 frontiers | Frontiers in Microbiology

Check for updates

# Comparative genomics of *Stutzerimonas balearica* (*Pseudomonas balearica*): diversity, habitats, and biodegradation of aromatic compounds

Francisco Salvà-Serra[1,2,3]*, Danilo Pérez-Pantoja[4],
Raúl A. Donoso[4,5], Daniel Jaén-Luchoro[2,3],
Víctor Fernández-Juárez[6], Hedvig Engström-Jakobsson[2],
Edward R. B. Moore[2,3], Jorge Lalucat[1] and
Antoni Bennasar-Figueras[1]*

[1]Microbiology, Department of Biology, University of the Balearic Islands, Palma de Mallorca, Spain,
[2]Department of Infectious Diseases, Institute of Biomedicine, Sahlgrenska Academy, University of
Gothenburg, Gothenburg, Sweden, [3]Culture Collection University of Gothenburg (CCUG), Institute of
Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden, [4]Programa
Institucional de Fomento a la Investigación, Desarrollo e Innovación, Universidad Tecnológica
Metropolitana, Santiago, Chile, [5]Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile,
[6]Marine Biological Section, Department of Biology, University of Copenhagen, Helsingør, Denmark

*Stutzerimonas balearica* (*Pseudomonas balearica*) has been found principally
in oil-polluted environments. The capability of *S. balearica* to thrive from the
degradation of pollutant compounds makes it a species of interest for potential
bioremediation applications. However, little has been reported about the
diversity of *S. balearica*. In this study, genome sequences of *S. balearica* strains
from different origins were analyzed, revealing that it is a diverse species with
an open pan-genome that will continue revealing new genes and functionalities
as the genomes of more strains are sequenced. The nucleotide signatures and
intra- and inter-species variation of the 16S rRNA genes of *S. balearica* were
reevaluated. A strategy of screening 16S rRNA gene sequences in public databases
enabled the detection of 158 additional strains, of which only 23% were described
as *S. balearica*. The species was detected from a wide range of environments,
although mostly from aquatic and polluted environments, predominantly related
to petroleum oil. Genomic and phenotypic analyses confirmed that *S. balearica*
possesses varied inherent capabilities for aromatic compounds degradation. This
study increases the knowledge of the biology and diversity of *S. balearica* and will
serve as a basis for future work with the species.

# 1. Introduction

The species *Stutzerimonas balearica* (Gomila et al., 2022; Lalucat et al., 2022) (formerly, *Pseudomonas balearica* Bennasar et al., 1996) was proposed after the analysis and characterization of two denitrifying and salt-tolerant (i.e., growth in 8.5% NaCl, w/v) strains of the genomovar 6 of *Pseudomonas stutzeri* (currently, *Stutzerimonas stutzeri*) (Rossello et al., 1991; Rosselló-Mora et al., 1994; Bennasar et al., 1996; Lalucat et al., 2022), which had been isolated as 2-methylnaphthalene degrading strains from a wastewater treatment plant and from polluted marine sediment (Rossello et al., 1991). Since then, several strains of *Stutzerimonas balearica* have been reported through the years, such as *S. balearica* strain st101, isolated as a phenanthrene degrader from the rhizosphere of *Spartina patens* at an oil refinery site and reported also as a naphthalene degrader (Dutta, 2001; Mulet et al., 2008), *S. balearica* strain Z8, isolated from petrochemical waste water (Ahmadi et al., 2017), *S. balearica* strain EC28, isolated from corroded steel of a floating production storage and off-loading vessel, Salgar-Chaparro et al. (2020), and three strains which were isolated from a submarine volcano (Bravakos et al., 2021).

Previous reports indicate that *S. balearica* is a species with diverse properties and biodegradation capabilities, which is found principally in marine and polluted environments (Rossello et al., 1991; Bennasar et al., 1996; Dutta, 2001; Mulet et al., 2008; Ahmadi et al., 2017; Salgar-Chaparro et al., 2020; Bravakos et al., 2021). These observations make *S. balearica* a species of interest for bioremediation of areas polluted with recalcitrant compounds, such as aromatic hydrocarbons, some of which are among the most prevalent and persistent pollutants in the environment and are listed in the top 10 of the Substance Priority List of the Agency for Toxic Substances and Disease Registry [Agency for Toxic Substances and Disease Registry (ATSDR), Division of Toxicology and Human Health Sciences, 2019]. However, the heterogeneity of the former *P. stutzeri* group (also known as *P. stutzeri* complex) (Mulet et al., 2010), and the limited discriminatory power of the 16S rRNA gene, have frequently resulted in misclassifications and in strains not being identified at the species level (Lalucat et al., 2006; Gomila et al., 2022; Li et al., 2022; Uddin et al., 2022). For instance, routine protocols for identification, such as commercial API strips (BioMérieux, France), identify *S. balearica* strains as *S. stutzeri* (Uddin et al., 2022). This may result in *S. balearica* strains not being detected or being misidentified. An

example is *S. balearica* YC-YH1, which was erroneously reported as *S. stutzeri* (Shi et al., 2015). Precise and accurate species identification is important because different taxonomic groups may have different capabilities and ecological adaptations (García-Valdés et al., 2010). For instance, *S. stutzeri* has been commonly reported in clinical samples; indeed a previous study (Scotta et al., 2012) reported that almost 80% of 138 analyzed clinical isolates of *Stutzerimonas* belonged to *S. stutzeri* (formerly, *P. stutzeri* genomovar 1), while no isolates were assigned to 11 of the 19 genomovars that had been described at that time, which suggests that different genomovars (i.e., phylogenomic species) of the genus *Stutzerimonas* may have different pathogenic potential. In the original description of *P. balearica* (currently, *S. balearica*), the authors defined 37 signature nucleotide positions of the 16S rRNA gene that differentiated *P. balearica* from the other six genomovars of *P. stutzeri* that had been described at that time (Bennasar et al., 1996). However, 15 additional genomovars and multiple closely-related species have been described since then. Therefore, it is unclear how many of these signature nucleotide positions are still relevant for identifying *S. balearica*. At this point, it is important to clarify that *S. balearica* is a member of the recently proposed genus *Stutzerimonas*, family *Pseudomonadaceae*, which includes species of the former *Pseudomonas stutzeri* phylogenetic group and currently comprises 51 phylogenomic species (Gomila et al., 2022; Lalucat et al., 2022).

The limitations in the detection and identification of *S. balearica* have hindered our understanding of the diversity, the biology and of the environmental distribution of the species. However, the development and the substantial reduction in the costs of whole-genome sequencing (Loman and Pallen, 2015) have resulted in a vast increase in the number of publicly available bacterial genome sequences (Sayers et al., 2022), which facilitates accurate species identification (Goris et al., 2007) and the exploration of the genomic diversity and physiological potential of species (Loman and Pallen, 2015).

The aim of this study was to analyze the genome sequences of strains of *S. balearica* from diverse sources and origins, together with the genome sequences of different phylogenomic species of *Stutzerimonas* and type strains of other species of the former *P. stutzeri* group to: firstly, explore the genomic diversity of *S. balearica* in its phylogenetic context; secondly, determine the current conservation and specificity of the 16S rRNA gene sequence signature nucleotide positions presented previously (Bennasar et al., 1996), thirdly, screen publicly available 16S rRNA gene sequences and determine the incidence and distribution of *S. balearica* at a larger scale; fourthly, determine the biodegradative capabilities of *S. balearica* toward aromatic compounds, based on genomic analyses and phenotypic testing.

# 2. Materials and methods

## 2.1. Strains and whole-genome sequences

A total of 18 genome sequences of *S. balearica* were included in the study, 11 derived from isolated strains and seven derived from metagenomic datasets (i.e., metagenome-assembled genomes, MAGs) (Supplementary Table 1). Of the 11 isolate-derived genome sequences, one was determined in this study (*S. balearica* strain SAGV3-2SA2 = CCUG 72049) and 10 were obtained from GenBank

---

Abbreviations: ANIb, average nucleotide identity based on BLAST; BDBH, bidirectional best-hit; CARD, Comprehensive Antibiotic Resistance Database; CCUG, Culture Collection University of Gothenburg; CDS, coding sequence; COG, cluster of orthologous groups; CRISPR, clustered regularly interspaced short palindromic repeats; dDDH, digital DNA–DNA hybridization; DFAST, DDBJ Fast Annotation and Submission Tool; EC, Enzyme Commission; GGDC, Genome-to-Genome Distance Calculator; GO, gene ontology; GTDB, Genome Taxonomy Database; Gv, genomovar; ICE, integrative and conjugative element; IME, integrative and mobilizable element; MAG, metagenome-assembled genome; NCBI, National Center for Biotechnology Information; MLSA, multilocus sequence analysis; NGS, Next-generation sequencing; $OD_{600}$, optical density at 600nm; OMCL, Ortho Markov Cluster; PGAP, Prokaryotic Genome Annotation Pipeline; Pgs, phylogenomic species; RGI, Resistance Gene Identifier; RND, resistance-nodulation-cell division; SH-aLRT, Shimodaira-Hasegawa-like approximate likelihood-ratio test; SMR, small multidrug resistance; T4SS, type IV secretion system; VFDB, Virulence Factors Database.

(Sayers et al., 2022), where nine were listed as *S. balearica* and one (that of strain YC-YH1) as *S. stutzeri*, although classified as *S. balearica* in the Genome Taxonomy Database (GTDB) (Parks et al., 2018). A recent study also confirmed this misclassification (Li et al., 2022).

MAGs of *S. balearica* were searched for in NCBI GenBank (Sayers et al., 2022) and GTDB (Parks et al., 2018). Additionally, a set of 52,515 published MAGs (Nayfach et al., 2021) was screened, using BLAST v2.9.0+ (Altschul et al., 1990) and the *rpoD* gene sequence of *S. balearica* DSM 6083$^T$ as query sequence. The completeness and contamination levels of the MAGs were estimated, using CheckM v1.2.2 (Parks et al., 2015), implemented in DFAST_QC v.0.2.6 (Tanizawa et al., 2016). The searches resulted in seven MAGs of *S. balearica*, of which six were obtained from GenBank and one (*S. balearica* 3300027365_7) from the set of 52,515 MAGs (Nayfach et al., 2021). Of the six obtained from GenBank, two were classified as *S. balearica*, two as *S. stutzeri* and two as *Pseudomonas* sp., although all six were listed as *S. balearica* in GTDB.

Additionally, 76 genome sequences of strains listed as *S. stutzeri* in GenBank (Sayers et al., 2022) and the genome sequences of 19 additional type strains of species reported as members of the genus *Stutzerimonas* or the former *Pseudomonas stutzeri* group were also included in the study (Supplementary Table 1). Of these 19 additional type strains, five were listed originally as *S. stutzeri* in GenBank: the type strain of *S. stutzeri* and the type strains of the recently proposed species *Stutzerimonas perfectomarina*, *Stutzerimonas frequens*, *Stutzerimonas degradans*, and *"Stutzerimonas decontaminans,"* i.e., the former genomovars 2, 5, 7, and 4, respectively (Gomila et al., 2022; Mulet et al., 2023). Of the 113 genome sequences, 22 are complete genome sequences (including three *S. balearica*) and 91 are of draft status, with a number of contigs/scaffolds ranging from 1 to 298, among isolate-derived genome sequences, and ranging from 41 to 469 among MAGs.

