in Pesticides: Implications for Human and Environmental Health. Environmental Health Perspectives. 2006;114(12):1803-1806. doi:10.1289/ehp.9374.
6. Igbedioh SO. Effects of Agricultural Pesticides on Humans, Animals, and Higher Plants in Developing Countries. Archives of Environ Health 1991;46:218-224.
7. Wang K, Yan PS, Cao LX. Chitinase from a novel strain of Serratia marcescens JPP1 for biocontrol of aflatoxin: Molecular characterization and production optimization using response surface methodology. BioMed Research International 2014;1-8.
8. Someya N, Nakajima M, Watanabe K, Hibi T, Akutsu K. Potential of Serratia marcescens strain B2 for biological control of rice sheath blight. Biocontrol Science and Technology 2005;15:105–109.
9. Benitez T, Rincon AM, Limon MC, Codon AC. Biocontrol mechanisms of Trichoderma strains. International Microbiology 2004;7:249-260.
10. Pal KK, Gardener BM,. Biological Control of Plant Pathogens. The Plant Health Instructor 2006; DOI: 10.1094/PHI-A-2006-1117-02.
11. Palumbo JD, Yuen GY, Jochum CC, Tatum K, Kobayashi DY. Mutagenesis of beta-1,3-glucanase genes in Lysobacter enzymogenes strain C3 results in reduced biological control activity toward Bipolaris leaf spot of tall fescue and Pythium damping-off of sugar beet. Phytopathology 2005;95:701-707.
12. Ramamoorthy V, Viswanathan R, Raguchander T, Prakasam V, Smaiyappan R. Induction of systemic resistance by plant growth promoting rhizobacteria in crop plants against pests and diseases. Crop. Prot 2001;20:1-11.

13. Lavania M, Chauhan PS, Chauhan SVS, Singh HB, Nautiyal CS. Induction of Plant Defense Enzymes and Phenolics by Treatment With Plant Growth – Promoting Rhizobacteria Serratia marcescens NBRI1213. Current Microbiology 2006;52:363–368.
14. Lim YL, Yong D, Ee R, Krishnan T, Tee KK, Yin WF, Chan KG. Complete genome sequence of Serratia fonticola DSM 4576 T, a potential plant growth promoting bacterium. Journal of Biotechnology 2015;214:43-44.
15. Afzal I, Iqrar I, Shinwari ZK, Yasmin A. Plant growth-promoting potential of endophytic bacteria isolated from roots of wild Dodonaea viscosa L. Plant Growth Regulation 2017;1-10.
16. Zaheer A, Mirza BS, Mclean JE, Yasmin S, Shah TM, Malik KA, Mirza MS. Association of plant growth-promoting Serratia spp. with the root nodules of chickpea. Research in Microbiology 2016;6:510–520.
17. Abebe-Akele F, Tisa LS, Cooper VS, Hatcher PJ, Abebe E, Thomas WK. Genome sequence and comparative analysis of a putative entomopathogenic Serratia isolated from Caenorhabditis briggsae. BMC Genomics 2015;16:1.
18. Vicente CS, Nascimento FX, Ikuyo Y, Cock PJ, Mota M, Hasegawa K. The genome and genetics of a high oxidative stress tolerant Serratia sp. LCN16 isolated from the plant parasitic nematode Bursaphelenchus xylophilus. BMC Genomics 2016;17:1.
19. García-Fraile P, Chudíčková, M.; Benada, O.; Pikula, J.; Kolařík, M. Serratia myotis sp. nov. and Serratia vespertilionis sp. nov., isolated from bats hibernating in caves. International Journal of Systematic and Evolutionary Microbiology, v. 65, n. 1, p. 90- 94, 2015.
20. Roy P, Ahmed NH, Grover RK. Non-pigmented strain of Serratia marcescens: An unusual pathogen causing pulmonary infection in a patient with malignancy. Journal of Clinical and Diagnostic Research: JCDR 2014;8.
21. Bonnin RA, Girlich D, Imanci D, Dortet L, Naas T. Draft genome sequence of the Serratia rubidaea CIP 103234T reference strain, a human-opportunistic pathogen. Genome Announcements, 2015;3:e01340-15.
22. Grimont F, Grimont PAD. The genus *Serratia*. In: Ballows, A. (Ed.), The Prokaryotes—A Handbook on the Biology of Bacteria:Ecophysiology, Isolation, Identification, Applications, second ed., vol. III. Springer, New York, NY, 1992;2823–2848.
23. Kamensky M, Ovadis M, Chet I, Chernin L. Soil-borne strain IC14 of Serratia plymuthica with multiple mechanisms of antifungal activity provides biocontrol of Botrytis cinerea and Sclerotinia sclerotiorum diseases. Soil Biology and Biochemistry 2003;35:323-331.
24. Giri AV, Anandkumar N, Muthukumaran G, Pennathur G. A novel medium for the enhanced cell growth and production of prodigiosin from Serratia marcescens isolated from soil. BMC microbiology 2004;41.
25. Lavania M, Nautiyal CS. Solubilization of tricalcium phosphate by temperature and salt tolerant Serratia marcescens NBRI1213 isolated from alkaline soils. African Journal of Microbiology Research 2013;7:4403-4413.
26. Gavini F, Ferragut C, Izard D, Trinel PA, Leclerc H, Lefebvre B, Mossel DAA. Serratia fonticola, a new species from water. International Journal of Systematic and Evolutionary Microbiology 1979;29:92-101.
27. Henriques I, Ramos RTJ, Baraúna RA, De Sá PH, Almeida DM, Carneiro AR, Egas C. Draft genome sequence of Serratia fonticola UTAD54, a carbapenem-resistant strain isolated from drinking water. Genome announcements 2013;1:e00970-13.
28. Kampfer P, Glaeser SP. Serratia aquatilis sp. nov., isolated from drinking water systems. Int. J. Syst. Evol. Microbiol 2016;66:407-413.
29. Bencini MA, Yzermana EPF, Bruina JP, Den Boer JW. Airborne dispersion of Serratia marcescens as a model for spread of Legionella from a whirlpool. The Royal Institute of Public Health 2008;962–964.
30. Kobayashi DY, El-Barrad N. Selection of bacterial antagonists using enrichment cultures for the control of summer patch disease in Kentucky Bluegrass. Current Microbiology. 1996;32:106–111.

31. Press C, Wilson M, Tuzun S, Kloepper JW. Salicylic acid produced by *Serratia marcescens* is not the primary determinant of induced systemic resistance in cucumber or tobacco. Molecular Plant–Microbe Interactions 1997;10:761–768.
32. Someya, N, Kataoka N, Komagata T, Hirayae K, Hibi T. Biological control of cyclamen soilborne diseases by *Serratia marcescens* strain B2. Plant Disease 2000;84:334–340.
33. Roberts DP, Lohrke SM, Meyer SLF, Buyer JS, Bowers JH, Baker CJ, Li W, Souza JT, Lewis JA, Chung S. Biocontrol agents applied individually and in combination for suppression of soilborne diseases of cucumber. Crop Protection 2005;24:141-155
34. Roberts DP, McKenna LF, Lakshman, DK, Meyer SLF, Kong H, de Souza JG, Lydon J, Baker CJ, Chung S. Suppression of damping-off of cucumber caused by *Pythium ultimum* with live cells and extracts of *Serratia marcescens*. Soil Biol. Biochem. 2007;39:2275-2288.
35. Roberts DP, Lakshman DK, Maul JE, McKenna LF, Buyer JS, Fan B. Control of damping-off of organic and conventional cucumber with extracts from a plant-associated bacterium rivals a seed treatment pesticide. Crop Prot. 2014;65:86-94.
36. Roberts DP, Lakshman DK, McKenna LF, Emche SE, Maul JE, Bauchan G. Seed treatment with ethanol extract of *Serratia marcescens* is compatible with *Trichoderma* isolates for control of damping-off of cucumber caused by *Pythium ultimum.* Plant Dis. 2016;100:1278-1287.
37. Roberts DP, McKenna LF, Buyer JS. Consistency of control of damping-off of cucumber is improved by combining ethanol extract of *Serratia marcescens* with other biologically based technologies. Crop Protection 2017;96:59-67.
38. Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics, 2015.
39. Rodriguez-R LM, Konstantinidis KT. Bypassing Cultivation To Identify Bacterial Species. Microbe. 2014;9:111-118.
40. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2010; 2:142-148. http://dx.doi.org/10.4056/sigs.541628
41. Traub WH. Antibiotic susceptibility of *Serratia marcescens* and *Serratia liquefaciens*. Chemotherapy 2000;46:315–321.
42. Stock I, Grueger T, Widemann B. Natural antibiotic susceptibility of strains of *Serratia marcescens* and the *S. liquefaciens* complex: S. *liquefaciens* sensu stricto, *Sproteamaculans* and *S. grimesii*. International Journal of Antimicrobial Agents 2003;22:35–47.
43. Sandner-Miranda L, Vinuesa P, Soberón-Chávez G, Morales-Espinosa R. Complete Genome Sequence of Serratia marcescens SmUNAM836, a Nonpigmented Multidrug-Resistant Strain Isolated from a Mexican Patient with Obstructive Pulmonary Disease. Genome Announc 2016;4: e01417-15.
44. Iguchi A, Nagaya Y, Pradel E, Ooka T, Ogura Y, Katsura K, Kurokawa K, Oshima K, Hattori M, Parkhill J, Sebaihia M, Coulthurst SJ, Gotoh N, Thomson NR, Ewbank JJ, Hayashi T. Genome evolution and plasticity of Serratia marcescens, an important multidrug-resistant noso- comial pathogen. Genome Biol Evol 2014;6:2096–2110.
45. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc, 2010.
46. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinfor- matics database and analysis resource. Nucleic Acids Res 2014;42:D581–91. https://doi.org/10. 1093/nar/gkt1099.
47. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. BMC genomics. 2013;14:1.
48. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–829.
49. Peng Y, Leung Hcm, Yiu Sm, Chin Fyl. IDBA–a practical iterative de Bruijn graph de novo assembler. Research in Computational Molecular Biology. 2010;6044: 426-440.

50. Bankevich A, et al. SPAdes: a new genome assembly algorithm and Its Applications to Single-Cell Sequencing. Journal of Computational Biology. 2012;19:455–477.
51. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: Quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–1075.
52. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code Biol Med. 2011;6:11.
53. Piro VC, Faoro H, Weiss VA, Steffens MB, Pedrosa FO, Souza EM, Raittz RT. FGAP: an automated gap closing tool. BMC Research Notes. 2014;7:371.
54. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.J. Mol. Biol. 1999;215:403-410.
55. Antipov D, Hartwick N, Shen M, Raiko M, Lapidus A, Pevzner PA. plasmidSPAdes: assembling plasmids from whole genome sequencing data. Bioinformatics, 2016;32:3380-3387.
56. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep 2015; doi:10.1038/srep08365.
57. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M, Barrell B. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16:944–945.
58. Wasmuth EV, Lima CD. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017;45:D158–D169.
59. Lagesen K, Hallin PF, Rødland E, Stærfeldt HH, Rognes T, Ussery DW. RNammer: consistent annotation of rRNA genes in genomic sequences. Nucleic Acids Res. 2007.
60. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Research. 2016;44:W54–W57.
61. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 2011;8:785–6
62. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001;305:567–80.
63. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research. 2016;44:D279-D285.
64. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, et al. BASys: a web server for automated bacterial genome annotation. Nucleic Acids Res. 2005;33:W455–9.
65. Martínez-García PM, Ramos C, Rodríguez- Palenzuela P. T346Hunter: A Novel Web-Based Tool for the Prediction of Type III, Type IV and Type VI Secretion Systems in Bacterial Genomes. PLoS ONE. 2015;10:e0119317. doi:10.1371/journal. pone.011931
66. Bendtsen JD, Jensen LJ, Blom N, von Heijne G, Brunak S. Feature based prediction of non-classical and leaderless protein secretion. Protein Eng. Des. Sel. 2004;17:349-356.
67. Bendtsen JD, Kiemer L, Fausbøll A, Brunak S. Non-classical protein secretion in bacteria. BMC Microbiology. 2005;5:58.
68. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, Mcarthur AG, Langille MG, Brinkman FS. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Research. 2015;43:W104-W108. doi: 10.1093/nar/gkv401.
69. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 2007;35:W52-7
70. Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. Top E, ed. Fems Microbiology Reviews. 2009;33:376-393. doi:10.1111/j.1574-6976.2008.00136.x.

71. Murdoch SL, Trunk K, English G, Fritsch MJ, Pourkamari E, Coulthurst. The opportunistic pathogen Serratia marcescens utilizes type VI secretion to target bacterial competitors. J Bacteriol. 2011;193:6057–6069.
72. Rath D, Amlinger L, Rath A, Lundgrenb M. The CRISPR-Cas immune system: Biology, mechanisms and applications. Biochimie, 2015;117:119-128.
73. Cerdeño AM, Bibb MJ, Challis GL. Analysis of the prodiginine biosynthesis gene cluster of Streptomy- ces coelicolor A3(2): new mechanisms for chain initiation and termination in modular multienzymes. Chem. Biol. 2001;119:1–13.
74. Harris AKP, et al. The Serratia gene cluster encoding biosynthesis of the red antibiotic, prodigiosin, shows species- and strain-dependent genome context variation. Microbiology. 2004;150:3547–3560.
75. Latge JP. The cell wall: a carbohydrate armour for the fungal cell. Mol Microbiol 2007;66:279–290.
76. Bansode VB, Bajekal SS. Characterization of chitinases from microorganisms isolated from Lonar lake. Indian Journal of Biotechnology 2006;5:357–363.
77. Ordentlich A, Elad Y, Chet I. The role of the chitinase of Serratia marcescens in Biocontrol of Sclerotium rolfsii. Phytopatology, 1988;78:84-88.
78. Zarei M, Aminzadeh S, Zolgharnein H, Safahieh A, Daliri M, Noghabi1 KA, Ghoroghi A, Motallebi A. Characterization of a chitinase with antifungal activity from a native Serratia marcescens B4A. Brazilian Journal of Microbiology 2011;42:1017–1029.
79. Scavino AF, Pedraza RO. The Role of Siderophores in Plant Growth-Promoting Bacteria. In: Maheshwari D., Saraf M., Aeron A. (eds) Bacteria in Agrobiology: Crop Productivity. Springer, Berlin, Heidelberg 2013.
80. Sharma A, Johri BN. Growth promoting influence of siderophore-producing Pseudomonas strains GRP3A and PRS9 in maize (Zea mays L.) under iron limiting conditions. Microbiological Research 2003;158:243-248.
81. Ridl J, Suman J, Fraraccio S, et al. Complete genome sequence of Pseudomonas alcaliphila JAB1 (=DSM 26533), a versatile degrader of organic pollutants. Standards in Genomic Sciences. 2018;13:3. doi:10.1186/s40793-017-0306-7.
82. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequences (MIGS) specification. Nat Biotechnol. 2008;26:541–7
83. Woese CR. Towards a natural system of organisms: Proposal for the domains Archea, Bacteria and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.
84. Garrity GM, Bell JA, Lilburn T. Phylum XIV. Proteobacteria phyl. Nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, 2nd ed, Volume 2, Part B. New York: Springer; 2005: p. 1.
85. Garrity GM, Bell JA, Lilburn T. Class III. Gammaproteobacteria class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, Springer 2005.
86. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. Int J Syst Evol Microbiol. 2005; 55:2235–8.
87. Garrity GM, Holt JG. Taxonomic outline of the Archaea and Bacteria. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's Manual of Systematic Bacteriology, vol. 1. 2nd ed. New York: Springer; 2001;155–66.
88. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.
89. Brenner DJ, Farmer JJ. Family I. Enterobacteriaceae Rahn 1937. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology, 2nd ed, Volume 2, Part B. New York: Springer; 2005;587.
90. Rahn O. New principles for the classification of bacteria. Zentralblatt für Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Abteilung II. 1937; 96:273–86.
91. Bizio B. Lettera di Bartolomeo Bizio al chiarissimo canonico Angelo Bellani sopra il fenomeno della polenta porporina. Biblioteca Italiana o sia Giornale di Letteratura. [Anno VIII]. Scienze e Arti. 1823;30:275–95

92. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.
93. Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 2013;30:2725-2729.
94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 2013;30:772-780.
95. Grant JR, Arantes AS, Stothard P. Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics. 2012;13:202.