## 2.2. DNA extraction and whole-genome sequencing

*S. balearica* strain SAGV3-2SA2 was cultivated on Columbia Blood Agar Base plus 5% defibrinated horse blood (Substrate Unit, Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Sweden), at 30°C, for 24 h. Genomic DNA was extracted, using a previously described protocol (Salvà-Serra et al., 2018). Subsequently, isolated DNA was used to prepare a standard Illumina library at Eurofins Genomics (Konstanz, Germany), following an optimized protocol and using standard Illumina adapter sequences. The library was sequenced, using an Illumina NovaSeq 6,000 system (Illumina, Inc., San Diego, CA, United States), to generate paired-end reads of 151 bp, distributed in 1.4 Gb.

## 2.3. Assembly and annotation of genome sequences

The Illumina sequence reads of *S. balearica* strain SAGV3-2SA2 were quality-evaluated, using FastQC 0.10.1 (Andrews, 2010), trimmed, using Sickle v1.33 (Joshi and Fass, 2011), and assembled *de novo*, using SPAdes v3.11.1 (Bankevich et al., 2012). The quality of the assembly was assessed, using QUAST v4.5 (Gurevich et al., 2013). The

genome sequence was subsequently annotated, using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.2 (Tatusova et al., 2016). All 106 genome sequences derived from isolated strains were downloaded from the NCBI Reference Sequence Database (RefSeq) (O'Leary et al., 2016), where they were annotated with PGAP. Of the seven MAGs, six were obtained from GenBank, annotated with PGAP. Of these, two were re-annotated on-line, using the DDBJ Fast Annotation and Submission Tool (DFAST) v1.2.6 (Tanizawa et al., 2017), since not all PGAP output files were available in GenBank. The other MAG (*S. balearica* 3300027365_7) was not available in GenBank (it was obtained from a previously published dataset of MAGs) (Nayfach et al., 2021) and, therefore, was also annotated, using DFAST. The DFAST annotations of these three MAGs were used for downstream functional annotations. The DFAST-generated GBFF files are available as Supplementary Files (Supplementary Files 1–3). Additionally, to normalize the annotations for the pan-genomic analyses, all 113 genome sequences were annotated, using Prokka v1.14.6 (Seemann, 2014).

## 2.4. Average nucleotide identity determinations

The average nucleotide identities based on BLAST (ANIb) (Goris et al., 2007) between all 113 genome sequences (all vs. all), were calculated, using JSpeciesWS (Richter et al., 2016). Calculations were done bidirectionally, and average values were determined. The ANIb values were used to construct a dendrogram and a heatmap, using Heatmapper, with an average linkage method for clustering and Pearson's distance correlation (Babicki et al., 2016).

## 2.5. Phylogenomic species assignment

Two approaches were used to determine or confirm the species or phylogenomic species affiliations of the genome sequences included in this study. On the one hand, ANIb values were calculated between all the 113 genome sequences, which included multiple type strains and non-type reference strains of phylogenomic species of *Stutzerimonas*. Genome sequences sharing an ANIb value of greater than or equal to 95% were considered to belong to the same phylogenomic species. The phylogenomic species assignments were confirmed by calculating the digital DNA–DNA hybridization (dDDH) values between the assigned genome sequences and the genome sequences of the respective type or reference strains. The dDDH values were determined, using Genome-to-Genome Distance Calculator (GGDC) v3.0 (Meier-Kolthoff et al., 2013, 2022). On the other hand, a comparative analysis of partial *rpoD* gene sequences was performed, as previously described (Scotta et al., 2012). Briefly, the complete *rpoD* gene sequences were extracted from the genome sequences. The partial *rpoD* gene sequences of the reference strains of 21 phylogenomic species of *Stutzerimonas* (previously, the 21 genomovars of *S. stutzeri*), were downloaded from the PseudoMLSA database (Bennasar et al., 2010) and also included. *Pseudomonas aeruginosa* CCM 1960$^T$ was used as outgroup. The *rpoD* sequences were aligned, using MUSCLE v3.8.425 (Edgar, 2004), implemented in AliView v1.25 (Larsson, 2014), and the distance matrix was built in MEGA7 (Kumar et al., 2016), using the Jukes-Cantor model (Jukes

and Cantor, 1969). The phylogenetic tree was generated by neighbor-joining, bootstrap analysis (1,000 replications) and displayed, using iTOL (Letunic and Bork, 2021).

## 2.6. Core and pan-genome determinations

The protein FASTA files generated by Prokka v1.14.6 were used for proteome comparisons. All-versus-all proteome searches were performed, using BLASTP v2.9.0+. Protein clustering was performed, using the software GET_HOMOLOGUES v17112020 (Contreras-Moreira and Vinuesa, 2013), with three different algorithms: bidirectional best-hit (BDBH), COGtriangles (Kristensen et al., 2010) and Ortho Markov Cluster (OMCL) (Li et al., 2003). Two proteins were considered to belong to the same family if they showed at least 70% sequence identity over at least 70% of the length of the longest sequence in comparison. Based on the BDBH clustering, the core and pan-genome curves were represented and adjusted to the exponential models of Tettelin et al. (2005) and Willenbrock et al. (2007).

"Strict" consensus core genomes of single-copy orthologous sequences were determined, using the intersections of the three clustering algorithms, and used to construct phylogenomic trees. Briefly, the amino acid sequences of single-copy orthologous clusters were, first, independently aligned, using Clustal Omega v1.2.3 (Sievers et al., 2011). The resulting alignments were concatenated and processed, using Gblocks v0.91b to remove poorly aligned positions (Castresana, 2000; Talavera and Castresana, 2007). The resulting trimmed alignments were used to construct phylogenetic trees, using the software PhyML v3.1 (Guindon et al., 2010) and a maximum-likelihood estimation. A Shimodaira-Hasegawa-like approximate likelihood-ratio test (SH-aLRT) was used for branching statistical support (Anisimova and Gascuel, 2006). The phylogenomic trees were displayed, using iTOL (Letunic and Bork, 2021).

The pan-genomes were determined, using the intersections of clusters determined with COGtriangles and OMCL. The pan-genome clusters were classified according to their presence in all genomes (core), in 95% or more of the genomes (soft-core), in a large number of genomes but in less than 95% of them (shell), and present only in one or two genomes (cloud), as described previously (Koonin and Wolf, 2008; Contreras-Moreira and Vinuesa, 2013). The pan-genome matrix was used to determine the pan-genes of *S. balearica* (i.e., genes that are present in the 11 isolate-derived genome sequences of *S. balearica* and absent in all the 95 non-*S. balearica* genome sequences) (Contreras-Moreira and Vinuesa, 2013).

## 2.7. Evaluation of the 16S rRNA gene signature nucleotide positions of *Stutzerimonas balearica*

The 16S rRNA gene sequences were extracted from the isolate-derived genome sequences included in this study, using RNAmmer v1.2 (Lagesen et al., 2007), and aligned with the partial sequences of the reference strains of the former 21 genomovars of *P. stutzeri* (currently, phylogenomic species of *Stutzerimonas*), using MUSCLE v3.8.425 (Edgar, 2004), implemented in AliView v1.25 (Larsson, 2014). The seven MAGs analyzed did not contain any 16S rRNA gene sequence. The genome sequence of *S. balearica* YC-YH1 also did not

contain any 16S rRNA gene sequence. Alternatively, a publicly-available and nearly-complete sequence was included (GenBank accession number: KJ786450.1). The genome sequences of some type strain genome sequences also lacked an assembled 16S rRNA gene sequence. In those cases, a 16S rRNA gene sequence was obtained from the List of Prokaryotic Names with Standing in Nomenclature (LPSN) (Parte, 2014). According to EzBiocloud, *Pseudomonas furukawaii* is the species presenting the highest 16S rRNA gene sequence similarity (97.8%) with that of *S. balearica*. Therefore, *P. furukawaii* (GenBank accession number AP014862.1), despite being a member of the *Pseudomonas resinovorans* group (Lalucat et al., 2020, 2022), was also included in the analysis.

Subsequently, the alignment was inspected manually to determine: first, if the 37 signature nucleotide positions are conserved among *S. balearica* strains; second, if the signature nucleotide positions are exclusive of *S. balearica*. Additionally, a distance matrix was constructed in MEGA7 (Kumar et al., 2016), using the Jukes-Cantor model (Jukes and Cantor, 1969), to determine the intra- and inter-species sequence variability of the 16S rRNA gene.

## 2.8. Search of 16S rRNA gene sequences of *Stutzerimonas balearica*

The partial 16S rRNA gene sequence of *S. balearica* DSM 6083[T] (positions 50–650 bp, which encompass all 37 signature nucleotide positions; GenBank accession number: CP007511.1) was queried, using BLASTN, against the NCBI Nucleotide collection (nr/nt) database (Sayers et al., 2022), to find 16S rRNA gene sequences of additional strains of *S. balearica*. Sequences with similarities greater than or equal to 99.6% and greater than or equal to 90% of query coverage were downloaded and aligned, using MUSCLE v3.8.425 (Edgar, 2004), implemented in AliView v1.25 (Larsson, 2014). Sequences containing all 37 signature nucleotide positions of *S. balearica* were considered to be from members of the species.

## 2.9. Functional annotations of genome sequences

The protein sequences predicted by PGAP or DFAST (if PGAP was not available) were annotated, using Blast2GO (Conesa et al., 2005), InterProScan v5.54–87.0 (Jones et al., 2014) and eggNOG-Mapper v2.1.0 (Huerta-Cepas et al., 2017), with the eggNOG v5.0.2 database (Huerta-Cepas et al., 2019). Transfer of eggNOG annotations was limited to one-to-one orthology and to experimental evidence (i.e., prioritizing precision). The taxonomic scope was limited to *Gammaproteobacteria* (NCBI Taxonomy ID: 1236). The annotated Gene Onthology (GO) terms were validated to remove redundant terms based on the GO True Path Rule and filtered, using the class *Gammaproteobacteria* (Taxonomy ID: 1236). Subsequently, GOs were mapped to Enzyme Commission (EC) numbers, which were used together with EggNOG to annotate KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (Kanehisa et al., 2020). Blast2GO, EggNOG and InterProScan annotations were performed, using the platform OmicsBox v2.0.36 (BioBam Bioinformatics S.L., Valencia, Spain). The protein sequences were also classified into Clusters of Orthologous Groups (COG) categories (Galperin et al., 2015), using

eggNOG-Mapper v2.1.6 (Huerta-Cepas et al., 2017) with the eggNOG v5.0 orthology resource (Huerta-Cepas et al., 2019), using the conditions listed above. Protein sequences of certain genes of interest were also analyzed, using InterPro 91.0 (Blum et al., 2020). Antibiotic resistance genes were searched, using the tool Resistance Gene Identifier (RGI) v5.2.1 of the Comprehensive Antibiotic Resistance Database (CARD) v3.1.4 (Alcock et al., 2020). Biocide and metal resistance genes were searched, using the BacMet v2.0 database (Pal et al., 2013) and the BacMet script *BacMet-Scan.pl*, against the database of "Experimentally confirmed resistance genes." Virulence factors were searched, using VFanalyzer, of the Virulence Factors Database (VFDB) (Liu et al., 2022). Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems were predicted, using the on-line tool CRISPRone (Zhang and Ye, 2017). Prophages were predicted, using PHASTER (Arndt et al., 2016). Putative integrative and conjugative elements (ICEs) and integrative and mobilizable elements (IMEs) were searched, using ICEfinder v1.0 (Liu et al., 2018). As ICEfinder can only analyze one single sequence per submission, the contigs or scaffolds of draft genome sequences and MAGs were merged into a single sequence per genome, using UGENE v43.0 (Okonechnikov et al., 2012), and contig/scaffolds were separated from each other by ten N's. Natural transformation genes were searched for, using VFanalyzer (Liu et al., 2022), BLAST v2.5.0+ (Altschul et al., 1990), and the genome sequence of *Stutzerimonas nitrititolerans* strain DSM 10701 (= JM300; the former reference strain of the genomovar 8 of *S. stutzeri*) as a reference (Busquets et al., 2012). Genes encoding regulatory proteins were predicted, using P2RP v2.7 (Barakat et al., 2013).

## 2.10. *In silico* identification of pathways for degradation of aromatic compounds

Well-characterized gene products of key central and peripheral aromatic pathways (Ferrero et al., 2002; Pérez-Pantoja et al., 2008, 2012) were used as initial seeds to search for orthologous sequences throughout the genomes and MAGs of *S. balearica* strains, using BLAST v2.7.1 (Altschul et al., 1990). Proteins displaying at least 30% amino acid identity with query sequences were selected for additional analysis, including sequence similarity search in the protein database of GenBank (Sayers et al., 2022), identification of conserved domains, using the Conserved Domain database v3.19 (Lu et al., 2019), and manual inspection of gene neighborhoods, using Artemis v17.0.1 (Carver et al., 2011). Once a key orthologous sequence of central and/or peripheral aromatic pathway was identified, a complementary gene search of the whole pathway was performed, using the same strategy as before. A schematic representation of presence/absence of genes was generated, using Heatmapper (Babicki et al., 2016).

## 2.11. Growth assays of aromatic compounds

The six strains of *S. balearica* available at the Culture Collection University of Gothenburg (CCUG) were grown in mineral salts medium (14 g/l $Na_2HPO_4 \cdot 12H_2O$, 2 g/l $KH_2PO_4$, 50 mg/l $Ca(NO_3)_2 \cdot 4H_2O$, 1 g/l $(NH_4)_2SO_4$, 200 mg/l $MgSO_4 \cdot 7H_2O$, 278 μg/L $FeSO_4 \cdot 7H_2O$, 70 μg/L $ZnCl_2$, 100 μg/L $MnCl_2 \cdot 4H_2O$, 62 μg/L $H_3BO_3$,

190 μg/L $CoCl_2 \cdot 6H_2O$, 17 μg/L $CuCl_2 \cdot 2H_2O$, 24 μg/L $NiCl_2 \cdot 6H_2O$, 36 μg/L $Na_2MoO_4 \cdot 2H_2O$), supplemented with 2.5 mM benzoate, phenol, salicylate, naphthalene, phenylalanine and tyrosine as sole carbon and energy source. Naphthalene was dissolved in acetone. As growth control, acetone was used without the addition of naphthalene. The cultures were incubated in borosilicate glass tubes at 30°C for 5 days and optical densities at 600 nm ($OD_{600}$) were measured, using a Synergy HTX multimode plate reader (BioTek, Winooski, VT, United States). Cultures were inoculated with 100-fold dilutions from overnight cultures grown on R2A broth. At least three biological replicates were performed for each growth measurement.

# 3. Results and discussion

## 3.1. ANIb and phylogenomic species assignment of genome sequences

The ANIb values were calculated between all the 113 genome sequences included in this study (Supplementary Table 2), which revealed a total of 33 genomic species clusters and confirmed the species identities of the 18 genome sequences of *S. balearica* (Figure 1). It is notable that, of the 16 confirmed genome sequences of *S. balearica* obtained from GenBank, three were misclassified as *S. stutzeri* and two were listed as *Pseudomonas* sp., which exemplifies the problem of the poor classifications or misclassifications of strains of *S. balearica*. The lowest ANIb value between *S. balearica* strains was 96.54%, while the highest was 100%, i.e., between strains KOL14W2010, KOL14W2549549, and KOL14.W25.495.B5. These three strains were isolated from the same geographical area at Kolumbo submarine volcano (Greece), which probably explains the high level of genomic sequence similarity (Mandalakis et al., 2019). When excluding the values between these clonal strains, the highest ANIb value within *S. balearica* was 98.46%. Meanwhile, the highest ANIb between *S. balearica* DSM 6083$^T$ and a non-*S. balearica* strain of the dataset of this study was observed to be 81.24% (*Stutzerimonas* sp. strain 24a75, pgs 17, ref.), and the highest ANIb value with a non-*S. balearica* type strain was with *S. degradans* DSM 50238$^T$ (81.08%), followed by *S. stutzeri* CGMCC 11803$^T$ (80.93%). Among other strains of *Stutzerimonas* species, the highest inter-pgs ANIb value was 92.51% (between *Stutzerimonas chloritidismutans* AW-1$^T$, and *Stutzerimonas* sp. strain NF13, pgs 19), while the lowest was 74.85% (*Stutzerimonas urumqiensis* T3$^T$, and *Stutzerimonas zhaodongensis* NEAU-ST5-21$^T$).

The ANIb values between the 20 type strains included in this study ranged from 71.24% (*S. zhaodongensis* NEAU-ST5-21$^T$ and *Pseudomonas matsuisoli* JCM 30078$^T$) to 95.84% (*Stutzerimonas kunmingensis* DSM 25974$^T$ and *S. chloritidismutans* AW-1$^T$). The high ANIb value indicates that *S. kunmingensis* and *S. chloritidismutans* are, in fact, synonymous, which is in agreement with previous observations (Hesse et al., 2018; Lalucat et al., 2020). Indeed, Gomila et al. emended the description of *S. chloritidismutans,* which includes strains of former *P. stutzeri* genomovar 3, of which *S. kunmingensis* is a later heterotypic synonym (Cladera et al., 2006; Gomila et al., 2022). This is supported also in this study by the high ANIb values (>96%) between the type strains of *S. chloritidismutans*, *S. kunmingensis* and the reference strain of the former *P. stutzeri* genomovar 3. If the synonyms are excluded, the highest ANIb value between type strains of *Stutzerimonas* is 89.29% (between *S. frequens* DNSP21$^T$ and
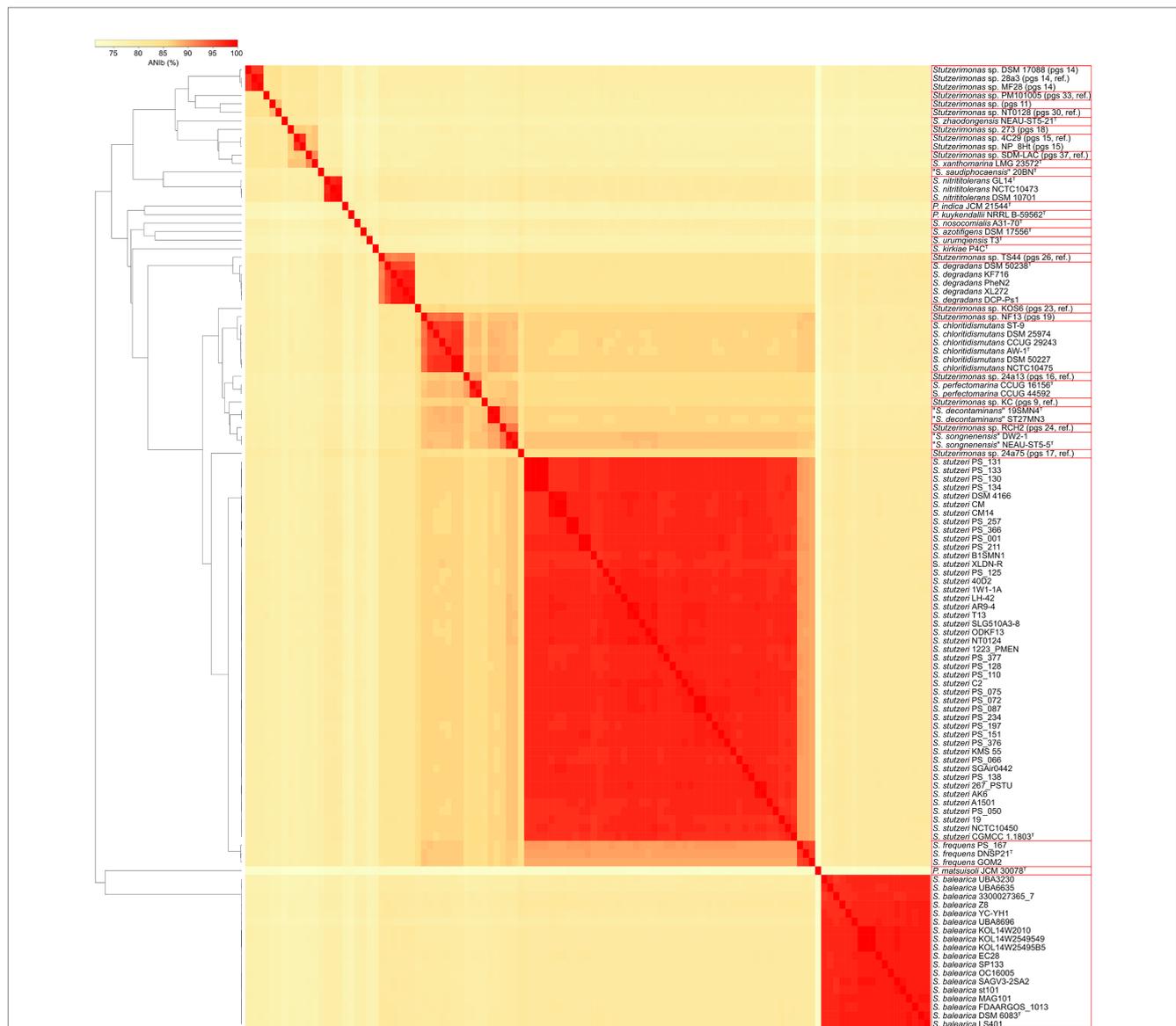
**FIGURE 1**
Dendrogram and heatmap, based on the ANIb values calculated between all 113 genome sequences included in this study. The phylogenomic species affiliations are indicated in parenthesis. Red squares indicate clusters of strains sharing greater than or equal to 95% of ANIb value (i.e., same genomic species). Pgs, phylogenomic species; ref., reference.