ARTICLE 2 - Genus-wide Taxonomy Review and Comparative Analyses of *Serratia*: Data from Complete Genomes

mBio journal format (preliminary version).

Genus-wide Taxonomy Review and Comparative Analysis of *Serratia*: Data from Complete Genomes

Larissa C. Ferreira[a] Daniel P. Roberts[b] Jorge T. de Souza[a#]

[a]Department of Plant Pathology, Federal University of Lavras, Lavras, Minas Gerais, Brazil

[b]Sustainable Agricultural Systems Laboratory, United States Department of Agriculture–Agricultural Research Service, Beltsville, Maryland, USA

Running Head: Genus-wide comparative analysis of *Serratia*

#Address correspondence to Jorge T. de Souza, jgeteodoro@gmail.com

ABSTRACT

Bacteria from the genus *Serratia* are ubiquitous in the environment and have been detected in soil, water, and air and in association with plants and animals; including humans in hospital settings. These bacteria occupy diverse lifestyles as pathogens on plants or animals or are beneficial to plants promoting growth and controlling disease. *In silico* comparisons of all complete genomes of *Serratia* available in GenBank are presented here along with descriptions of their genetic organization, their secretion systems, and their secondary metabolite biosynthetic gene clusters, chitinase genes and CRISPR arrays. A taxonomic review of the genus *Serratia* based on dDDH, ANI, 16S identity, phylogenetic trees of seven housekeeping genes (individually and concatenated), as well as phylogenomic tree construction with whole genomes detected two misidentifications, supported a recent proposal of a novel *Serratia* species, confirmed the taxonomic placement of most strains and indicated that certain *Serratia* genomes publicly deposited in GenBank are incorrectly named. From these genomes we constructed updated pan-, core- and accessory-genomes. Analysis revealed an open pan-genome with 546 core genes. Secretion system analysis revealed two T6SS loci, one T4SS and one non-flagellar T3SS. In general, genetic organization of *Serratia* secretion systems is complex and non-standard. Secondary metabolite BGC prediction revealed that 41% of compounds encoded by *Serratia* genomes are antibiotics, with some having anti-tumor

properties. Four of the seven *Serratia* species have CRISPR arrays while five species encode chitinases.

IMPORTANCE

Serratia, an opportunistic nosocomial pathogen, are gram-negative bacteria also important in environmental and agricultural scenarios. Moreover, *Serratia* strains attract industry's attention due to their ability to produce several antimicrobial compounds and surfactants. Here we present a genus-wide analysis in *Serratia*. In this study, we used complete genome sequences to analyze and to present the state-of-art of *Serratia*. Our results provide insights from *Serratia* at the genus level.

KEYWORDS Nosocomial pathogen, plant growth promoting bacteria, entomopathogen, average nucleotide identity, DDH, antibiotics, secretion system

INTRODUCTION

Bacterial studies has been greatly influenced by DNA sequencing advancements and improved capacity to analyze large data sets. As a result, the number of bacterial genomes assembled, annotated and publicly available in databases has increased exponentially (1). From this phenomenon arises numerous possible ways of analyzing these new genomes. For instance, bacterial taxonomy evolved throughout the years to the point that *in silico* genome-based analyses provide equally or more accurate results compared to those obtained through conventional methods (2-4). These *in silico* metrics include mainly digital DNA-DNA hybridization (dDDH) and average nucleotide/amino acid identity (ANI/AAI) (4-7). The use of these and other metrics to aid bacterial species delineation studies have been widely adopted by microbiologists thus following the trend towards incorporating these approaches into prokaryotic taxonomy (8, 9). In fact, the widespread application of genomics methods in prokaryotic taxonomy and systematics revealed several misidentifications and proposals of new taxa and reclassifications (10-12). Beyond taxonomy and systematics, genomic analysis provided better understanding of genotypic/phenotypic characteristics inherent in organisms individually studied and also contributed to elucidate questions at the Kingdom level (13-17).

Here, we applied a range of genomic approaches to study the *Serratia* genus. This taxon of gram-negative bacteria displays very diverse biological functions and lifestyles, including pathogenicity to humans, animals and plants (18-20). However, there are also non-pathogenic *Serratia* strains able to promote plant growth and to control plant pathogens (21, 22). Interestingly, all these phenotypes can be found in strains from the same species, i.e. *S. marcescens* (23-26). These bacteria carry tools that enable them to

exhibit this wide phenotypic variation, for example, production of several antibiotics, surfactants, chitinolytic enzymes and effector molecules (19, 27-31). One example of antibiotic produced by *Serratia* isolates is the secondary metabolite, red pigment called prodigiosin (32, 33). This powerful compound holds the ability to kill gram-negative and positive bacteria, fungi, nematodes, insects and even cancer cells (34-39). Its potential applications in medicine and agriculture is enormous. The prodigiosin biosynthetic gene cluster (BGC) is composed by 14 genes and is widespread within *Serratia* genomes (40, 41). In addition to the production of multi-target substances, secretion system (SS) is other tool employed by *Serratia* organisms to thrive in nature. These bacteria are known to target bacterial competitors by using their type VI SS apparatus to deliver toxins into adjacent prokaryotic cells (42, 43). Secretion systems are machineries in bacterial cells also related to pathogenicity, DNA-conjugation, and general transport of molecules into the environment, prokaryotic and eukaryotic cells (44-47). Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are an adaptive bacterial immune system (48). Alongside with the features cited above, the presence of immune system also provide fitness competition advantage (48, 49). Besides CRISPR exploitation and relevance as a genomic editing tool, it has also an important natural biological function of providing bacterial protection against phages (50). These features are advantageous for bacterial survival and competition in clinical scenarios as well as in plant associations (such as growth promotion and biocontrol of phytopathogens). Therefore, the presence of these features have an effect in their success in nature and, in the long run, in their evolutionary arms race.

Currently, there are 17 *Serratia* species recognized within the genus. However, 78% of *Serratia* genomes sequenced and available in GenBank are *S. marcescens* (51). This species is the best studied whereas other *Serratia* spp. are left aside in many studies. This is the first study to include all *Serratia* complete genomes and to make comparisons at a genus level. We aimed to review the taxonomy placement of *Serratia* complete genomes deposited and to provide their species boundaries. In addition, we presented and investigated secretion systems and secondary metabolites gene clusters and the presence of chitinases and CRISPR arrays in the genomes studied.

RESULTS

Taxonomy review and new classification criteria of the *Serratia* genus. In order to verify the taxonomic identity of the strains selected to be used in this study, a heatmap and corresponding phylogram were constructed using dDDH values (Fig. 1).

Two misidentifications were found. The strains *S. symbiotica* ‘Cinara cedri’ and STs have dDDH, ANI and 16S identity values lower than the cutoff for species delineation in each of these criteria (Table S2-6). In average, the ‘Cinara cedri’ and STs dDDH values are respectively 21.73 and 34.23% when compared to the other *Serratia* spp. genomes. Likewise, the average nucleotide identity varied from 66.65 to 74.92% in comparison to the genomes sampled. The 16S gene identity matrix corroborates with these results. Based on a multi-criteria analysis of dDDH, GC variation, ANIb, ANIm and 16S phylogenetic tree (Table S2-6 and Fig. S1), we strongly suggest that *S. symbiotica* strains ‘Cinara cedri’ and STs do not belong in the *Serratia* genus and should be reclassified. For this reason, we excluded both strains from the subsequent analyses.

The uncharacterized *Serratia* sp. strains were allocated into three *Serratia* species used in this study, except for strain YD25. The strains FS14 and SCBI were grouped within *S. marcescens*, FGI94 within *S. rubidaea* and AS12 and AS13 within *S. plymuthica*. The dDDH values were above the cutoff for all these strains, except FS14 and SCBI, which values varied from 57.6 to 89.3% and 60.7 to 88.8% respectively. Even though some values were lower than 70%, all the other criteria cited above, indicate these strains belong with *S. marcescens* species. The closest species to the strain YD25 is *S. marcescens* according to genomic indexes. For instance, the highest dDDH value is 58.7%, being in average, 56.46% similar to the *S. marcescens* strains. Likewise, the pairwise ANIb values varied from 93.51 to 94.32% among *S. marcescens* strains. The same result was observed in the phylogenomic tree with whole genome sequences and in the phylogenetic trees constructed with *ftsZ*, *groES*, *gyrB*, *recA* and *rpoD* individual and concatenated genes (MLSA) (Fig. S2-6, Fig. 2). In contrast, in the phylogenetic trees of the *groEL* gene and of the 16S containing type strains for all *Serratia* species, YD25 was grouped with *S. marcescens* (Fig. S1,7,8). Recently, Su et al. (52) proposed a novel *Serratia* species, *S. surfactantfaciens*, based on YD25 genomic analysis combined with phylogenetic and phenotypic analyses. Our results show that YD25 is very close/similar to *S. marcescens* genomes but not enough to be in this species and, therefore, we support the findings of Su et al. (52).

According to the GGDC subspecies delineation criteria (53), six subgroups were delineated within the *S. marcescens* species, one in *S. plymuthica*, *S. liquefaciens* and *S. rubidaea*. The strains UMH11, UMH10, UMH1 and UMH12 were grouped together with *S. marcescens* subsp. *marcescens* strain Db11 - their dDDH was higher than 92.5%. The second subgroup was composed by the *S. marcescens* strains UHM9, UHM3,

SmUNAM836, SMB2099 and SM39 (dDDH>90.4%), the third by the strains CAV1492 and UMH5 (dDDH=84%), the fourth by UHM7, UHM2, RSC14 UHM6 and SCBI (dDDH>89.3%), the fifth by U36365, B3R3, FS14 and WW4 (dDDH>83.4%) and the sixth by UMH8 and N4-5 (DDH=90.3%). The *S. plymuthica* strains 3Re4-18, 3Rp8, 4Rx13 and S13 formed a subgroup with dDDH higher than 89.7%. All strains classified as the *S. liquefaciens* and *S. rubideae* used in this study have dDDH percentages above the 79% cuttoff, namely dDDH>90.7% for *S. liquefaciens* and equal to 79.4% between *S. rubidaea* strains 1122 and FGI94.

To understand the species boundaries of *Serratia* spp., we analyzed four genomic indexes and four genomic features (Table 1). *Serratia* genomes comprise one chromosome and some individuals harbor extrachromosomal plasmids. The GC content, number of genes and protein varied in accordance with the genome size range. *Serratia* mean genome size is around 5.144 Mb, with *S. fonticola* strain GS2 being the largest - 6.3258 Mb - and *S. rubidaea* strain FGI94 the smallest - 4.85822 Mb. *Serratia* GC content varied from 53.6 to 60.2%. The *in silico* DNA-DNA hybridization varied from 20.3 to 100%. The other genomic metrics were also as high as 100%. Among strains from the same species, the lowest dDDH, ANIb, ANIm and 16S identity values were respectively 55.4, 93.51, 94.22 and 98.65%. All these values were from *S. marcescens*, which is the species most represented in the genus with 24 complete genome sequences. In addition, their GC ratio has the largest variation (1.57%) among all the strains. These wide ranges is due its phenotypic diversity among strains and because the amount of genomes sampled.

TABLE 1 *Serratia* species boundaries based on genomic properties and indexes.

Indexes	***Serratia* species**							
	S. marcescens	*S. plymuthica*	*S. fonticola*	*S. liquefaciens*	*S. rubidaea*	*S. proteamaculans*	*S. ficaria*	*S. surfactantfaciens*
Genomes	24	8	3	3	2	1	1	1
Size (Mb)	5.024 - 5.828	5.403 - 5.546	5.725 - 6.325	5.282 - 5.326	4.858 - 4.922	5.495	5.209	5.117
GC (%)	58.63 - 60.2	55.7 - 56.2	53.6 - 53.8	53.6 - 53.8	58.9 - 59.2	55.05	60.1	59.6
Gene	4734 - 5649	5060 - 5240	5229 - 5915	5006 - 5044	4540 - 4613	5198	4874	4866
Protein	4320 - 5354	4804 - 5024	5064 - 5580	4726 - 4863	4363 - 4459	4999	4673	4713
dDDH (%)	55.4 - 100	57.1 - 96.8	63.5 - 65.4	90.7 - 90.8	64	-	-	-
ANIb (%)	93.51-100	93.79 - 100	94.83 - 95.28	98.67 - 98.75	97.36	-	-	-
ANIm (%)	94.22 - 99.99	94.47 – 99.99	95.56 - 95.86	98.92 - 98.95	97.75	-	-	-
16S (%)	98.65 - 100	99.03 -100	99.55 - 99.87	99.94 - 99.16	99.67	-	-	-

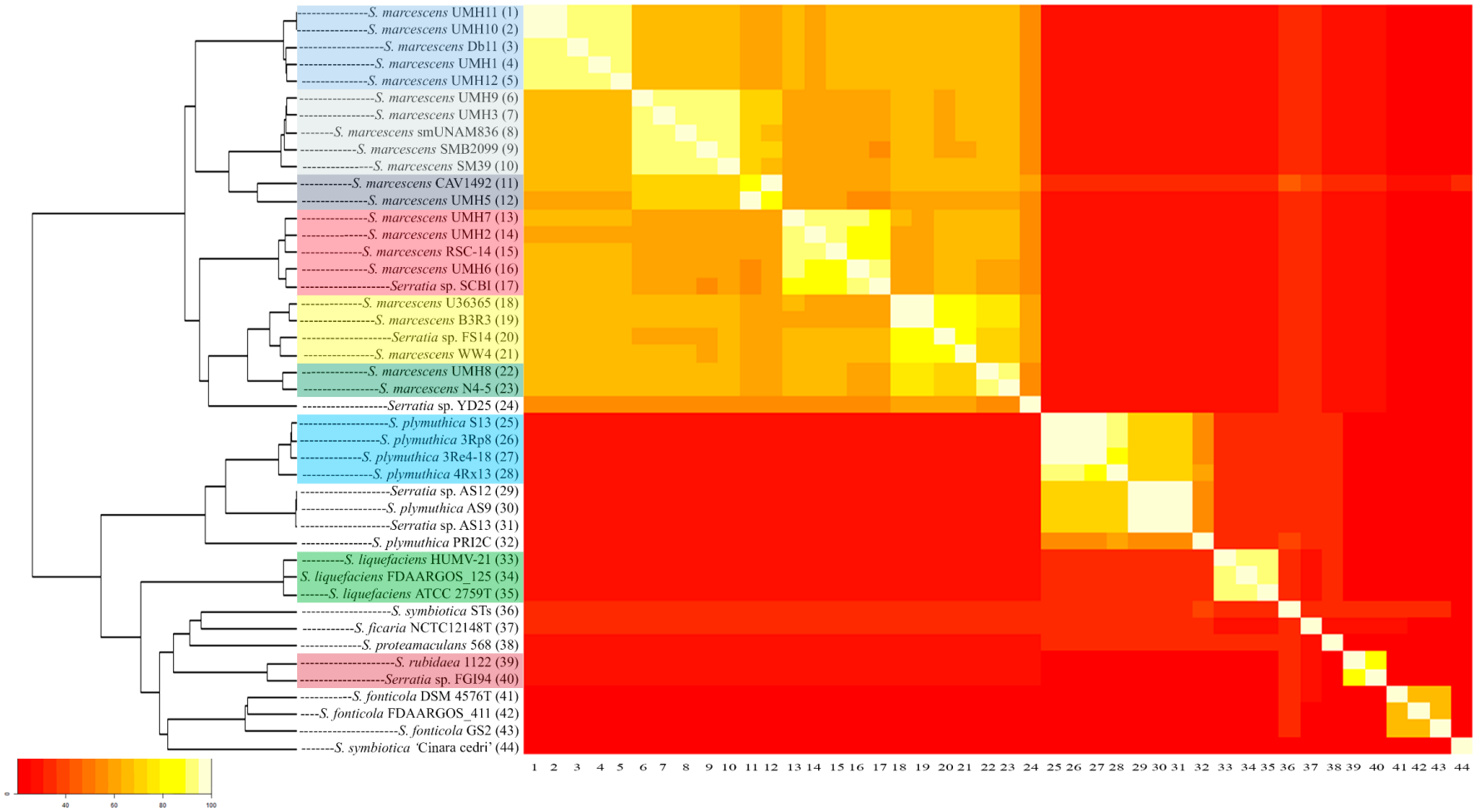

FIG 1 Heatmap constructed with dDDH values (formula 2) and corresponding phylogram. Numbers in x-axis represent the strains indicated in the y-axis. Highlighted clades represent genomic groups of strains with dDDH values higher than 79%.

Among the genes and approaches used for phylogenetic analysis and species delineation, the gene *recA* yielded results similar to those obtained through genomic methods (Fig. 2). Therefore, this gene could be used in labs as a cheaper and accurate alternative method to distinguish *Serratia* species as opposed to sequencing multiple genes or the whole genome.

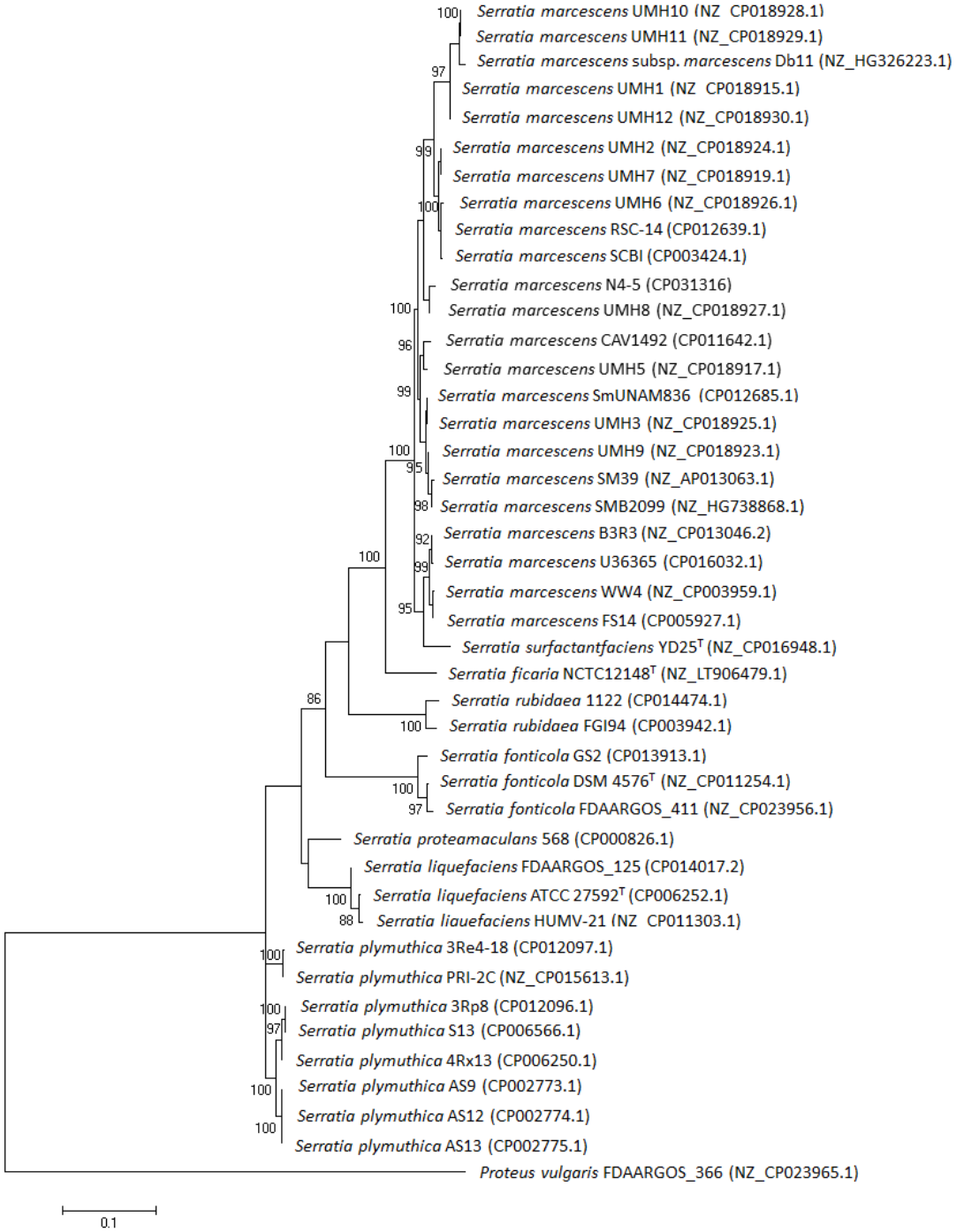

FIG 2 Phylogenetic tree indicating the current taxonomic placement and identity of 42 *Serratia* spp. complete genomes. The phylogenetic tree was constructed in MEGA 6.06 based on complete nucleotide sequences of the *recA* gene (890 bp) aligned in MAFFT (87-89). The tree construction method was Maximum Likelihood with the TN93 model

incorporating invariant sites and gamma distribution. Bootstrap values were calculated with 1,000 resamplings and values higher than 70% are shown on the appropriate branching points. The letter ^{T} refers to the type strain of the species.

Serratia pan-, core- and accessory genomes. The present study conducted an *in silico* pan-genome analysis among 42 *Serratia* spp. genomes isolated from distinct geographic locations and with diverse lifestyles. As expected, the analysis revealed an open pan-genome (B_{pan}=0.327163) with 12,881 gene families and the core genome size decreased as more strains were added (Fig. 3A). The core genome of *Serratia* spp. comprises 546 genes, distributed in 13 families. These conserved genes were categorized into five COG categories responsible for basic cellular functions, of which 54.43% are involved in amino acid transport and metabolism (E) and 15% of protein functions are related to transcription (K) and cell wall/membrane/envelope biogenesis (M) (Fig. 3B). These findings are consistent with results obtained from comparative analysis between 10 *Serratia* strains (41). On the other hand, Basharat and Yasmin (54) used a different approach to calculate the *Serratia* spp. pan and core-genome. Their analysis included draft genomes (100 strains total) and several tools. They found a core genome of 972 genes/100 genomes, twice as much the results obtained in this present study using 42 genomes.

The accessory genome (dispensable set of genes) varied from 1,864 genes (*S. rubidaea* FGI94) to 4,469 (*S. marcescens* N4-5). Analyzing further the unique (strain-specific/singletons) genes, we observed that among the five genomes with the highest amount of unique genes three strains are from *S. fonticola* species and two from *S. rubidaea*. The strain *S. fonticola* FDAARGOS_411 has the largest amount of unique genes (448) while the strains *S. marcescens* UMH10, *S. plymuthica* AS9, AS13 do not harbor singletons. Accordingly, *S. fonticola* displays the highest numbers of genes and proteins as seen in Table 1.

Approximately 50% of the *Serratia* unique genes and the accessory genome are categorized into the same five COG categories in which 15.08% of unique genes are in the general function category (R), 13.07% in transcription (K), 8.04% are related with both carbohydrate transport and metabolism (G) and function unknown (S) and 7.40% of genes are responsible for amino acid transport and metabolism (E). The accessory genes accounted for similar percentages (Fig. 3B). Among all the genomes, a total of 622 accessory sequences, 448 singletons and 3 core genes were identified with atypical GC content. These three conserved genes code for two glutamine synthases (amino acid

synthase activity) and an APC family permease (transmembrane transporter activity). Analysis of exclusively absent sequences showed that 93% of the organisms have six exclusively missing genes (data not shown).

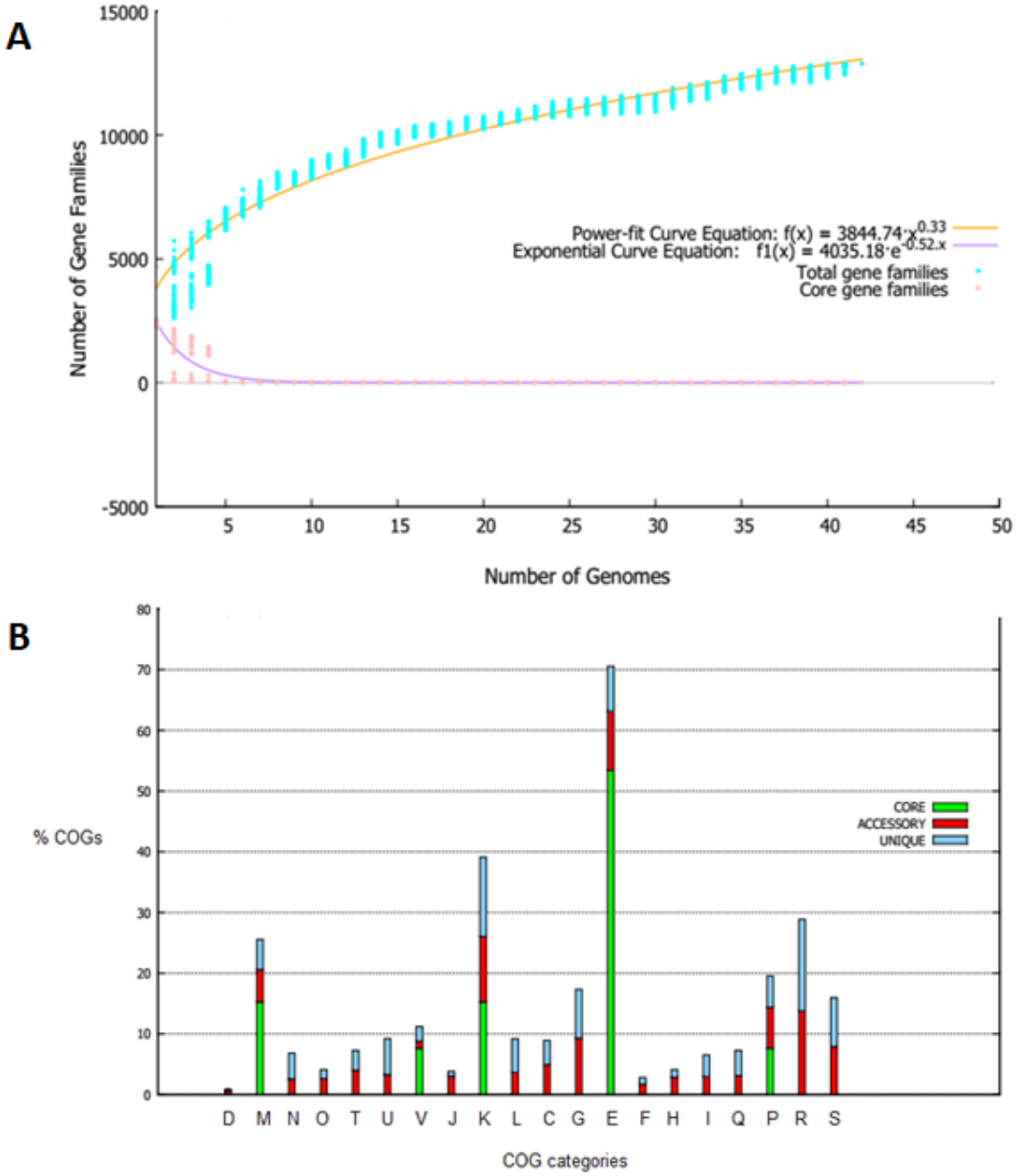

FIG 3 Pan- and core-genome curves (A), calculated with 42 *Serratia* spp. complete genome sequences and its COG distribution (B). The COG categories are: [D] Cell cycle control, cell division, chromosome partitioning; [M] Cell wall/membrane/envelope biogenesis; [N] Cell motility; [O] Post-translational modification, protein turnover, and chaperones; [T] Signal transduction mechanisms; [U] Intracellular trafficking, secretion, and vesicular transport; [V] Defense mechanisms; [J] Translation, ribosomal structure and biogenesis; [K] Transcription; [L] Replication, recombination and repair; [C] Energy production and conversion; [G] Carbohydrate transport metabolism; [E] Amino acid transport and metabolism; [F] Nucleotide transport and metabolism; [H] Coenzyme transport and metabolism; [I] Lipid transport and metabolism; [Q] Secondary metabolites biosynthesis, transport and catabolism; [P] Inorganic ion transport and metabolism; [R] General function prediction only; [S] Function unknown.

***Serratia* secretion systems have complex nonstandard genetic organization.** Secretion systems are important bacterial machineries for expelling proteins and genetic material into the environment and/or other prokaryotic and eukaryotic cells. Here we present the genetic organization displayed in two type VI secretion system (T6SS) loci, a type IV (T4SS) and a non-flagellar type III secretion system (NF-T3SS) in organisms within the *Serratia* genus (Fig. 4). The type VI SSs are different from each other, in terms of genomic order of conserved genes and presence/number of these genes. These diferences are also seen among species and organisms. In the T6SS with approximately 15 core components (T6SS-15), eighteen *S. marcescens* genomes show synteny thus forming four groups with the same genetic organization (Fig. 4A). Contrary to what was expected, these synteny groups are not composed by strains from the same genomic subgroup highlighted in figure 1. The same result was observed in both syntenic groups in the second T6SS loci (Fig. 4B). The *S. marcescens* strains B3R3, UMH2, UMH7, FS14 and WW4 have the same organization pattern for both T6SS loci. Among the *S. marcescens* strains, the genes *VgrG, VCA0113, VCA0114, vasF, vasK, impM, VCA0119* and *VCA0107* follows the same organization pattern in T6SS-15 (Fig. 4A). Strain RSC-14 is the only *S. marcescens* that does not have the T4SS *virB7* gene flaking the *VCA0113* gene. The posterior half of this locus varies in number and position of non-core gene components. For instance, strain U36365 is the only that has the gene *vasK* duplicated. The genes *VCA0113, VCA0114, vasF* are present and together in all *Serratia* spp. genomes for the T6SS-15 locus, whereas *S. plymuthica* PRI-2C is the only with a non-core gene between *VCA0114* and *vasF* genes. *Serratia fonticola* strain DSM4576 is the only microorganism that does not harbor a copy of *VgrG* gene in the T6SS-15 locus and, interestingly, it is also the only one *vasH* gene in both T6SS loci. VasH is a well known T6SS core component responsible for this system's transcriptional regulation (55). Five "exogenous" genes were found in the both T6SS loci: *virB7*, *virB11* and *traP* from T4SS and *fleS* amd *motB* from NF-T3SS. The genes *VgrG* and *vasH* from T6SS were found in one T4SS cluster (Fig. 4C), *vasH* was also found in the *S. rubidaea* T3SS clusters (Fig. 4D). Both genes *fleS* and *motB* were not present in *Serratia* spp. NF-T3SS.