*S. stutzeri* CGMCC 11803[T]). Additionally, when looking at the ANIb values between the 20 type strains, one notices that *P. matsuisoli* exhibits the lowest values. This species was considered to be a member of the *P. stutzeri* group by Lalucat et al. after performing a four-gene multilocus sequence analysis (MLSA) (Lalucat et al., 2020). However, *P. matsuisoli* was the most distantly-related species, together with *Pseudomonas indica* and *Pseudomonas kuykendallii*, which were previously also considered to be members of the former *P. stutzeri* group (Lalucat et al., 2020). Indeed, these three species have been recently shown to not be members of the former *P. stutzeri* group and to form 'orphan' groups of individual single species (Girard et al., 2021; Lalucat et al., 2022).

　　Worth noting is that *S. stutzeri* strain DSM 17088 was supposed to be the reference strain of pgs 12 (Sikorski et al., 2005; Gomila et al., 2022), but instead was affiliated with pgs 14. Partial *rpoD* sequencing

of the CCUG deposit of this strain (i.e., CCUG 50543) revealed the same problem (GenBank accession number: OP081805.1). This suggests that a mix-up probably happened before the strain was deposited in these two culture collections. The ANIb analysis also confirmed that strain DW2-1 is a member of "*Stutzerimonas songnenensis*," which is supported by the *rpoD* comparative analysis (Supplementary Figure 1) and in agreement with Li et al. (2022).

　　The inclusion of partial *rpoD* sequences of all the reference strains of the 21 former *P. stutzeri* genomovars allowed the assignment of two strains to phylogenomic species for which the genome sequences for the reference strains were not available for this study. Partial *rpoD* data also indicated that *S. zhaodongensis* is equivalent to the former genomovar 20, which supports the results from Gomila et al. (2022). However, in three cases, the *rpoD* analysis contradicted the ANIb results. One of these cases was the type strain of

*S. kunmingensis* (DSM 25974[T]), which is synonymous with *S. chloritidismutans*, according to the ANIb analyses, but was assigned to *Stutzerimonas* pgs 19 by the *rpoD* analysis. Two other cases were the strains PS_066 and SGAir0442, which, according to ANIb analyses are members of *S. stutzeri*, but clustered with strain KOS6 (pgs 23, ref.) in the *rpoD* analysis. For the rest of strains, the *rpoD* analysis was in agreement with the ANIb results, although the observed discrepancies highlight the importance of relying, when possible, on whole-genome comparisons rather than single gene analyses.

In total, 75 of the 81 strains originally classified as *S. stutzeri* in GenBank were assigned to 15 of the 21 former *P. stutzeri* genomovars and one strain was assigned to "*S. songnenensis*." These 75 strains included two *S. perfectomarina* strains (including the type strain), four *S. chloritidismutans* strains (including the type strain), three *S. frequens* strains (including the type strain), five *S. degradans* strains (including the type strain), two *S. nitrititolerans* strains, and two "*S. decontaminans*" strains (including the type strain), i.e., the former genomovars 2, 3, 5, 7, 8 and 4, respectively (Gomila et al., 2022; Mulet et al., 2023). The remaining six strains had been assigned previously to recently proposed phylogenomic species of *Stutzerimonas* (Gomila et al., 2022).

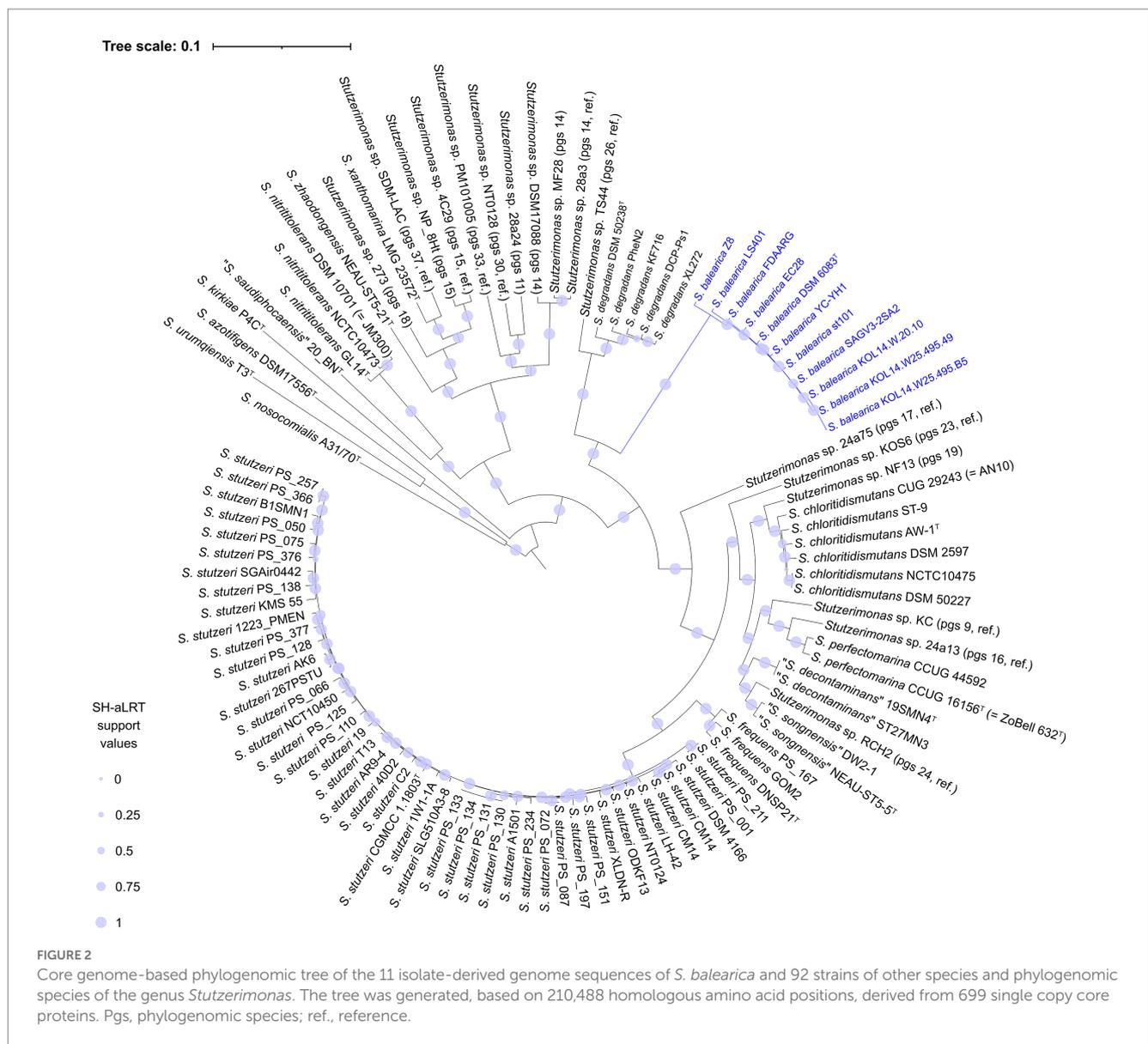## 3.2. Core and pan-genome analyses of *Stutzerimonas balearica*

A pan-genome analysis was carried out, including the Prokka-annotated protein sequences of the 113 genome sequences used in the ANIb calculations. The analysis revealed a "strict" consensus core genome of single copy proteins, which was formed by 535 proteins when excluding the seven MAGs and 699 proteins when excluding the type strains of *P. indica*, *P. kuykendallii* and *P. matsuisoli*. These three type strains were reported previously to be members of the *P. stutzeri* group (i.e., *Stutzerimonas*) after a four-gene MLSA (Lalucat et al., 2020), but later shown to form orphan groups, by phylogenomic analyses (Girard et al., 2021; Lalucat et al., 2022). The observed numbers are in line with the 524 single-copy orthologous genes shared between the 123 strains of the *P. stutzeri* complex and the four outgroup strains analyzed by Li et al. (2022). These are also in accordance with the 666 core genes detected by Gomila et al. when outgroup strains of the family *Pseudomonadaceae* were included with the analysis of 200 strains (Gomila et al., 2022). The 699 proteins, totaling 210,488 amino acid positions, were used to construct a phylogenomic tree (Figure 2). In this 103-genomes phylogenomic tree, *S. balearica* strains cluster together with strains of *S. degradans* and *Stutzerimonas* sp. strain TS44 (pgs 26, ref.), but forming a distinct and separate branch. A second phylogenomic tree was also constructed, using a "strict" consensus single-copy core genome of 127 protein sequences, totaling 35,374 amino acid positions, including the MAGs and the type strains of *P. indica*, *P. kuykendallii* and *P. matsuisoli* (Supplementary Figure 2). The two trees showed highly similar structures and, in both cases, all strains of *S. balearica* clustered together in a single monophyletic branch, closely-related to *S. degradans* and *Stutzerimonas* sp. strain TS44. The 113-genomes tree also shows that the type strains of *P. indica*, *P. kuykendallii* and *P. matsuisoli* were the three strains most distantly related from the type strain of *S. stutzeri*, i.e., the type species of the genus *Stutzerimonas*,

which confirms the results from Girard et al. (2021) and Lalucat et al. (2022).

The BDBH clustering algorithm estimated a core genome size of nearly 900 gene clusters (estimations ranging from 858 to 893 gene clusters) for the 103 isolate-derived genome sequences of *Stutzerimonas* (i.e., the seven MAGs and the three non-*Stutzerimonas* genomes excluded) (Figure 3A). The Tettelin and the Willenbrock models estimated the size of the core genome of the 103 genome sequences in 1,035 and 957 gene clusters, respectively (Figure 3A). If 1,000 were added, according to these models, the sizes of the core genome would be 1,032 and 687 gene clusters, respectively. These estimations are in line with the 1,104 and 1,054 core genes predicted after including 123 (27 genomic species) (Li et al., 2022) and 200 genome sequences (45 genomic species) (Gomila et al., 2022), respectively. The core genome estimations may vary depending on the algorithms used, clustering parameters, the number of strains, their diversity and the quality of the genome sequences and annotation, but all in all, demonstrate that the members of the genus *Stutzerimonas* share a relatively small core genome, which represents approximately a 15–25% of the average number of coding sequences per genome annotated by PGAP in the *Stutzerimonas* genomes included in this study. This is in line with what has been found among other members of the family *Pseudomonadaceae* (Udaondo et al., 2016; Gomila et al., 2017; Freschi et al., 2019; Whelan et al., 2021). However, larger scale analyses including more genomes (and if possible complete) and more genomic species are necessary to get more exact estimations. For instance, after the inclusion of more than 1,300 genome sequences, Freschi et al. determined that the core genome of *P. aeruginosa* was formed by 665 genes, which represents approximately 10% of a typical *P. aeruginosa* genome (Freschi et al., 2019).