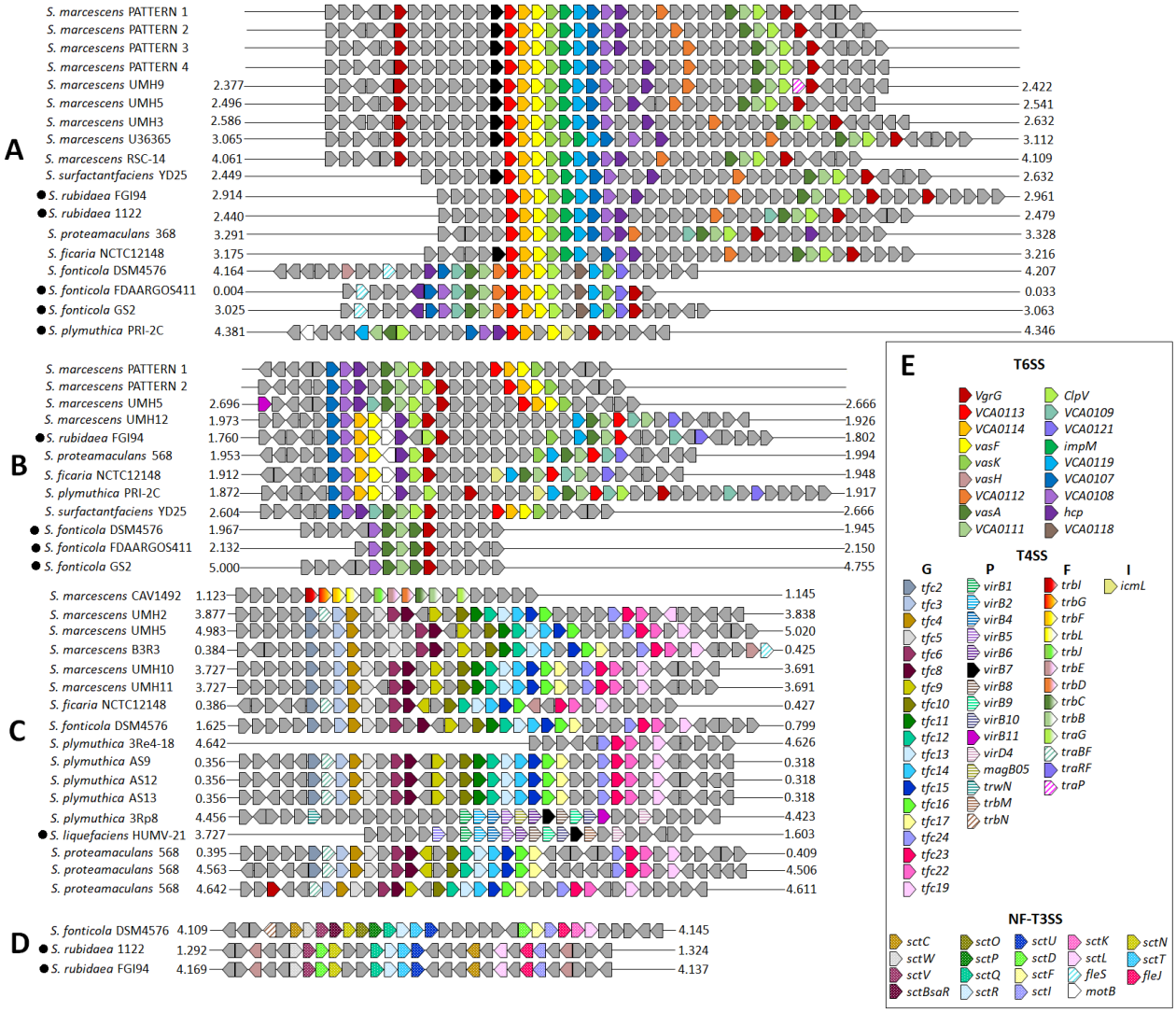

FIG 4 Complexity of *Serratia* spp. T6SS, T4SS and NF-T3SS genetic arrangement. (A) Type VI secretion system with approximately 15 core components, T6SS-15. Strains in pattern 1: N4-5 and UMH1. Strains in pattern 2: UMH8, UMH10, UMH11 and UMH12. Strains in pattern 3: B3R3, UMH2, UMH7, FS14, SCBI and WW4. Strains in pattern 4: SM39, SMB2099, smUNAM836 and UMH9. (B) Type VI secretion system with approximately 11 core components, T6SS-11. Strains in pattern 1: B3R3, RSC-14, UMH2, UMH6, UMH7, FS14 and WW4. Strains in pattern 2: U36365, CAV1492 and SCBI. (C) Type IV secretion system factors G, P, F and I. (D) Non-flagellar type III secretion system. (E) Color code for the genes in each secretion system cluster. Dark gray represents genes in the context, non-core components of the secretion system. Bold dots indicate which clusters are not in Genomic Islands. Numbers flanking the clusters indicate the genome positions (Mb) of the secretion system identified.

Similar to the type VI SS, the type IV also showed an inter and intraspecies variation. There are four T4SS factors present in *Serratia* strains, namely G-, P-, F- and I-type systems. The most frequent factor is T4SS-G, 80% of the genomes. The F-type system only one organism and the P- in two. However, the gene *traBF* from F-factor was only found in microrganisms whith G-factor. The *icmL* gene is only one I-type system

and it was found in two strains in T6SS-11 and one in T6SS -15. *S. proteamaculans* 568 is the only genome that has more than one copy (three) of T4SS. Thirty nine genomes harbor at least one type IV or VI secretion systems and only three genomes have the non-flagellar type III secretion system apparatus, namely *S. fonticola* DSM4576 and both *S. rubidaea* strains. The *S. rubidaea* isolates have exact same organization of the non-flagellar T3SS. The strain *S. fonticola* DSM4567 is the only that has the three secretion system types studied here, from wich only the type VI was not found in genomic island.

Among the plant beneficial isolates only two have both T6SS and T4SS (*S. fonticola* DSM4567 and *S. proteamaculans* 568), the others have either one or the other except the strains *S. plymuthica* S13 and 4Rx13 that do not have T6SS and T4SS machineries. Ninety percent of clinical isolates harbor at least one T6SS locus and 46.67% of genomes with T4SS are clinical. On the other hand, the insect pathogenic strains have only the T6SS. Our results show the presence of T6SS in most lifestyles, namely clinical, plant pathogen, plant associated, insect pathogen, symbiont and environmental (Table 2).

***Serratia* genomes code multiple secondary metabolites BGC with wide range of biological functions.** An *in silico* prediction of secondary metabolite biosynthetic gene clusters (BGC) revealed 17 potentially produced metabolites. Among these, seven are antibiotics namely althiomycin, bacillomycin, microcin H47, PM100117/PM100118, prodigiosin, pyrrolnitrin, and zeamine (56-63). The antitumor macrolides PM100117/PM100118 were present in all 42 genomes studied. The BGC similarity was 21% for all strains, except *S. liquefaciens* HUMV-21 which was 15%. This means that 21% of genes found in *Serratia* genomes PM100117/PM100118 clusters are similar to those in the BGC database. The similarity of genes in the zeamine, pyrrolnitrin and althiomycin biosynthetic clusters is 100% for all the strains that code these antibiotics. Similarly, bacillomycin biosynthetic cluster in *Serratia* has 20% of the genes similar to the BGC database and it is coded by all *S. marcescens* strains. On the other hand, prodigiosin genes similarity varies from 12% to 100%. There is a negative correlation between prodigiosin, pyrrolnitrin and althiomycin. The strains that have genes in biosynthetic prodigiosin cluster do not harbor the pyrrolnitrin and althiomycin BGCs, and vice-versa. Microcin H47 is only coded by *S. marcescens* strains. Some strains do not have Microcin H47 BGC, including the strains FS14 and WW4 (same subgroup), UMH5 and all organisms in the fourth genomic subgroup (strains UMH7, UMH2, RSC-14, UMH6 and SCBI). The isolates *S. fonticola* GS2, DSM4576, FDAARGOS_411, *S*

liquefaciens ATCC27592, FDAARGOS_125, HUMV-21 and *S. proteamaculans* 568 do not harbor any other antibiotic gene cluster besides PM100117/PM100118.

Other metabolites are cell-associated compounds, such as capsular polysaccharide (CPS), colanic acid, lipopolysaccharide and O-antigen. These compounds are often related with pathogenesis, biofilm formation and cell adhesion (64-69). Capsular polysaccharide BGC is common throughout *Serratia* species. However, the isolate 3Re4-18 is the only *S. plymuthica* without this polysaccharide. The second subgroup within *S. marcescens*, strains SM39, SMB2099, smUNAM836, UMH3 and UMH9 does not code for this polysaccharide. The O-antigen BGC was found in 40 genomes from which antiSMASH identified two clusters in 24 organisms (FIG 5, Table 2).

The remaining six metabolites have different functions, such as siderophore (turnerbactin and vibriobactin) (70, 71), vitamin K2 (menaquinone), lipodepsipeptides (taxlllaid)(72); antioxidative carotenoids (APE Ec)(73) and volatile organic compound (sodorifen)(74). Taxlllaid, isolated from the enthomopathogenic bacteria *Xenorhabdus indica*, was present in all genomes except *S. marcescens* UMH5 and *S. rubidaea* strains FGI94 and 1122. Aryl polyenes (APE Ec) is present in all *Serratia* genomes; however, in *S. marcescens* the strain Db11 is the only one with this metabolite BGC. In the other species, *S. fonticola* strain DSM4576 is the only isolate without APE Ec genes. Strain FS14 is the only genome besides the *S. plymuthica* strains to code sodorifen and it has 50% similarity whereas all the *S. plymuthica* genomes have 100% similarity.

Sixteen metabolites were exclusively present in one organism. Here we highlight oocydin A coded by *S. plymuthica* strain 4Rx13 (91% similarity), an anti-tumor and anti-oomycete compound; indigoidine, a blue-pigment anti-oxidant with antimicrobial acitivity coded by *S. proteamaculans* 568 (80%) and toxoflavin, a phytotoxin with antibiotics properties coded by *S. ficaria* NCTC12148 with 50% of genes similar to BGC (75-78). The other 13 compounds similarity were under 40% (data not shown).

Additional advantageous features in *Serratia* spp. CRISPR arrays make up the bacterial defense mechanism against exogenous DNA/RNA, such as phages (48). We report the presence of this immune system in 10 *Serratia* strains, namely: *S. fonticola* DSM4576 (3 arrays), *S. marcescens* B3R3 (1), CAV1492 (2), N4-5 (2), *S. plymuthica* 4Rx13 (1), AS9 (1), PRI-2C (2), AS12 (1), AS13 (1) and *S. rubidaea* FGI94 (4). The arrays were not found in genomic islands; therefore, they are naturally occurring features in these *Serratia* genomes. However, some of these arrays were not reported in the genome annotation available at GenBank. Other questionable arrays were predicted in

these genomes cited above and in six others (*S. fonticola* FDAARGOS_411 and GS2, *S. marcescens* UMH12, FS14 and SCBI and *S. proteamaculans* 568).

Chitinases are enzymes that also play key role in bacterial survival in nature (79). These compounds provide bacteria that produce it great competitive advantage against fungi. Among the seven *Serratia* species studied here, five harbor the chitinase genes *chiA* and *chiB*. *Serratia fonticola* and *S. rubidaea* genomes do not harbor these genes. This means that 88% of *Serratia* strains potentially produce chitinases.

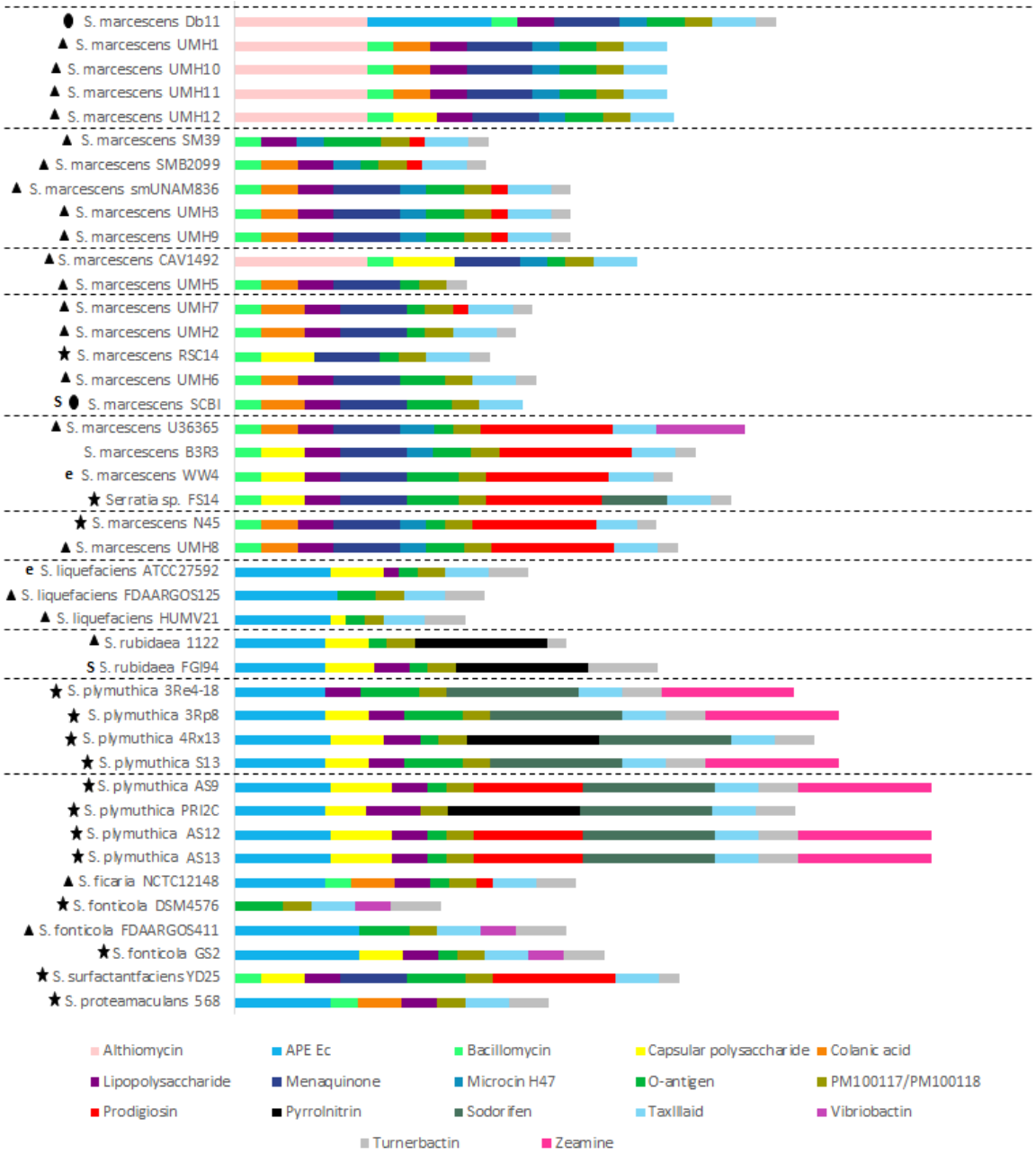

FIG 5 Secondary metabolites biosynthetic gene clusters distribution in *Serratia* spp. Bar size represents the percentage similarity of genes with BGCs. Strains between dashed lines have dDDH higher than 79%. Star symbol indicate strains associated and/or beneficial to plants, triangle are clinical strains, bold dots are entomopathogens, letters "S" and "e" are symbionts and environmental, respectively.

DISCUSSION

Bacteria are dynamic beings, able to modify constantly their gene content in order to adapt and to thrive in nature (80, 81). These modifications result in genome discrepancies beyond what is expected for interspecific variation but it is also observed in strains from the same species (82). Therefore, this intraspecific variation should be taken into consideration when using genome-based analysis to delineate species. Rosselló-Móra (83) when discussing prokaryotic taxonomy based on genomic metrics highlighted that the DDH value of 70% could not be used as an absolute cutoff for species delineation. According to this author, DDH values between 60 and 70% still encompass genomes belonging to same species, which our results showed to be true for *Serratia*.

Bacterial genome variation affects also the comparative analysis. We did not observe a clear-cut relationship between closely related *Serratia* genomes with their genetic organization of SSs, presence and similarity of secondary metabolite BGCs and CRISPR arrays and to the lifestyle each strain exhibits (Table 2). However, there is a trend in genetic organization of the secretion systems in strains from the same species. Even though Bacteria are promiscuous organisms - with the ability of gaining/losing genetic information fairly easy (84) - we still were able to find relatevily large conserved regions in the SS clusters. We propose that horizontally transferred secretion systems are adpatative acquisitions that provide competitive advantage to organisms that hold these apparatus. Similarly, there is also a trend in distribution of secondary metabolite BGC among strains from the same genomic subgroup. The discrepancy between genome content and phenotypes could be due to the fact that the strains were isolated from different sources and geographic locations. Yatsunenko et al (85) when analyzing human gut microbiome isolated from different sites and countries, observed pronounced variations in the assemblies and their proteomes. Therefore, the origin of each strain is a weighty factor in comparative genomics. Besides origin, the acessory genome that each organism carries can justify the variety of lifestyles exhibited. Overall, even though some species were not represented in this study the *Serratia* complete genomes available enabled a better understanding of the genus.

TABLE 2 Overall *Serratia* lifestyles distribution and their attributes.

Features \ Lifestyles	Clinical	Plant Pathogen	Plant Associated	Insect Pathogen	Symbiont	Animal Associated	Environmental
T6SS	19/21	1/1	08/15	2/2	2/2	0/0	1/2
T4SS	07/21	1/1	07/15	0/2	0/2	0/0	0/2
NF-T3SS	01/21	0/1	01/15	0/2	1/2	0/0	0/2
CRISPR	01/21	1/1	07/15	0/2	1/2	0/0	0/2
Chitinases	18/21	1/1	13/15	2/2	1/2	0/0	2/2
Althiomycin	05/21	0/1	00/15	1/2	0/2	0/0	0/2
Prodigiosin	09/21	1/1	06/15	0/2	0/2	0/0	1/2
Pyrrolnitrin	01/21	0/1	02/15	0/2	1/2	0/0	0/2
CPS	04/21	1/1	11/15	0/2	1/2	0/0	2/2
O-antigen	21/21	1/1	13/15	2/2	2/2	0/0	2/2

MATERIAL AND METHODS

Whole genome sequences. All complete genome sequences of strains classified within the genus *Serratia* that were available in GenBank database by 24th November of 2017 were used in the study (Table S1). Draft genomes were not included in the analysis, therefore the species *Serratia aquatilis, S. entomophila, S. glossinae, S. grimesii, S. myotis, S. nematodiphila, S. odorifera, S. oryzae* and *S. quinovorans* were not represented in this study. The nomenclature of the strains used is the same as it is on GenBank.

Phylogenetic and phylogenomic analysis. The nucleotide sequences of the genes 16S, *ftsz, groEL, groES, gyrB, recA, rpoD* of the 44 genomes were aligned using MAFFT online service (86, 87). Type strains were included only in the 16S analysis. The phylogenetic trees og the seven genes concatenated (MLSA) and for each gene individually were constructed in MEGA6 (88) using the maximum likelihood method with 1000 bootstrap replicates. The analysis of bacterial species based on whole genome sequence similarity was performed by comparing the strains using the genome-to-genome distance calculator 2.1 (http://ggdc.dsmz.de/distcalc2.php) to obtain DNA-DNA Hybridization *in silico* (dDDH), with cutoff of 60% for species delineation (4). A heatmap and the corresponding phylogram were constructed on R based on dDDH values from GGDC's formula 2. Additionally, the average nucleotide identity (ANI) based on BLAST and on MUMmer (ANIb and ANIm) were obtained by pairwise genome comparisons using the web-tool JspeciesWS (http://jspecies.ribohost.com/jspeciesws/)(89). For ANI analyzes the cutoff was 94% to delineate genomes belonging to the same species (7). The PEPR program (Phylogenomic Estimation with Progressive Refinement) was used to generate a phylogenomic tree. This is an automated system for generation of phylogenomic trees from amino acid sequences by a maximum likelihood algorithm

(RAxML). It identifies the common orthologs among all genomes, filters out genes that have been transferred horizontally, aligns, concatenates sequences, and generates a tree.

Pan- and core-genome definitions. The Bacterial Pan Genome Analysis tool (BPGA) was used to construct the updated *Serratia* Pan, Core and Accessory genome and its categorization into COG functions (90). BPGA calculates curve fitting using the power-law regression model for the pan-genome data ($Y_{pan} = A_{pan} \cdot x^{Bpan} + C_{pan}$) and an exponential curve fit model for the core-genome data ($Y_{core} = A_{core} \cdot e^{-Bcore \cdot x} + C_{core}$). B_{pan} parameter will estimate weather the pan-genome is open or closed, i.e $0 < B_{pan} < 1$ indicates the pan-genome is open (91). BPGA scripts were also used for atypical GC content analysis and exclusive gene absence. Pan and core-genome profile curves and COG distribution were plotted with gnuplot integrated in the BPGA pipeline.

Comparative genomics. The 42 *Serratia* spp. genomes were compared in search for differences and similarities regarding their production of secondary metabolites, secretion systems genetic organization, presence of chitinases and CRISPR arrays. The type III, IV and VI secretions systems (T3SS, T4SS and T6SS) were predicted by the software T346Hunter (92). The secondary metabolite biosynthetic gene clusters (BGC) were identified using antiSMASH 4.0.2 (93). From the 55 secondary metabolites identified, 17 were selected for analysis and are displayed in figure 5. The selection criteria were: to be present in at least 10% of the strains and the cluster should have at least 20% of the genes similar to those in the known biosynthetic cluster found in the BGC database. Chitinases sequences, *chiA* and *chiB*, were searched against the proteome of all 42 organisms using BLASTp (94). Chitinases *chiC* and *chiD* are 87% similar (100% cover) therefore to avoid biased results of these two genes were left out of the analysis. The CRISPR arrays and the genomic islands (GIs) were identified using the web-based tools CRISPRfinder and IslandViewer 3, respectively (95, 96). Some of the attributes identified above were compared (in terms of presence/absence) between strains exhibiting same lifestyle category, such as Clinical, Plant Pathogen, Plant Associated, Insect Pathogen, Symbiont, Animal Associated and Environmental (Table 2). Strains categorized into “Clinical” were either isolated from hospitalized patients or found to be human pathoges. Organisms associated with plants are endophytes, rhizobacteria, plant growth promoters, biocontrol agents and nitrogen-fixing bacteria. Strains associated symbiotically with fungi, insects, etc. were categorized as symbionts. “Animal Associated” are strains found in association with animals, for example mammals and

invertebrates. The “Environmental” category includes strains isolated from water, soil and air or that do not display any specific lifestyle described above.

REFERENCES

1. Land M, Hauser L, Jun S-R, et al. 2015 Insights from 20 years of bacterial genome sequencing. Functional & Integrative Genomics 15:141-161. doi:10.1007/s10142-015-0433-4.
2. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. 106:19126-31. doi: 10.1073/pnas.0906412106.
3. Hedge J, Wilson DJ. 2014. Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not. mBio 5:e02158-14. doi:10.1128/mBio.02158-14.
4. Auch AF, Von Jan M, Klenk H-P, Goker M. 2010. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Standards in genomic sciences 2:117-134. doi:10.4056/sigs.531120.
5. Thompson CC, Chimetto L, Edwards RA, Swings J, Stackebrandt E, Thompson FL. 2013. Microbial genomic taxonomy. BMC Genomics 14:913. doi:10.1186/1471-2164-14-913.
6. Konstantinidis KT, Tiedje JM. 2005. Towards a Genome-Based Taxonomy for Prokaryotes. Journal of Bacteriology 187:6258-6264. doi:10.1128/JB.187.18.6258-6264.2005.
7. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. PNAS 102:2567-2572. doi:10.1073/pnas.0409727102.
8. Chun J, Rainey FA. 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. International Journal of Systematic and Evolutionary Microbiology 64:316–324. doi:10.1099/ijs.0.054171-0.
9. Garrity GM. 2016. A new genomics-driven taxonomy of Bacteria and Archaea: are we there yet? J Clin Microbiol 54:1956–1963. doi:10.1128/JCM.00200-16.
10. Colston SM, Fullmer M, Beka L, Lamy B, Gogarten JP, Graf J. 2014. Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case. mBio 5(6):e02136-14. doi:10.1128/mBio.02136-14.
11. Safni I, Cleenwerck I, De Vos P, Fegan M, Sly L, Kappler U. 2014. Polyphasic taxonomic revision of the Ralstonia solanacearum species complex: proposal to emend the descriptions of Ralstonia solanacearum and Ralstonia syzygii and reclassify current R. syzygii strains as Ralstonia syzygii subsp. syzygii subsp. nov., R. solanacearum phylotype IV strains as Ralstonia syzygii subsp. indonesiensis subsp. nov., banana blood disease bacterium strains as Ralstonia syzygii subsp. celebesensis subsp. nov. and R. solanacearum phylotype I and III strains as Ralstonia pseudosolanacearum sp. nov. Int J Syst Evol Microbiol 64:3087-3103. doi:10.1099/ijs.0.066712-0.
12. Kishi LT, Fernandes CC, Omori WP, Campanharo JC, Lemos EG de M. 2016. Reclassification of the taxonomic status of SEMIA3007 isolated in Mexico B-11A Mex as Rhizobium leguminosarum bv. viceae by bioinformatic tools. BMC Microbiology 16:260. doi:10.1186/s12866-016-0882-5.

13. Li P, Kwok AHY, Jiang J, Ran T, Xu D, Wang W, et al. 2015. Comparative Genome Analyses of Serratia marcescens FS14 Reveals Its High Antagonistic Potential. PLoS ONE 10: e0123061. doi:10.1371/journal.pone.0123061.
14. Beja O, Koonin EV, Aravind L, Taylor LT, Seitz H, Stein JL, Bensen DC, Feldman RA, Swanson RV, DeLong EF. 2002. Comparative Genomic Analysis of Archaeal Genotypic Variants in a Single Population and in Two Different Oceanic Provinces. Applied And Environmental Microbiology 335–345. doi:10.1128/AEM.68.1.335–345.2002.
15. Makarova KS, Koonin EV. 2003. Comparative genomics of archaea: how much have we learned in six years, and what's next? Genome Biology 4:115. doi:10.1186/gb-2003-4-8-115.
16. Fouts DE, Matthias MA, Adhikarla H, et al. 2016. What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira. Small PLC, ed. PLoS Neglected Tropical Diseases 10:e0004403. doi:10.1371/journal.pntd.0004403.
17. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL. 2017. Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin Microbiol Rev 30:1015–1063. https://doi.org/10.1128/ CMR.00016-17.
18. Ursua PR, Unzaga MJ, Melero P, Iturburu I, Ezpeleta C, Cisterna R. 1996. Serratia rubidaea as an invasive pathogen. J Clin Microbiol 34: 216–217.
19. Grimont F, Grimont PAD. 2006. The Genus Serratia, p 219-244. *In* Dworkin M., Falkow S., Rosenberg E., Schleifer KH., Stackebrandt E. (eds) The Prokaryotes. Springer, New York, NY.
20. Wang X-Q, Bi T, Li X-D, Zhang L-Q, Lu S-E. 2015. First Report of Corn Whorl Rot Caused by Serratia marcescens in China. J Phytopathol 163:1059–1063.
21. Fürnkranz M, Adam E, Müller H, Grube M, Huss H, Winkler J, Berg G. 2012. Promotion of growth, health and stress tolerance of Styrian oil pumpkins by bacterial endophytes. Eur. J. Plant Pathol. 134:– 509–519.
22. Saha D, Purkayastha GD, Saha A. 2012. Biological Control of Plant Diseases by Serratia Species: A Review or a Case Study, p 99-115. *In* Frontiers on Recent Developments in Plant Science, vol 1.
23. Khanna A, Khanna M, Aggarwal A. 2013. Serratia Marcescens- A Rare Opportunistic Nosocomial Pathogen and Measures to Limit its Spread in Hospitalized Patients. Journal of Clinical and Diagnostic Research 7:243-246. doi:10.7860/JCDR/2013/5010.2737.
24. Gerc AJ, Song L, Challis GL, Stanley-Wall NR, Coulthurst SJ. 2012. The Insect Pathogen Serratia marcescens Db10 Uses a Hybrid Non-Ribosomal Peptide Synthetase-Polyketide Synthase to Produce the Antibiotic Althiomycin. Cornelis P, ed. PLoS ONE7:e44673. doi:10.1371/journal.pone.0044673.
25. Rascoe, J.; Berg, M.; Melcher, U.; Mitchell, F.; Bruton, B. D.; Pair, S. D.; Fletcher, J. Identfication, phylogenetic analysis and biological characterization of Serratia marcescens causing cucurbit yellow vine disease. Phytopathology, v. 93, p. 1233–1239, 2003.
26. Roberts DP, Lakshman DK, McKenna LF, Emche SE, Maul JE, Bauchan G. 2016. Seed treatment with ethanol extract of Serratia marcescens is compatible with Trichoderma isolates for control of damping-off of cucumber caused by Pythium ultimum. Plant Dis 100:1278-1287.

27. Matsuyama T., Tanikawa T., Nakagawa Y. 2011. Serrawettins and Other Surfactants Produced by Serratia. *In* Soberón-Chávez G. (eds) Biosurfactants. Microbiology Monographs, vol 20. Springer, Berlin, Heidelberg.
28. Fadhil L, Kadim A, Mahdi A. 2014. Production of Chitinase by Serratia marcescens from Soil and Its Antifungal Activity. Journal of Natural Sciences Research 4:80-86.
29. Alcoforado D J, Coulthurst SJ. 2015. Intraspecies competition in Serratia marcescens is mediated by type VI-secreted Rhs effectors and a conserved effector-associated accessory protein. J Bacteriol 197:2350–2360. doi:10.1128/JB.00199-15.
30. Di Venanzio G, Stepanenko TM, García Véscovi E. 2014. Serratia marcescens ShlA Pore-Forming Toxin Is Responsible for Early Induction of Autophagy in Host Cells and Is Transcriptionally Regulated by RcsB. Payne SM, ed. Infection and Immunity 82:3542-3554. doi:10.1128/IAI.01682-14.
31. Kalbe C, Marten P, Berg G. 1996. Strains of the genus Serratia as beneficial rhizobacteria of oilseed rape with antifungal properties. Microbiol Res 151:433-439.
32. Williams RP. 1973. Biosynthesis of Prodigiosin, a Secondary metabolite of Serratia marcescens. Applied Microbiology 25:396-402.
33. Kurbanoglu EB, Ozdal M, Ozdal OG, Algur OF. 2015. Enhanced production of prodigiosin by Serratia marcescens MO-1 using ram horn peptone. Brazilian Journal of Microbiology 46:631-637. doi:10.1590/S1517-838246246220131143.
34. Montaner B, Perez-Thomas R. 2001. Prodigiosin induced apoptosis in human colon cancer cells. Life Sci 68:2025-2036.
35. Patil CD, Patil SV, Salunke BK, et al. 2011. Prodigiosin produced by Serratia marcescens NMCC46 as a mosquito larvicidal agent against Aedes aegypti and Anopheles stephensi. Parasitol Res 109:1179. doi:10.1007/s00436-011-2365-9.
36. Rahul S, Chandrashekhar P, Hemant B, Chandrakant N, Laxmikant S, Satish P. 2014. Nematicidal activity of microbial pigment from Serratia marcescens. Natural Product Research 28.
37. Soto-Cerrato V, Llagostera E, Montaner B, et al. 2004. Mitochondria-mediated apoptosis operating irrespective of multidrug resistance in breast cancer cells by the anticâncer agent prodigiosin. Biochem Pharmacol 68:1345-1352.
38. Suryawanshi RK, Patil CD, Borase HP, Narkhede CP, Salunke BK, Patil SV. 2015. Mosquito larvicidal and pupaecidal potential of prodigiosin from Serratia marcescens and understanding its mechanism of action. Pesticide Biochemistry and Physiology 123:49–55.
39. Lapenda JC, Silva PA, Vicalvi MC. et al. 2015. Antimicrobial activity of prodigiosin isolated from Serratia marcescens UFPEDA 398. World J Microbiol Biotechnol 31:399-406. doi.org/10.1007/s11274-014-1793-y.
40. Harris AKP, et al. 2004. The Serratia gene cluster encoding biosynthesis of the red antibiotic, prodigiosin, shows species- and strain-dependent genome context variation. Microbiology 150:3547–3560.
41. Li P, Kwok AHY, Jiang J, Ran T, Xu D, Wang W, et al. 2015. Comparative Genome Analyses of Serratia marcescens FS14 Reveals Its High Antagonistic Potential. PLoS ONE 10:e0123061. doi:10.1371/journal.pone.0123061.
42. Murdoch SL, Trunk K, English G, Fritsch MJ, Pourkarimi E, Coulthurst SJ. 2011. The opportunistic pathogen Serratia marcescens utilizes type VI secretion to target bacterial competitors. J Bacteriol 193:6057-6069.