The analysis also revealed that the genome sequences derived from 103 isolated strains of *Stutzerimonas* species included in this study exhibit an open pan-genome (Figure 3B), formed by 15,484 gene clusters, according to the estimation of the OMCL-COGtriangles algorithms (Figures 3B,C). When the seven MAGs of *S. balearica* were included, the size increased to 16,628 gene clusters. These pan-genome sizes are in line with the studies published by Li et al. (2022) and Gomila et al. (2022), which reported 13,261 and 18,654 gene clusters, respectively. The Tettelin exponential model estimated that if 1,000 genomes were analyzed, the pan-genome would comprise 62,647 gene clusters, which indicates that the isolation and sequencing of additional strains of *Stutzerimonas* will reveal large amounts of novel genes. However, large-scale studies, such as that of Freschi et al. (2019) are needed to get more precise estimations. Moreover, studies aiming to sequence novel strains of *Stutzerimonas* would probably end up revealing novel genomic species and, thus, increase the knowledge of the diversity of *Stutzerimonas*. It is also notable that nearly 10,000 of the 15,484 gene clusters were found in only one strain (Figure 3C), which confirms the extremely high diversity of the genus *Stutzerimonas*.

The analysis of the 11 isolate-derived genome sequences of *S. balearica* revealed a core genome of 3,071 gene clusters (BDBH estimations) (Figure 3D). The estimated core genome sizes, according to the Tettelin and Willenbrock models, were 3,167 and 3,138 gene clusters, respectively. According to these models, the sizes of the core genome would decrease to 3,163 and 3,037, respectively, if 1,000 genomes were analyzed. In any case, this should be confirmed by including additional strains of *S. balearica* in the analyses. The

**FIGURE 2**
Core genome-based phylogenomic tree of the 11 isolate-derived genome sequences of *S. balearica* and 92 strains of other species and phylogenomic species of the genus *Stutzerimonas*. The tree was generated, based on 210,488 homologous amino acid positions, derived from 699 single copy core proteins. Pgs, phylogenomic species; ref., reference.

analyses also revealed that *S. balearica* has an open pan-genome (Figure 3E), formed by 6,100 gene clusters according to the OMCL-COGtriangles algorithms (Figure 3F). The consensus "strict" single-copy core genome (i.e., the intersection of the BDBH, COG and OMCL algorithms) was formed by 2,987 gene clusters (Figure 3G). When the MAGs were included, the size of the pan-genome estimated by the OMCL-COGtriangles algorithms increased to 7,867 gene clusters (Figures 3H,I). In the analysis of the 11 isolate-derived genome sequences of *S. balearica*, the Willenbrock model estimated that, if 100 genomes were analyzed, the pan-genome would be formed by 17,144 gene clusters. If the seven MAGs were included in the analysis, the predicted pan-genome for 100 genome sequences increased slightly to 17,661 gene clusters. This small increase, which occurred despite the relatively low completeness of four of the MAGs, might be due to contamination during the binning process, even if CheckM indicated low levels of contamination (<3%) (Parks et al., 2015). More contiguous or even complete MAGs with less contamination could be obtained by using long-read sequencing in

metagenomic studies (Chen et al., 2020; Sereika et al., 2022). The presence of the three clonal strains in the pan-genomic analysis may also have an effect on the exponential models and cause an underestimation of the slope of the pan-genome, but in any case, these results indicate that *S. balearica* has a clearly open pan-genome and that sequencing additional strains will continue revealing new genes.

The pan-genome analyses also revealed 24 genes which were present in all of the 11 isolate-derived genome sequences of *S. balearica* but absent in the 95 non-*S. balearica* genome sequences (i.e., pan-genes; Supplementary Table 4). Annotations revealed that 17 of the genes encode membrane-bound proteins. Four and seven of the genes were grouped in two respective gene clusters. In the first gene cluster, two genes encode hypothetical proteins that could not be assigned to any function, another one a transcription factor, and a fourth one an EAL domain-containing protein, which could potentially serve as a phosphodiesterase or regulate cell motility and biofilm formation (Youssef et al., 2017). The same gene cluster also contains a gene encoding a carbon storage regulator protein, which is

**FIGURE 3**

Core and pan-genome representations of *Stutzerimonas* and *S. balearica*. **(A)** Estimate of the core genome size of the 103 isolate-derived *Stutzerimonas* genome sequences fitted to the Tettelin (red) and Willenbrock (blue) exponential decay models. **(B)** Estimate of the pan-genome size of the 103 isolate-derived *Stutzerimonas* genome sequences fitted to the Tettelin exponential model. **(C)** Distribution of the OMCL-COGtriangles pan-genome matrix of the 103 isolate-derived *Stutzerimonas* genome sequences into cloud, shell, soft core, and core compartments. **(D)** Estimate of the core genome size of the 11 isolate-derived genome sequences of *S. balearica*. **(E)** Estimate of the pan-genome size of the 11 isolate-derived genome sequences of *S. balearica*. **(F)** Distribution of the OMCL-COG triangles pan-genome matrix of the 11 isolate-derived genome sequences of *S. balearica*. **(G)** Venn diagram of the core genomes generated by the three clustering algorithms: BDBH, COG triangles, and OMCL. **(H)** Estimate of the pan-genome size of the 11 isolate-derived genome sequences and the seven MAGs of *S. balearica*. **(I)** Distribution of the OMCL-COGtriangles pan-genome matrix of the 11 isolate-derived genome sequences and seven MAGs of *S. balearica*.

not exclusive of *S. balearica*. The second gene cluster includes genes encoding subunits of a NADH-quinone oxidoreductase. In *P. aeruginosa*, the presence of three enzymes for catalyzing the NADH:quinone oxidoreductase step of the respiratory chain, which confer resilience on its energy production systems, was shown in a previous study (Hreha et al., 2021). Another pan-gene, which is not located in the NADH-quinone oxidoreductase gene cluster, also encodes a protein potentially involved in the electron transport chain,

possibly a cytochrome C oxidase subunit. Other genes encode proteins with domains for various activities. These include a methyltransferase domain, a protein with an acetyltransferase domain, and possible S-methyl-5-thioribose-1-phosphate isomerase activity. Another gene encodes a dTDP-4-dehydrorhamnose reductase, potentially involved in dTDP-L-rhamnose biosynthesis and thus in the synthesis of lipopolysaccharide and rhamnolipids (i.e., surfactants) (Olvera et al., 1999). Finally, a gene codifies for a

membrane lipoprotein and multiple genes encode hypothetical proteins.

## 3.3. Clusters of orthologous groups functional annotation of *Stutzerimonas balearica*

The accessioned protein sequences annotated by PGAP or DFAST were assigned to COG functional categories. The distribution of the COG categories was similar among all the *S. balearica* strains and, in all of them, about a 20% of the proteins were assigned to the category "function unknown" (Figure 4), highlighting once again the need for studies to characterize and unveil the function of yet undescribed proteins. Other categories with large numbers of assigned proteins were "energy production and conversion," "amino acid transport and metabolism," and "inorganic ion transport and metabolism," which might be due to the wide metabolic versatility of *Stutzerimonas* (Lalucat et al., 2006). The categories "signal transduction mechanisms" and "transcription" were also among those with the highest numbers of assigned proteins, which is in line with the large regulatory network previously reported in *Pseudomonas* (Stover et al., 2000). The category "cell wall/membrane/envelope biogenesis" also encompassed numerous proteins (>200 per strain), which is reasonable, considering that *S. balearica* is a widely distributed and versatile species that deals with numerous environmental conditions and is exposed to a wide range of stressing conditions.

## 3.4. Habitats and environmental distribution of *Stutzerimonas balearica*

The description of *S. balearica*, presented 37 species-specific 16S rRNA gene signature nucleotide positions distributed in 18 regions of the 16S rRNA gene (Bennasar et al., 1996). Those signatures were determined by comparing the 16S rRNA gene sequences of 2 *S. balearica* strains with the consensus 16S rRNA gene sequences of the former *P. stutzeri* strains, which, at that time, were classified in six genomovars. However, since then, many species and phylogenomic species of *Stutzerimonas* and species of *Pseudomonas* have been described and more strains of *S. balearica* obtained. For these reasons, the current relevance of the described signature nucleotide positions for detecting and differentiating strains of *S. balearica* is unclear.

The comparison of the sequences included in this study revealed that the highest 16S rRNA gene sequence identity between *S. balearica* DSM 6083[T] and the type strain of another species was 97.7% (*P. furukawaii* KF707[T], in agreement with EzBiocloud), while the lowest sequence identity among strains of *S. balearica* was 99.6%. The analysis also revealed that only one signature position remains exclusive to *S. balearica* (*E. coli* position 224). This was partly due to the relatively high similarities of the 16S rRNA gene sequences of the type strain of *S. balearica* and the type strains of *S. azotifigens*, *P. furukawaii*, *P. indica*, *P. matsuisoli*, *S. urumqiensis*, particularly in the hypervariable region V1, which harbors 18 of the 37 signature nucleotide positions. However, none of the non-*S. balearica* strains contained
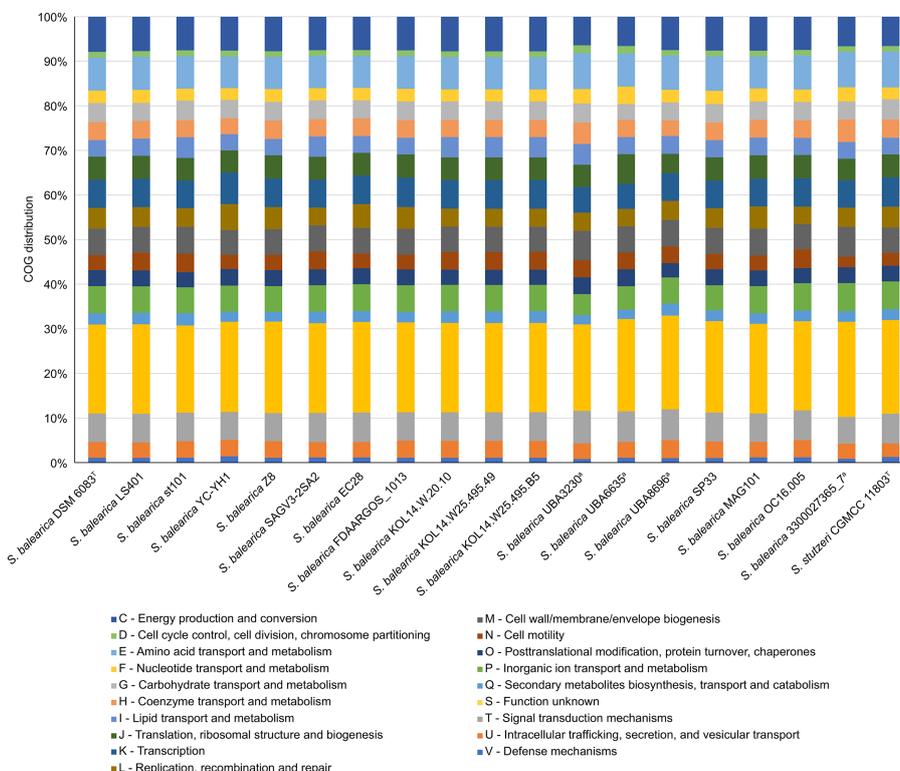


FIGURE 4
COG functional categories distribution of the accessioned proteins of the isolate-derived genome sequences and MAGs of *S. balearica* and the type strain of *S. stutzeri*. [a]MAGs with a CheckM completeness estimate lower than 95%.

the combination of all 37 signature nucleotide positions, while the 16S rRNA gene sequences derived from all 10 *S. balearica* genomes and the nearly-complete sequence of strain YC-YH1 maintained all 37 signature nucleotide positions. These results indicate that the 16S rRNA gene sequences can be used effectively to differentiate *S. balearica* from other closely-related species. This allowed us to establish a two-step strategy to screen for and detect strains of *S. balearica* in public sequence databases, based on sequence identity and identification of the 37 signature nucleotide positions (Figure 5A).
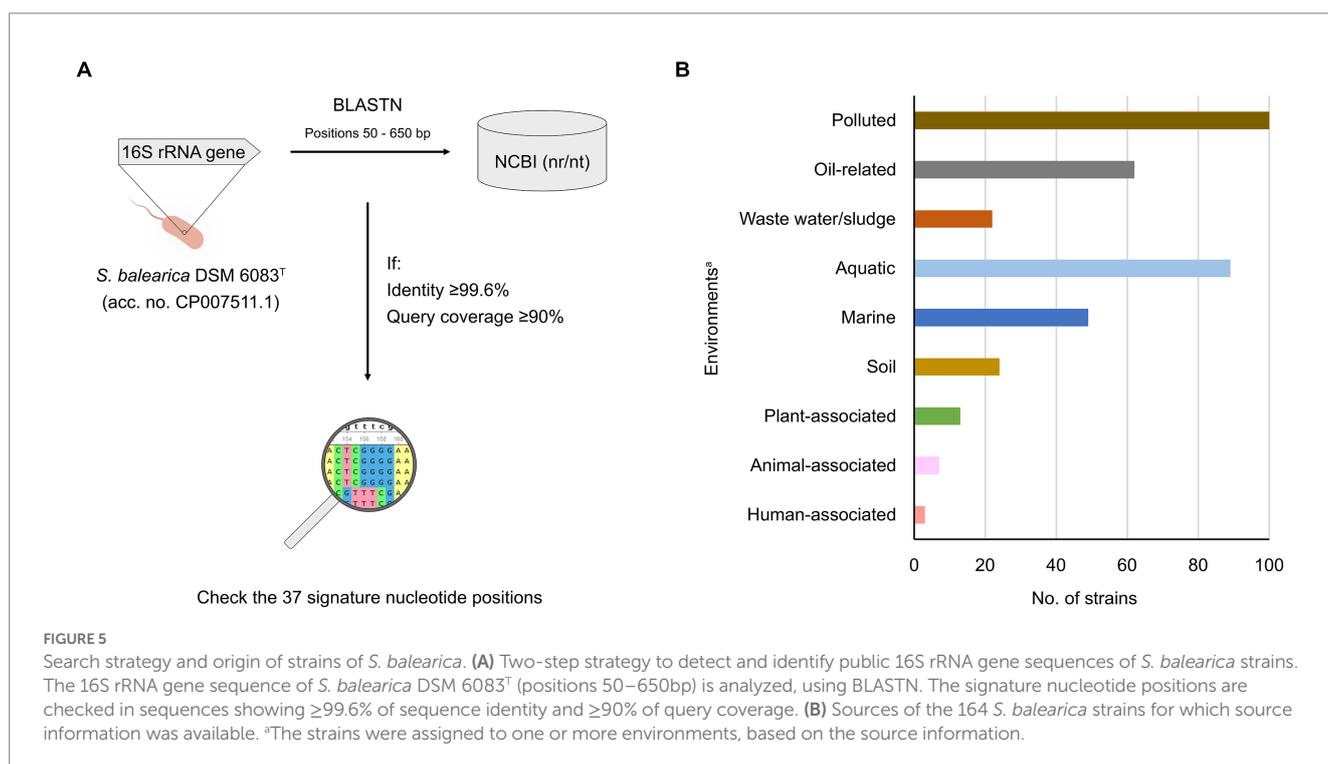
The BLASTN of the 16S rRNA gene sequence of *S. balearica* DSM 6083[T] against NCBI Nucleotide collection (nr/nt) database yielded 173 sequences with 99.6% or greater sequence identity and with 90% or greater query coverage. Of these, 158 were strains without publicly available genome sequence that contained all 37 signature nucleotides. All of these sequences were considered to be derived from authentic *S. balearica* strains that, together with the 18 that were already included in the study, add up to as many as 176 *S. balearica* strains (Supplementary Table 3). This demonstrates that the proposed strategy is a useful method for identifying strains of *S. balearica* by analyzing 16S rRNA gene sequence in cases where genomic data is not available. To increase further the reliability of the identification, the strategy could be complemented with *rpoD* gene sequencing, which despite the discrepancies that have been observed for a few strains of some species of *Stutzerimonas*, has shown full agreement with the genomic-based taxonomic assignments of *S. balearica*.

The strategy is highly restrictive, as it allows only for strains exhibiting sequences with 99.6% or greater identity and all 37 signature nucleotide positions, to be considered as authentic *S. balearica*. A consequence of this is that authentic *S. balearica* strains probably have been missed, because their partial 16S rRNA gene sequences did not cover one or more signature nucleotide positions.

However, we consider that being strict and missing strains is more reliable than setting relaxed thresholds and having a higher risk of obtaining false positives.

It is notable that only 23% (37/158) of the sequences were originally labeled as *S. balearica*, while 68% (107/158) were assigned to higher rank taxa and 9% (14/158) were misclassified, 13 as *S. stutzeri* and one as *Mesobacillus subterraneus*. This shows, once again, the problems with lack of accuracy and misclassified sequences in public databases (Ashelford et al., 2005; Yarza et al., 2008), which is widely extended and was even shown to affect specialized 16S rRNA gene sequence databases, such as SILVA and RDP (Ribosomal Database Project) (Edgar, 2018), despite the availability of tools and curated databases, such as EzBiocloud, that can help to alleviate the problems of inaccurate and wrongly classified of 16S rRNA gene sequences (Yoon et al., 2017).

Information about the geographical origins was available for 116 of the 176 *S. balearica* strains (Supplementary Table 3). These strains, alone, cover more than 30 countries from around the world and illustrate the geographically ubiquitous distribution of *S. balearica*. Additionally, information about the sources was available for 164 of the strains and included a wide range of environments (Figure 5B), including diverse aquatic and terrestrial ecosystems and several plant, animal, and fungal hosts. Interestingly, 61% (100 of 164) originated from polluted or anthropogenic environments, which is in agreement with what previous reports have suggested (Rossello et al., 1991; Bennasar et al., 1996; Dutta, 2001; Mulet et al., 2008; Ahmadi et al., 2017; Salgar-Chaparro et al., 2020). Moreover, 13% (22 of 164) originated from waste waters or sludge and 38% (62 of 164) originated from petroleum oil-related sources. This could be due to sampling bias toward these kinds of environments or a selection for *S. balearica*, but, in any case, such a predominance of suchlike environments highlights the relevance and the potential of *S. balearica* strains for bioremediation applications.



**FIGURE 5**
Search strategy and origin of strains of *S. balearica*. **(A)** Two-step strategy to detect and identify public 16S rRNA gene sequences of *S. balearica* strains. The 16S rRNA gene sequence of *S. balearica* DSM 6083[T] (positions 50–650bp) is analyzed, using BLASTN. The signature nucleotide positions are checked in sequences showing ≥99.6% of sequence identity and ≥90% of query coverage. **(B)** Sources of the 164 *S. balearica* strains for which source information was available. [a]The strains were assigned to one or more environments, based on the source information.

Additionally, 54% of these strains (89 of 164) were obtained from aquatic or water-containing sources and, 55% of these (49 of 89), originated from marine environments, which is in accordance with the described capacity of *S. balearica* to grow in the presence of 8.5% NaCl (Bennasar et al., 1996). Moreover, Li et al. recently suggested that the ancestors of the *P. stutzeri* complex may have originated from high-osmolarity environments, due to the presence of an ectoine biosynthesis gene cluster present in all genomes of the *P. stutzeri* complex (currently *Stutzerimonas*) and absent in most of the other *Pseudomonas* genomes (Li et al., 2022). Meanwhile, 13 strains were associated with plants, one with fungal spores, seven with animals and, interestingly, three were derived from human samples, one from the skin of a healthy individual and two from clinical samples. Indeed, a recent study reported a clinical carbapenem-resistant *S. balearica* strain isolated from a tracheal aspirate and carrying a VIM metallo-β-lactamase (Uddin et al., 2022). This suggests that potential of *S. balearica* to cause opportunistic infections and to carry relevant antimicrobial-resistance factors should not be disregarded, especially considering that rapid phenotypic identification methods often misidentify *S. balearica* strains as *S. stutzeri* (Uddin et al., 2022). Indeed, *S. stutzeri* is often isolated from clinical samples (Köse et al., 2004; Sader and Jones, 2005; Scotta et al., 2012; Mulet et al., 2017), which highlights the importance of accurate species and phylogenomic species identification in the clinics. Altogether, these data demonstrate that *S. balearica* is a versatile bacterium that can be found in a wide range of environments, similar to that of the nearly universal presence of *S. stutzeri* and *P. aeruginosa* (Lalucat et al., 2006; Crone et al., 2019). In many cases, *S. balearica* strains have been isolated because of certain metabolic and physiological properties of interest, such as denitrification, oxidation of thiosulfate, oxidation of 2-chloroethanol and iron, halotolerance, biofilm formation, antimicrobial properties, degradation of pesticides and polyethylene, biosurfactant production, but especially for capacities for biodegradation of aromatic hydrocarbons (Supplementary Table 3). Additionally, several studies have reported on strains of *S. balearica* with other properties of interest, such as amylase production (Kizhakedathil and Subathra, 2021), nitrogen fixation (Beauchamp et al., 2006), degradation of tributyltin (used as a biocide in antifouling paints) (Sampath et al., 2012), methylmercury decomposition (Lee et al., 2012) or bioleaching of gold and silver (Kumar et al., 2018). However, the identifications of such strains could not be verified or did not fulfill the identification thresholds used in this study. Altogether, these observations reinforce the idea that *S. balearica* is a species with large potential, not limited to bioremediation applications, particularly in highly-polluted marine and salty environments.

## 3.5. Degradation of aromatic compounds by *Stutzerimonas balearica*

Strains of *Stutzerimonas* are well known for their potential in aerobic bioremediation of aromatic compounds (Lalucat et al., 2006) and, indeed, several strains of *S. balearica* have been reported to be degraders of aromatic compounds (Rossello et al., 1991; Dutta, 2001; Parreira et al., 2011). To draw a wider picture of the potential of *S. balearica* for bioremediation, gene products of central and peripheral pathways of catabolism of aromatic compounds were searched for in the 11 isolate-derived genome sequences and seven

MAGs. Additionally, growth assays to confirm their degradative capabilities were performed on six of these 11 strains (i.e., the six that were available at the CCUG).