43. Gerc AJ, Diepold A, Trunk k, Porter M, Rickman C, Armitage JP, Stanley-Wall NR, Coulthurst SJ.. 2015. Visualization of the Serratia Type VI Secretion System Reveals Unprovoked Attacks and Dynamic Assembly. Cell Rep 12:2131-2142
44. Masum MMI, Yang Y, Li B, et al. 2017. Role of the Genes of Type VI Secretion System in Virulence of Rice Bacterial Brown Stripe Pathogen Acidovorax avenae subsp. avenae Strain RS-2. International Journal of Molecular Sciences 18:2024. doi:10.3390/ijms18102024.
45. Green ER, Mecsas J. 2016. Bacterial Secretion Systems – An overview. Microbiology spectrum 4(1):VMBF-0012-2015. doi:10.1128/microbiolspec.VMBF-0012-2015.
46. Cianfanelli FR, Monlezun L, Coulthurst SJ. 2016. Aim, Load, Fire: The Type VI Secretion System, a Bacterial Nanoweapon. Trends in Microbiology 24:51-62.
47. Byndloss MX, Rivera-Chávez F, Tsolis RM, Baumlr AJ. 2016. How bacterial pathogens use type III and type IV secretion systems to facilitate their transmission. Current Opinion in Microbiology 35:1–7.
48. Sorek R, Kunin V, Hugenholtz P. 2008. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat. Rev. Microbiol 6:181. doi:10.1038/nrmicro1793 pmid:18157154.
49. Barrangou R, Marraffini LA. 2014. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Molecular cell 54:234-244. doi:10.1016/j.molcel.2014.03.011.
50. Deveau H, Garneau JE, Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64:475-93. doi:10.1146/annurev.micro.112408.134123.
51. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2005. GenBank**.** Nucleic Acids Research 33:D34–D38.
52. Su C, Xiang Z, Liu Y, et al. 2016. Analysis of the genomic sequences and metabolites of Serratia surfactantfaciens sp. nov. YD25T that simultaneously produces prodigiosin and serrawettin W2. BMC Genomics 17:865. doi:10.1186/s12864-016-3171-7.
53. Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy T, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk H-P. 2014. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. Standards in Genomic Sciences 10:2.
54. Basharat Z, Yasmin, A. 2016. Pan-genome Analysis of the Genus Serratia. arXiv preprint arXiv:1610.04160.
55. Kitaoka M, Miyata ST, Brooks TM, Unterweger D, Pukatzki S. 2011. VasH is a transcriptional regulator of the type VI secretion system functional in endemic and pandemic Vibrio cholerae. J Bacteriol 193:6471-82. doi:10.1128/JB.05414-11.
56. Arima K, Imanaka H, Kousaka M, Fukuda A, Tamura G. 1964. Pyrrolnitrin, a new antibiotic substance, produced by Pseudomonas**.** Agric Biol Chem 28:575–576.
57. Wu J, Zhang HB, Xu JL, Cox RJ, Simpson TJ, Zhang LH. 2010. 13C Labeling reveals multiple amination reactions in the biosynthesis of a novel polyketide polyamine antibiotic zeamine from Dickeya zeae. Chem Commun 46:333-335.
58. Zhou J, Zhang HB, Wu J, Liu Q, Xi P, Lee J, Liao J, Jiang Z, Zhang LH. 2011. A novel multidomain polyketide synthase is essential for zeamine production and the virulence of Dickeya zeae. Mol Plant-Microbe Interact 24:1156-1164.

59. Rapoport H, Holden HG. 1962. The synthesis of prodigiosin. J Am Chem Soc 84:635-42.
60. Laviña M, Gaggero C, Moreno F. 1990. Microcin H47, a chromosome-encoded microcin antibiotic of Escherichia coli. J Bacteriol 172:6585-8.
61. Peypoux F, Marion D, Maget-Dana R, Ptak M, Das BC, Michel G. 1985. Structure of bacillomycin F, a new peptidolipid antibiotic of the iturin group. Eur J Biochem 153:335-40.
62. Yamaguchi H, Nakayama Y, Takeda K, Tawara K, Maeda K, Takeuchi T, Umezawa H. 1957. A new antibiotic, althiomycin. J Antibiot 10:195-200.
63. Perez M, Schleissner C, Fernández R, Rodríguez P, Reyes F, Zuñiga P, Calle F de la, Cuevas C. 2016. PM100117 and PM100118, new antitumor macrolides produced by a marine Streptomyces caniferus GUA-06-05-006ª. The Journal of Antibiotics 69:388–394.
64. Beck von Bodman S, Farrand SK. 1995. Capsular polysaccharide biosynthesis and pathogenicity in Erwinia stewartii require induction by an N-acylhomoserine lactone autoinducer. J Bacteriol 177:5000-8. PMID: 7665477.
65. Roy D, Auger JP, Segura M, Fittipaldi N, Takamatsu D, Okura M, Gottschalk M. 2015. Role of the capsular polysaccharide as a virulence factor for Streptococcus suis serotype 14. Can J Vet Res 79:141-6.
66. Galanos, C. 1998. Endotoxin (Lipopolysaccharide (LPS)), p. 806-809. *In* Encyclopedia of Immunology, 2nd ed, Elsevier.
67. Hanna A, Berg M, Stout V, Razatos A. 2003. Role of Capsular Colanic Acid in Adhesion of Uropathogenic Escherichia coli. Appl Environ Microbiol 69:4474–4481. doi:10.1128/AEM.69.8.4474-4481.2003
68. Valvano MA. 1992. Pathogenicity and molecular genetics of O-specific side-chain lipopolysaccharides of Escherichia coli. Can J Microbiol 38:711-9.
69. Anderson MT, Mitchell LA, Zhao L, Mobley HLT. 2017. Capsule production and glucose metabolism dictate fitness during Serratia marcescens bacteremia. mBio 8:e00740-17. https://doi.org/10.1128/mBio.00740-17.
70. Griffiths GL, Sigel SP, Payne SM, Neilands JB. 1984. Vibriobactin, a siderophore from Vibrio cholerae. J Biol Chem 259:383-5. PMID:6706943.
71. Han AW, Sandy M, Fishman B, Trindade-Silva AE, Soares CA, Distel DL, Butler A, Haygood MG. 2013. Turnerbactin, a novel triscatecholate siderophore from the shipworm endosymbiont Teredinibacter turnerae T7901. PLoS One 8:e76151. doi:10.1371/journal.pone.0076151.
72. Kronenwerth M, Bozhüyük KA, Kahnt AS, Steinhilber D, Gaudriault S, Kaiser M, Bode HB. 2014. Characterisation of taxlllaids A-G; natural products from Xenorhabdus indica. Chemistry 20:17478-87.
73. Schoner TA, Gassel S, Osawa A, Tobias NJ, Okuno Y, Sakakibara Y, Shindo K, Sandmann G, Bode HB. 2016. Aryl Polyenes, a Highly Abundant Class of Bacterial Natural Products, Are Functionally Related to Antioxidative Carotenoids. ChemBioChem 17:247–253.
74. von Reuß, SH, Kai M, Piechulla B, Francke, W. 2010. Octamethylbicyclo[3.2.1]octadienes from the Rhizobacterium Serratia odorifera. Angew. Chem 122:2053–2054. doi:10.1002/ange.200905680.
75. Nagamatsu T. 2001. Syntheses, transformation, and biological activities of 7-azapteridine antibiotics: toxoflavin, fervenulin, reumycin and their analogs. Recent Res Devel Org Bioorg Chem 4:97–121.
76. Latuasan HE, Berends W. 1961. On the origin of the toxicity of toxoflavin. Biochim Biophys Acta 52:502–508.

77. Reverchon S, Rouanet C, Expert D, Nasser W. 2002. Characterization of indigoidine biosynthetic genes in Erwinia chrysanthemi and role of this blue pigment in pathogenicity. J Bacteriol 184:654–665.
78. Srobel G, Li JY, Sugawara F, Koshino H, Harper J, Hess WM. 1999. Oocydin A, a chlorinated macrocyclic lactone with potent anti-oomycete activity from Serratia marcescens. Microbiology 145:3557-64.
79. Hamid R, Khan MA, Ahmad M, et al. 2013. Chitinases: An update. Journal of Pharmacy & Bioallied Sciences 5:21-29. doi:10.4103/0975-7406.106559.
80. Ochman H, Davalos LM. 2006. The nature and dynamics of bacterial genomes. Science 311:1730–1733. doi: 10.1126/science.1119966.
81. Cohan FM, Koeppel AF. 2008. The origins of ecological diversity in prokaryotes. Curr Biol 18:R1024-34. doi:10.1016/j.cub.2008.09.014.
82. Lan R, Reeves PR. 2000. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol 8:396-401.
83. Rosselló-Móra R. 2006. DNA-DNA reassociation methods applied to microbial taxonomy and their critical evaluation, p 23–50. *In* Stackebrandt E, editor. Molecular Identification, Systematics, and Population Structure of Prokaryotes. Berlin: Springer-Verlag.
84. Thomas CM, Nielsen KM. 2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol 3:711-21.
85. Yatsunenko T, Rey FE, Manary MJ, et al. 2012 Human gut microbiome viewed across age and geography. Nature 486:222-227. doi:10.1038/nature11053.
86. Katoh, Rozewicki, Yamada 2017 (Briefings in Bioinformatics, in press). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization.
87. Katoh, Misawa, Kuma, Miyata 2002 (Nucleic Acids Res. 30:3059-3066) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.
88. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30:2725-2729.
89. Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. 2015. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929-31. doi:10.1093/bioinformatics/btv681.
90. Chaudhari NM, Gupta VK, Dutta C. 2016. BPGA – an ultra-fast pan-genome analysis pipeline. Scientific Reports 6:24373.
91. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, et al. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190:6881–6893. doi:10.1128/JB.00619-08.
92. Martínez-García PM, Ramos C, Rodríguez- Palenzuela P. 2015. T346Hunter: A Novel Web-Based Tool for the Prediction of Type III, Type IV and Type VI Secretion Systems in Bacterial Genomes. PLoS ONE 10:e0119317. doi:10.1371/journal. pone.011931.
93. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, et al. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters, Nucleic Acids Research 43:W237–W243.
94. Gish W, States DJ. 1993. "Identification of protein coding regions by database similarity search." Nature Genet 3:266-272.

95. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52-7.
96. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, Pereira SK, Waglechner N, Mcarthur AG, Langille MG, Brinkman FS. 2015. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Research 43:W104-W108. doi: 10.1093/nar/gkv401.

APENDIX A – Supplementary material Article 1

Supplementary Table S1

Table S1. Genomic relatedness values between the strain N4-5 and other *Serratia* spp. genomes calculated with genomic metrics.

Organism	DDH	ANI	AAI
S. marcescens B3R3	**74.2**	**96.41**	**97.66**
S. marcescens U36365	**74.3**	**96.58**	**97.82**
S. marcescens UMH8	**90.3**	**98.68**	**99.06**
S. marcescens WW4	**72.8**	**96.40**	**96.15**
S. liquefaciens ATCC 27592^T	34.8	82.88	85.77
S. ficaria NCTC12148^T	24.8	87.06	87.65
S. fonticola DSM4567^T	27.2	79.91	79.61

Values in percentage. ANI was calculated based on BLAST and DDH on formula 2. Values in bold indicate they are above the threshold to belong to same species, namely 70, 95 and 95%.

Supplementary Table S2

Table S2. List of the genomes with extra copies of the 5S rRNA gene.

Species	Strain	Acession nº	Source	Posição
S. marcescens	B3R3	CP013046.2	Wang et al., 2015	5º operon C
S. marcescens	CAV1492	CP011642.1	Sheppard et al., 2015, unpublished	6º operon
S. marcescens	Db11	NZ_HG326223.1	Iguchi et al., 2014	3º operon C
S. marcescens	RSC-14	CP012639.1	Khan et al., 2017	6º operon
S. marcescens	SM39	NZ_AP013063.1	Iguchi et al., 2014	3º operon C
S. marcescens	SMB2099	NZ_HG738868.1	Yao et al., 2017, unpublished	5º operon C
S. marcescens	SmUNAM836	CP012685.1	Sandner-Miranda et al., 2016	3º operon C
S. marcescens	U36365	CP016032.1	Sahni et al., 2016	5º operon C
S. marcescens	UMH1	NZ_CP018915.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH2	NZ_CP018924.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH3	NZ_CP018925.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH5	NZ_CP018917.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH6	NZ_CP018926.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH7	NZ_CP018919.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH8	NZ_CP018927.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH9	NZ_CP018923.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH10	NZ_CP018928.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH11	NZ_CP018929.1	Anderson et al., 2017	3º operon C
S. marcescens	UMH12	NZ_CP018930.1	Anderson et al., 2017	3º operon C
S. marcescens	WW4	NC_020211.1	Chung et al., 2013	5º operon C
S. marcescens	FDAARGOS_65	NZ_CP026050.1	Sichtig et al., 2018, unpublished	6º operon
S. marcescens	AR_0027	NZ_CP026702.1	Colan et al., 2018, unpublised	6º operon C
S. marcescens	AR_0091	NZ_CP027533.1	Colan et al., 2018, unpublised	3º operon C
S. marcescens	AR_0099	NZ_CP027539.1	Colan et al., 2018, unpublised	6º operon C
S. marcescens	AR_0124	NZ_CP028946.1	Colan et al., 2018, unpublised	2º operon C
S. marcescens	AR_0130	NZ_CP028947.1	Colan et al., 2018, unpublised	5º operon C
S. marcescens	AR_0123	NZ_CP028948.1	Colan et al., 2018, unpublised	6º operon
S. marcescens	AR_0121	NZ_CP028949.1	Colan et al., 2018, unpublised	6º operon
S. ficaria	NCTC12148	NZ_LT906479.1	NCTC	5º operon C

S. fonticola	DSM 4576	NZ_CP011254.1	Lim et al., 2015	2º operon C
S. fonticola	FDAARGOS_411	NZ_CP023956.1	Minogue et al., 2017, unpublished	6º operon C
S. fonticola	GS2	CP013913.1	Jung et al., 2016, unpublished	1º e 8º operon C
S. liquefaciens	ATCC 27592	CP006252.1	Nicholson et al., 2013	4º operon C
S. liquefaciens	FDAARGOS_125	CP014017.1	Goldberg et al., 2016, unpublished	6º operon
S. liquefaciens	HUMV-21	NZ_CP011303.1	Lázaro-Diez et al., 2015	1º operon
S. plymuthica	3Re4-18	CP012097.1	Adam et al., 2016	6º operon C
S. plymuthica	3Rp8	CP012096.1	Adam et al., 2016	6º operon
S. plymuthica	4Rx13	CP006250.1	Weise et al., 2013	5º operon C
S. plymuthica	AS9	CP002773.1	Neupane et al., 2012a	5º operon C
S. plymuthica	PRI-2c	NZ_CP015613.1	Schmidt et al., 2017	5º operon C
S. plymuthica	S13	CP006566.1	Muller et al., 2013	5º operon C
S. proteamaculans	568	CP000826.1	Copeland et al., 2007, unpublished	5º operon C
S. rubidaea	1122	CP014474.1	Yao et al., 2016	5º operon C
S. rubidaea	FGI94	CP003942.1	Aylward et al., 2013	5º operon C
Serratia sp.	AS12	CP002774.1	Neupane et al., 2012b	5º operon C
Serratia sp.	AS13	CP002775.1	Neupane et al., 2012c	5º operon C
Serratia sp.	YD25	NZ_CP016948.1	Su et al., 2016	3º operon C

Supplementary Figure S1

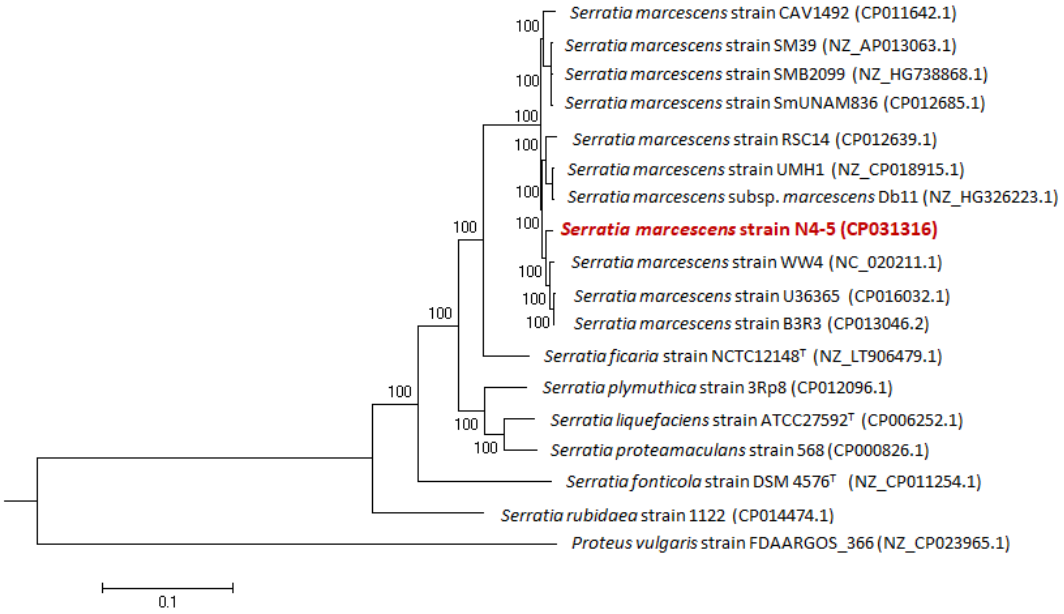

Figure S1. Phylogenomic tree constructed with *Serratia* spp. complete genome sequences. The tree was constructed using the Phylogenetic Tree Building Service available at PATRIC, using the RAxML algorithm (Stamakis 2006).

APENDIX B – Supplementary material Article 2

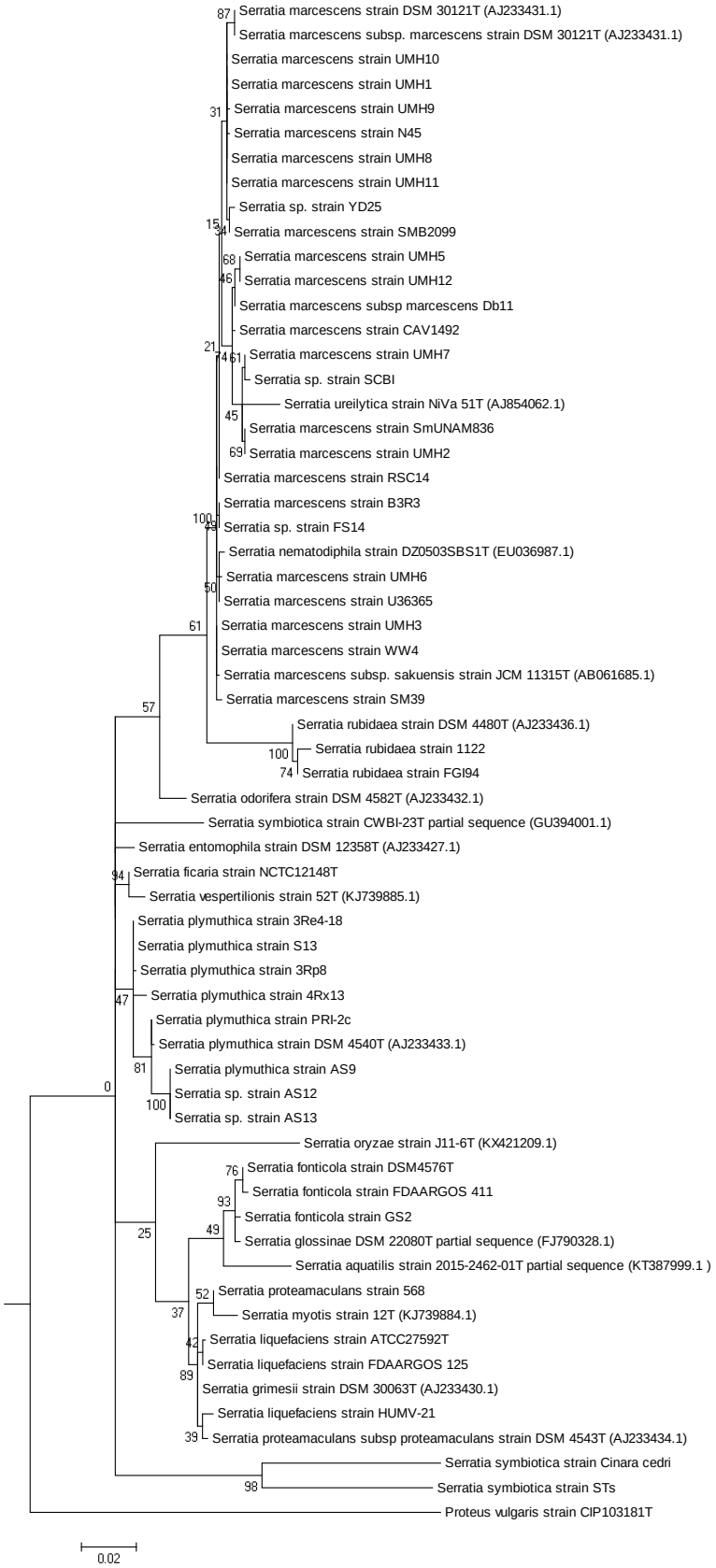

FIG S1 Phylogenetic tree based on nucleotide sequences of the 16S gene. The tree construction method was Maximum Likelihood with the Kimura 2-parameter model incorporating invariant sites and gamma distribution.

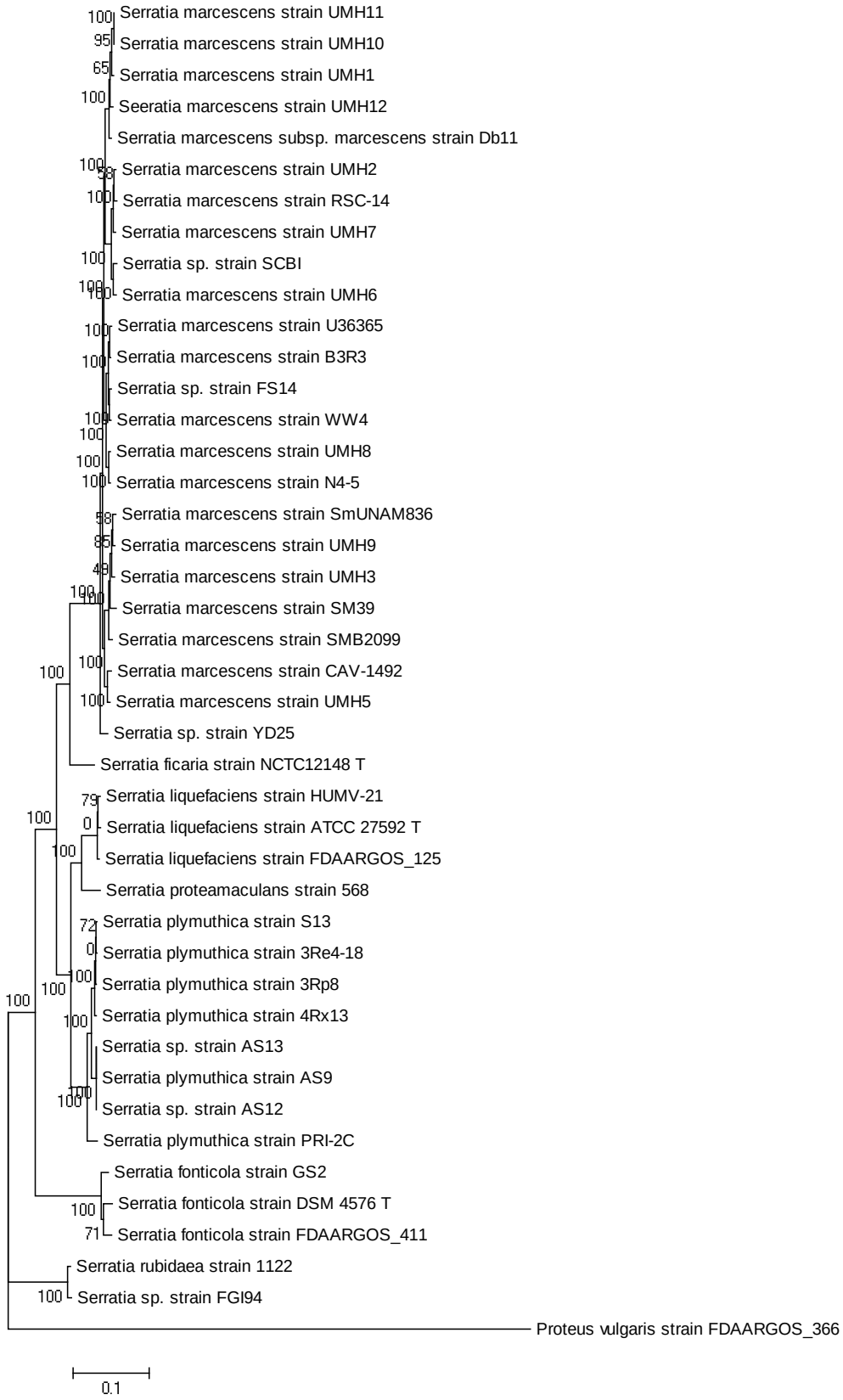

FIG S2 Phylogenomic tree constructed with 42 *Serratia* spp. complete genomes sequences. The tree was constructed using the Phylogenetic Tree Building Service available at PATRIC, using the RAxML algorithm (Stamakis 2006).

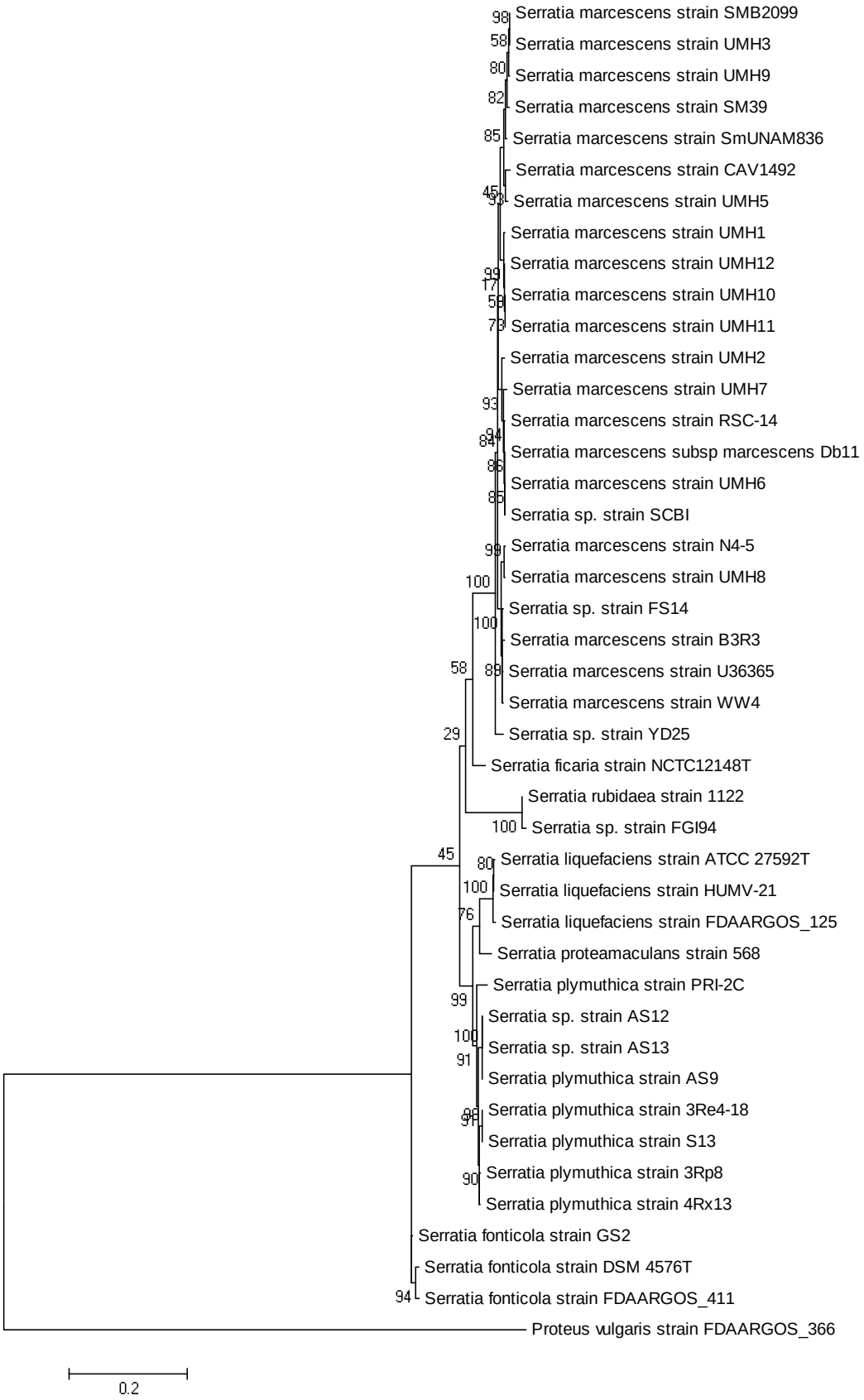

FIG S3 Phylogenetic tree based on nucleotide sequences of the *ftsz* gene. The tree construction method was Maximum Likelihood with the Kimura 2-parameter model incorporating invariant sites and gamma distribution.

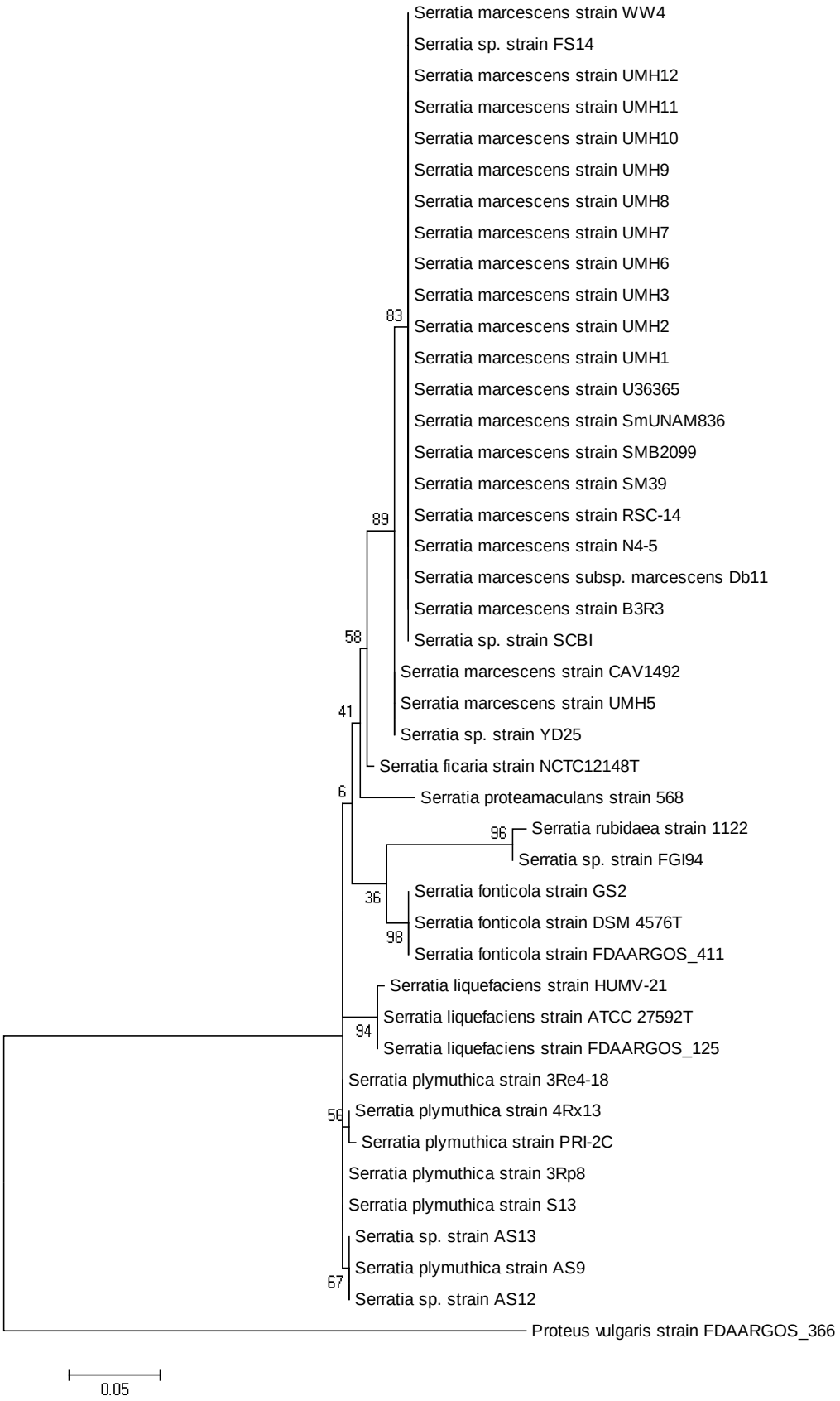

FIG S4 Phylogenetic tree based on nucleotide sequences of the *groES* gene. The tree construction method was Maximum Likelihood with the Kimura 2-parameter model with gamma distribution.