Seven of the 11 genome sequences from isolated strains and three MAGs codified for the complete catechol *ortho* ring-cleavage pathway (Figure 6), containing the key enzyme catechol 1,2-dioxygenase. This enzyme was described in well-known biodegradative strains, such as *Pseudomonas* sp. strain EST1001 and *Acinetobacter baylyi* strain ADP1, which catabolize a wide range of aromatic compounds (Neidle et al., 1988; Kivisaar et al., 1991). The presence of the catechol *ortho*-cleavage pathway was correlated with benzoate degradation (Figure 6), since the same genomes that contained the complete catechol degradation route harbored genes codifying for a three-component benzoate 1,2-dioxygenase (BenABC) and 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase (BenD) (Figure 6; Supplementary Table 4). This is usually required for complete benzoate degradation by bacterial species (Pérez-Pantoja et al., 2012). Exceptionally, the gene encoding the large subunit of benzoate dioxygenase (*benA*) was found to be frame-shifted in the genome sequence of *S. balearica* DSM 6083[T], suggesting a non-functional benzoate degradation pathway in this bacterium. Of these strains with genes for benzoate degradation, four could be tested for growth on benzoate as a sole carbon and energy source, of which only two (*S. balearica* strain LS401 and strain st101) were able to grow adequately (Table 1). *S. balearica* DSM 6083[T] did not grow on benzoate, which was expected, due to the frame-shift in the *benA* gene, while *S. balearica* strain CCUG 18844 (= FDAARGOS_1013) grew very weakly, despite harboring all necessary genes and being previously reported as a benzoate degrader (Pichinoty et al., 1977). However, this strain also grew very weakly on phenylalanine, tyrosine (Table 1) and, even, on non-aromatic compounds, such as succinate. Thus, a possible explanation is that the strain may harbor auxotrophic features that result in limited growth on minimal media.

Three isolated strains (*S. balearica* DSM 6083[T], LS401, and EC28) encoded for the entire catechol *meta*-cleavage pathway (Figure 6), harboring a key catechol 2,3-dioxygenase similar to those contained in plasmids pWW0, pNAH7 and pDK1. These plasmids are found in *Pseudomonas* species associated with the catabolism of toluene, naphthalene and phenol, respectively (Harayama et al., 1987; Bartilson and Shingler, 1989; Benjamin et al., 1991). Moreover, the genome sequences of two strains (*S. balearica* DSM 6083[T] and strain LS401), which harbor the catechol *meta*-cleavage pathway, codified all components for channeling naphthalene into catechol (Figure 6), and both strains were able to grow using naphthalene and salicylate as sole carbon and energy sources (Table 1), which is in agreement with a previous study (Rossello et al., 1991). *S. balearica* strain CCUG 66667 (= st101), despite being originally described as naphthalene degrader (Dutta, 2001), was not able to grow on naphthalene as a sole carbon and energy source and no genes for naphthalene degradation were detected in its genome sequence, as previously reported (Salvà-Serra et al., 2017). This suggests that *S. balearica* strain st101 may have lost the naphthalene degradation capacity, which could be related to mobile genetic elements, such as plasmids, transposons, or integrative conjugative elements, being unstable in the absence of selective pressure (Phale et al., 2019). As well, one of the 11 isolate-derived genome sequences (*S. balearica* EC28) and one MAG (*S. balearica* UBA3230) encoded for a multicomponent phenol 2-monooxygenase (Figure 6), comparable

**FIGURE 6**

Pathways for degradation of aromatic compounds in *S. balearica*. **(A)** Overview of putative peripheral and ring-cleavage pathways of aromatic compounds found in *S. balearica* genome sequences. Strains containing the full set of genes for a biodegradative route are indicated in each colored box. **(B)** Schematic representation of presence/absence (blue/white coloring) pattern of genes belonging to central and peripheral aromatic pathways in genome sequences of *S. balearica*. Red indicates incomplete or frame-shifted gene. [a]MAGs with a CheckM completeness estimate lower than 95%.

**TABLE 1** Growth assays performed on aromatic compounds with the six *S. balearica* strains that are available at the CCUG.

| *S. balearica* strain | Benzoate | Phenol | Salicylate | Naphthalene | Phenylalanine | Tyrosine |
|---|---|---|---|---|---|---|
| CCUG 44487[T] (= DSM 6083[T] = SP1401[T]) | − | − | + | + | +/− | + |
| CCUG 66666 (= LS401) | + | − | + | + | +/− | + |
| CCUG 66667 (= st101) | + | − | − | − | +/− | + |
| CCUG 70972 (= Z8) | − | − | − | − | +/− | + |
| CCUG 72049 (= SAGV3-2SA2) | − | − | − | − | +/− | + |
| CCUG 18844 (= DSM 46326 = FDAARGOS_1013) | +/− | − | − | − | +/− | +/− |

−: assays showed no growth after 5 days.

+/−: assays showed very weak growth, in all cases not higher than $OD_{600} = 0.1$.

+: assays showed distinct growth, in all cases higher than $OD_{600} = 0.4$.

to that found in *Pseudomonas* sp. strain CF600, which has been related to phenol catabolism through catechol *meta*-cleavage pathway (Shingler et al., 1989; Nordlund et al., 1990), and that found in *S. chloritidismutans* strain ST-9 (Michael et al., 2017). Additionally, all sequences except various MAGs, harbored the homogentisate ring-cleavage pathway (Figure 6), which is associated with catabolism of the aromatic amino acids tyrosine and phenylalanine (Arias-Barrau et al., 2004). In this context, all six strains were able to grow on tyrosine, but only weakly on phenylalanine as a sole carbon source (Table 1), similar to the phenotype reported in *Pseudomonas putida* strain KT2440, in which growth on phenylalanine as the sole carbon source was almost negligible, in contrast to growth with phenylalanine as a sole nitrogen source (Herrera and Ramos, 2007). Genes encoding for extra 45 peripheral and 18 central enzymes for degradation of aromatic compounds were also searched for but were not detected in the analyzed genome sequences (Supplementary Table 4). However, the pan-genomic analysis indicates that only a fraction of the predicted pool of genes of *S. balearica* has been included in this study and that more strains of *S. balearica* will reveal novel genes and pathways. Additionally, it should be noted that four of the seven MAGs included in this study have low completeness levels (67–68%), which indicates that significant fractions of their genomes were missing. This implies that further pathways and capabilities in *S. balearica* strains for biodegradation of aromatic compounds, complementary to homogentisate- and catechol *ortho*- and *meta*-cleavage pathways, might be revealed in future studies.

## 3.6. Additional genomic features of *Stutzerimonas balearica*

The analysis of the genome sequences of *S. balearica*, using the tool RGI (CARD), did not reveal any antibiotic resistance genes that seemed acquired horizontally (Supplementary Table 4). This could be expected as none of the 18 genome sequences is derived from an environment with apparent high antibiotic selective pressure. However, the natural transformation capacity reported for multiple members of *Stutzerimonas* and related species (Carlson et al., 1983; Lorenz and Sikorski, 2000), could potentially facilitate the acquisition of antibiotic resistance genes when exposed to antibiotic pressure, such as the reported VIM metallo-β-lactamase in a *S. balearica* clinical isolate (Uddin et al., 2022). Indeed, genes codifying for type IV pili components, which are essential for natural genetic transformation (Graupner et al., 2000, 2001), were found in all genome sequences (Supplementary Table 4). The gene *pilAI*, which is located upstream of *pilB*, in opposite direction, and is essential for pili formation (Graupner et al., 2000), was not detected by the software Vfanalyzer. However, a gene encoding a protein with low amino acid identity to PilAI (39% in *S. balearica* DSM 6083[T]) was found upstream of *pilB* in almost all genome sequences. The genes *comA*, also essential for natural transformation, and *exbB*, which inactivation has been shown to decrease the level of natural transformation (Graupner and Wackernagel, 2001), were found also in all isolate-derived genome sequences and five MAGs. This suggests that natural transformation might have played an important role in the evolution and diversification of *S. balearica*. However, future studies are necessary to elucidate the capacity of *S. balearica* for natural transformation.

The RGI analysis revealed subunits of efflux pumps of the resistance-nodulation-cell division (RND) and small multidrug resistance (SMR) families, which can play significant roles in resistance by extruding multiple antibiotics (Bay and Turner, 2016; Murakami, 2016). Meanwhile, the search of antibacterial biocide and metal resistance genes revealed a wide range of genes (Supplementary Table 4), which might be advantageous for strains of *S. balearica* to thrive in diverse polluted and anthropogenic environments. Many were widely distributed among the genomes and MAGs and, indeed, are intrinsic of *Pseudomonas* (e.g., the RND efflux pump MexEF-OprN, which in *P. aeruginosa* has been associated with extrusion of multiple antibiotics but also to organic solvents such as *n*-hexane and *p*-xylene) (Köhler et al., 1997; Li et al., 1998), while others were more strain-specific. For instance, *S. balearica* strain YC-YH1 was the only strain containing (and with high amino acid sequence identity to the reference sequence: greater than 98%) the genes coding for all components of TbtABM, an RND multidrug efflux pump that, in *S. stutzeri*, has been associated with resistance to the tributyltin and decreased susceptibility to various antibiotics (Jude et al., 2004). Meanwhile, *S. balearica* strain EC28 was the only strain carrying the gene *qacE*, which encodes an SMR transporter (Quaternary ammonium compound-resistance protein QacE; also with greater than 98% sequence identity to the reference sequence), which has been associated with resistance to intercalating dyes and quaternary ammonium compounds (Paulsen et al., 1993). A homolog of this protein was also found in the MAG *S. balearica* 3300027365_7, although with lower identity to the reference (73.6%).

Multiple factors that have been associated with virulence in *Pseudomonas* were also detected in *S. balearica* genome sequences (Supplementary Table 4). These included genes related with lipopolysaccharide synthesis, the main antigen (O-antigen) and a major and diverse virulence factor of *P. aeruginosa* (Pier, 2007), which has also been shown to be highly diverse in *Stutzerimonas* (Rosselló et al., 1992). Numerous protein sequences were linked to flagella biosynthesis, in agreement with the description of the species (Bennasar et al., 1996), and which has been linked to biofilm formation in *P. aeruginosa* (O'Toole and Kolter, 1998). Other sequences were related with type IV pili biogenesis, which confers twitching motility and which have also been linked to capacity for biofilm formation in *P. aeruginosa* (O'Toole and Kolter, 1998) and to genetic transformation in *P. stutzeri* (Graupner et al., 2000). The genome sequences also revealed multiple elements related with the regulation and biosynthesis of alginate, which in *P. aeruginosa* has been shown to play an important role in biofilm formation and to protect bacteria from external adversities (Nivens et al., 2001). Genes encoding a complete secretion system were found, homologous to the Xcp secretion system of *P. aeruginosa*, a type II secretion system associated with the secretion of numerous substrates encoded all over the chromosome, including toxins and enzymes (Filloux, 2011). In the case of *S. balearica* DSM 6083[T], nine of the genes were structured in a single cluster (*xcpR*, *xcpS*, *xcpT*, *xcpU*, *xcpV*, *xcpW*, *xcpX*, *xcpY*, *xcpZ*), while two others were co-located elsewhere in the chromosome (*xcpP*, *xcpQ*). The twelfth gene (*xcpA*), essential for the proper functioning of the Xcp system (Bally et al., 1992), was found in a different region of the chromosome. Additionally, several urease-related genes were detected, including a urea uptake system (*urtA*, *urtB*, *urtC*, *urtD*, *urtE*), urease subunits (*ureA*, *ureB*, *ureC*) and urease accessory proteins (*ureD*, *ureE*, *ureF*, *ureG*, *ureJ*).