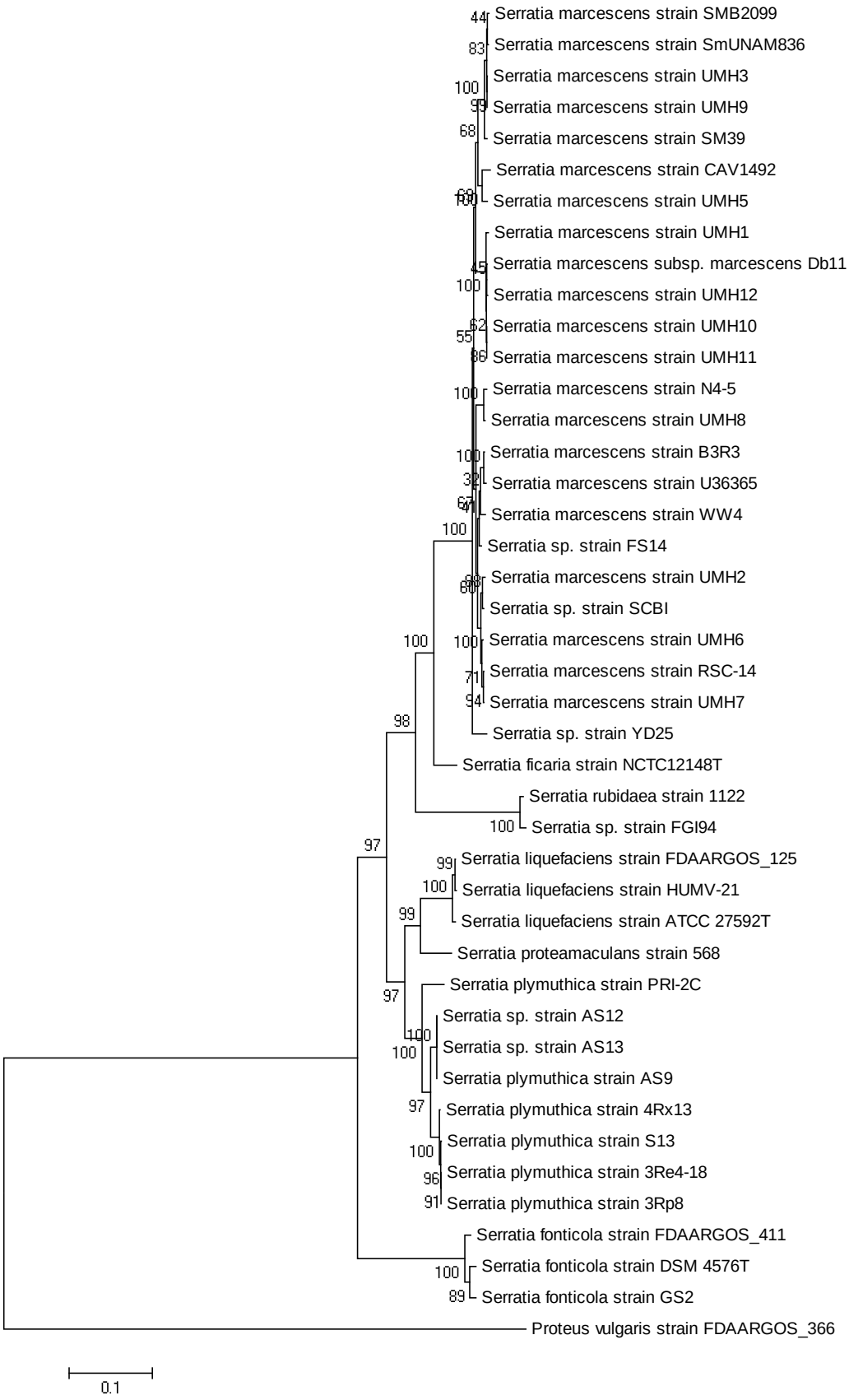

FIG S5 Phylogenetic tree based on nucleotide sequences of the *gyrB* gene. The tree construction method was Maximum Likelihood with the TN93 model incorporating invariant sites and gamma distribution.

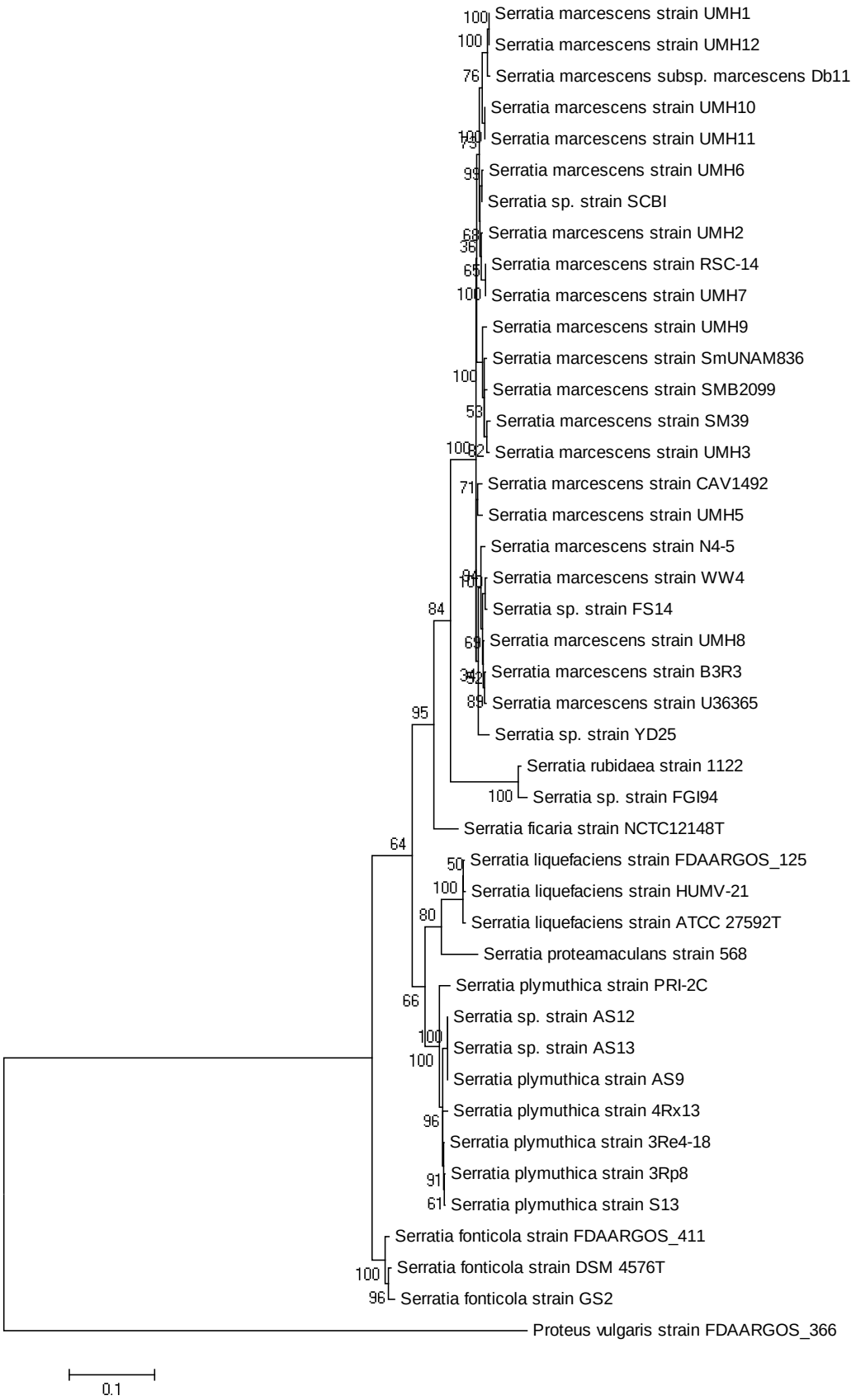

FIG S6 Phylogenetic tree based on nucleotide sequences of the *rpoD* gene. The tree construction method was Maximum Likelihood with the Kimura 2-parameter model incorporating invariant sites and gamma distribution.

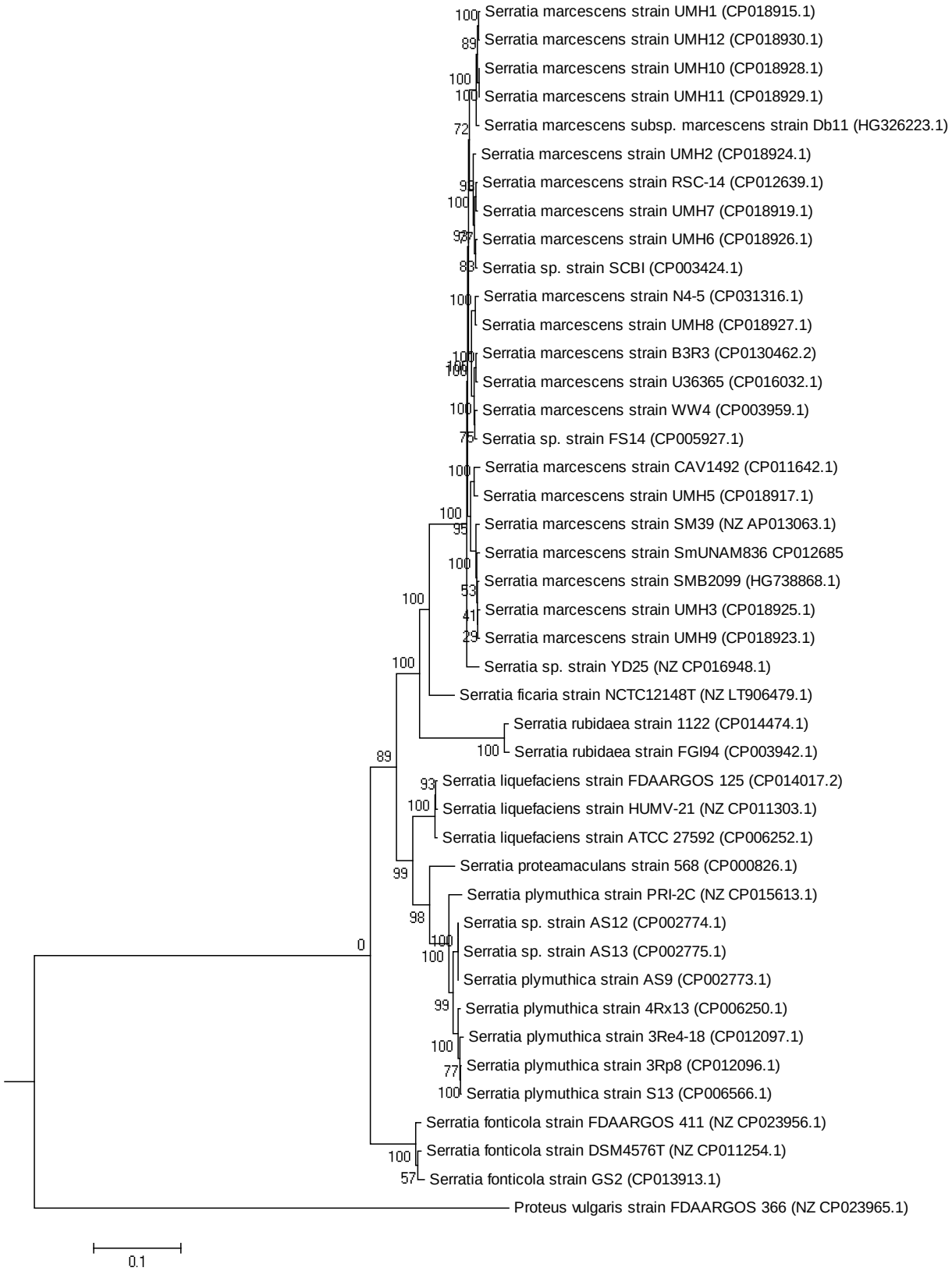

FIG S7 Phylogenetic tree based on concatenated nucleotide sequences of the genes 16S, *ftsz, groEL, groES, gyrB, recA* and *rpoD* (MLSA). The tree construction method was Maximum Likelihood with the GTR model incorporating invariant sites and gamma distribution.

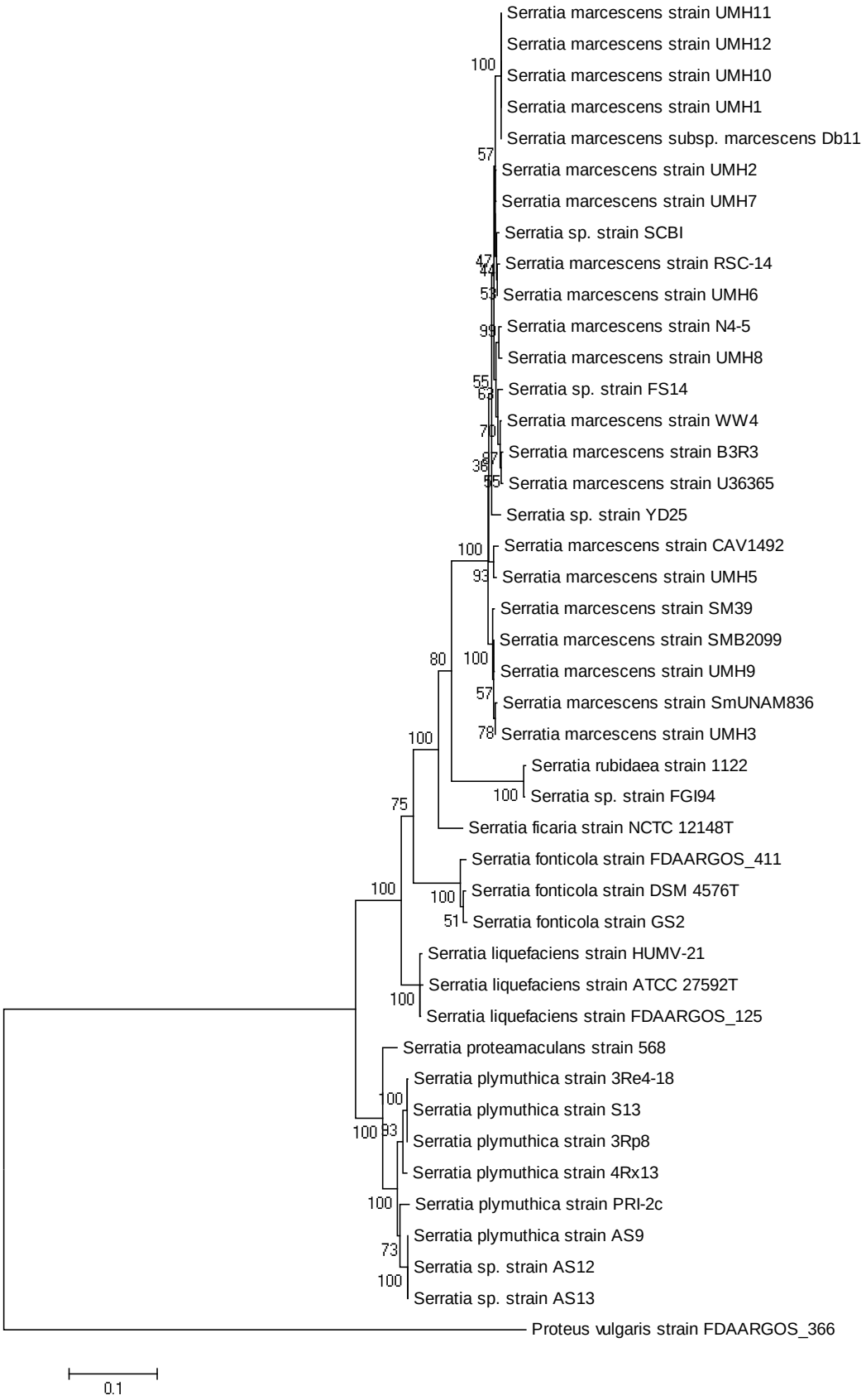

FIG S8 Phylogenetic tree based on nucleotide sequences of the *groEL* gene. The tree construction method was Maximum Likelihood with the Kimura 2-parameter model incorporating invariant sites and gamma distribution.