The genome sequences also revealed that multiple strains of *S. balearica* contain a CRISPR-Cas system (Supplementary Table 4), which are adaptive immune mechanisms that are present in approximately 10% of bacteria (Burstein et al., 2016). The genome sequence of *S. balearica* DSM 6083[T] revealed a full CRISPR-Cas system subtype I-E (Makarova et al., 2011), comprising nine CRISPR-associated genes and two CRISPR arrays. One CRISPR array was located upstream of the CRISPR-associated genes and contained 69 direct repeats and 68 non-redundant spacer sequences. The other was located downstream, with six direct repeats and five spacers. No leader sequence was detected for this array, which might be the reason of its low number of spacers (Alkhnbashi et al., 2016; Fernández Juárez, 2017). CRISPR-Cas systems with the same genetic structure and a varying number of spacer/repeats were found in five other genome sequences of *S. balearica* strains and three MAGs. The number of spacer/repeats in the other genome sequences ranged from 16/17 to 103/104. An additional potential CRISPR-Cas system of type IV was found in MAG101, containing a CRISPR array formed by 11 spacers and 12 repeats. Viruses are the most common biological entities in the marine environment and can be significant controlling agents on marine bacterial communities (Fuhrman, 1999). Therefore, CRISPR-Cas systems might be critical systems for certain strains of *S. balearica*. In fact, genome analyses also revealed intact prophages in nine of the 18 genome sequences of *S. balearica*, with sizes ranging from 34 to 83 kb (Supplementary Table 4).

Although no plasmids were revealed in the complete genomes, putative ICEs with type IV secretion systems (T4SS) were predicted in six of the isolate-derived genome sequences and four MAGs. The draft status of most of the genomes impeded the estimation of the size of multiple ICEs that were predicted to be split in multiple contigs or scaffolds. However, those predicted on a single contig or scaffold or in complete genome sequences exhibited sizes ranging from 58 to 269 kb (Supplementary Table 4). Multiple IMEs were also predicted, with sizes ranging from 2.3 to 15.6 kb. In *S. balearica* strain FDAARGOS_1013 (= CCUG 18844 = DSM 46326), three ICEs were predicted, totaling 404 kb (i.e., 9.1% of the genome). In the case of *S. balearica* DSM 6083[T], the ICE encoded the naphthalene upper pathway and the catechol *meta*-cleavage pathway. This was observed also in the genome of *S. chloritidismutans* AN10 (formerly, *P. stutzeri* genomovar 3) (Bosch et al., 1999, 2000; Hirose et al., 2021). Among others, ICEs and IMEs encode multiple transporter systems, transposase genes, toxin-antitoxin systems as well as numerous hypothetical proteins. In any case, predictions of these mobile genetic elements should be confirmed in future studies.

The genome sequences also revealed that each *S. balearica* strain carries between 306 and 361 regulatory genes (7.7–8.3% of the isolate-derived genome sequences) (Supplementary Table 4). These included tens of two-component systems, transcription factors (i.e., transcriptional regulators, one-component systems, response regulators and sigma factors) and other DNA-binding proteins. These findings are in accordance with previous reports in *Pseudomonas* (Stover et al., 2000) and suggest that *S. balearica* has a large capacity for adapting to varying environmental conditions and responding to external factors.

## 4. Conclusion

*S. balearica* is a diverse species with an open pan-genome that will continue to reveal new genes as the genomes of new strains are sequenced. The detection of 176 strains of the species revealed that *S. balearica* has been found in a wide range of environments, although principally in aquatic and polluted environments, most of which were associated with the oil industry. The genomes of strains of *S. balearica* indicate that the species has a diverse potential for biodegradation of aromatic compounds, which has been experimentally confirmed in six of the strains, and presents multiple features that increase our understanding of the diversity and the lifestyle of *S. balearica*.

## Data availability statement

The strains used in this study are available at the Culture Collection University of Gothenburg (CCUG, Gothenburg, Sweden; www.ccug.se). The draft genome sequence of *S. balearica* strain SAGV3-2SA2 (= CCUG 72049) is deposited and publicly available in DDBJ/ENA/GenBank under the accession no. RAXS00000000. The Illumina sequence reads have been deposited at the Sequence Read Archive (SRA) (Leinonen et al., 2011) under the accession number SRR14574958.

## Author contributions

FS-S, DP-P, JL, and AB-F: conceptualization. FS-S, DP-P, RD, DJ-L, and VF-J: methodology. FS-S, DP-P, RD, DJ-L, and AB-F: validation. FS-S and AB-F: original draft preparation. FS-S, DP-P, RD, DJ-L, VF-J, HE-J, EM, JL, and AB-F: review and editing. HE-J, EM, and AB-F: supervision. EM, JL, and AB-F: project administration. DP-P, EM, JL, and AB-F: funding acquisition. All authors contributed to the article and approved the submitted version.

Malmborg and the staff at the Substrate Department of the Department of Clinical Microbiology, Sahlgrenska University Hospital, are acknowledged for media preparation. The computations were partially performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project SNIC 2019/8-176.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher

## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1159176/full#supplementary-material

**SUPPLEMENTARY FIGURE 1**
Phylogenetic tree based on partial *rpoD* sequences (549 bp). The tree includes 111 of the 113 genome sequences included in this study; the other two did not contain any *rpoD* sequence in the assembly. The phylogenetic analysis also includes the reference *rpoD* sequences of most of the described species and phylogenomic species of *Stutzerimonas*. The GenBank accession numbers of the genome sequences are listed in Supplementary Table 1 and those of the reference sequences obtained from the PseudoMLSA database are indicated in parenthesis. For each strain, the

final taxonomic assignment is indicated. The final phylogenomic species assignment of each strain is indicated in parenthesis. The distances were calculated, using the Jukes-Cantor method, and the tree was constructed, using the neighbor-joining method. Bootstrap values of 50% or greater (from 1,000 replicates) are shown at the nodes. *P. aeruginosa* CCM 1960$^T$ was used as an outgroup. Pgs: phylogenomic species; gv.: genomovar; ref.: reference.

**SUPPLEMENTARY FIGURE 2**
Core genome-based phylogenomic tree of 11 isolate-derived genome sequences and seven MAGs of *S. balearica*, 92 strains of other species and phylogenomic groups of the genus *Stutzerimonas*, and the type strains of *P. indica*, *P. kuykendallii* and *P. matsuisoli*. The tree was generated, using 35,374 amino acid positions derived from 127 single copy core proteins. Pgs: phylogenomic species; ref.: reference.

**SUPPLEMENTARY TABLE 1**
List of the 113 genome sequences used in this study and their associated metadata.

**SUPPLEMENTARY TABLE 2**
ANIb values between the 113 genome sequences (all vs. all) included in this study. Each pairwise comparison was performed bidirectionally; the average value is displayed.

**SUPPLEMENTARY TABLE 3**
List of the 176 confirmed *S. balearica* strains and their associated metadata.

**SUPPLEMENTARY TABLE 4**
Genomic features of the genome sequences of *S. balearica*. Sheet S1: Pan-genes of *S. balearica*. Sheet S2: Genes encoding enzymes for catabolism of aromatic compounds detected in one or more genome sequences of *S. balearica*. Sheet S3: Genes for catabolism of aromatic compounds not detected in the genome sequences of *S. balearica*. Sheet S4: Antibiotic resistance genes detected, using Resistance Gene Identifier (RGI) and the Comprehensive Antibiotic Resistance Database (CARD). Sheet S5: Biocide and metal resistance genes predicted by BacMet. Sheet S6: Virulence determinants detected, using VFanalyzer and the VFDB. Sheet S7: CRISPR-Cas systems predicted by CRISPRone. Sheet S8: Prophages predicted by PHASTER. Sheet S9: Integrative and conjugative elements (ICEs) and integrative and mobilizable elements (IMEs) predicted by ICEfinder. Sheet S10: genes associated with natural transformation capacity. Sheet S11: Regulatory genes predicted, using P2RP.

**SUPPLEMENTARY FILE 1**
Metagenome assembled genome (MAG) sequence of *S. balearica* UBA3230, annotated using DFAST, in GBK format.

**SUPPLEMENTARY FILE 2**
Metagenome assembled genome (MAG) sequence of *S. balearica* UBA6635, annotated using DFAST, in GBK format.

**SUPPLEMENTARY FILE 3**
Metagenome assembled genome (MAG) sequence of *S. balearica* 3300027365_7, annotated using DFAST, in GBK format.

## References

Agency for Toxic Substances and Disease Registry (ATSDR), Division of Toxicology and Human Health Sciences (2019). ATSPDR's Substance Priority List. Available at: https://www.atsdr.cdc.gov/spl/index.html (Accessed October 3, 2022).

Ahmadi, M., Jorfi, S., Kujlu, R., Ghafari, S., Darvishi Cheshmeh Soltani, R., and Jaafarzadeh Haghighifard, N. (2017). A novel salt-tolerant bacterial consortium for biodegradation of saline and recalcitrant petrochemical wastewater. *J. Environ. Manag.* 191, 198–208. doi: 10.1016/j.jenvman.2017.01.010

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Alkhnbashi, O. S., Shah, S. A., Garrett, R. A., Saunders, S. J., Costa, F., and Backofen, R. (2016). Characterizing leader sequences of CRISPR loci. *Bioinformatics* 32, i576–i585. doi: 10.1093/bioinformatics/btw454

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

Arias-Barrau, E., Olivera, E. R., Luengo, J. M., Fernández, C., Galán, B., García, J. L., et al. (2004). The homogentisate pathway: a central catabolic pathway involved in the degradation of L-phenylalanine, L-tyrosine, and 3-hydroxyphenylacetate in *Pseudomonas putida*. *J. Bacteriol.* 186, 5062–5077. doi: 10.1128/JB.186.15.5062-5077.2004

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., and Weightman, A. J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl. Environ. Microbiol.* 71, 7724–7736. doi: 10.1128/aem.71.12.7724-7736.2005

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., et al. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44, W147–W153. doi: 10.1093/nar/gkw419

Bally, M., Filloux, A., Akrim, M., Ball, G., Lazdunski, A., and Tommassen, J. (1992). Protein secretion in *Pseudomonas aeruginosa*: characterization of seven *xcp* genes and

processing of secretory apparatus components by prepilin peptidase. *Mol. Microbiol.* 6, 1121–1131. doi: 10.1111/j.1365-2958.1992.tb01550.x

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barakat, M., Ortet, P., and Whitworth, D. E. (2013). P2RP: a web-based framework for the identification and analysis of regulatory proteins in prokaryotic genomes. *BMC Genomics* 14:269. doi: 10.1186/1471-2164-14-269

Bartilson, M., and Shingler, V. (1989). Nucleotide sequence and expression of the catechol 2,3-dioxygenase-encoding gene of phenolcatabolizing *Pseudomonas* CF600. *Gene* 85, 233–238. doi: 10.1016/0378-1119(89)90487-3

Bay, D. C., and Turner, R. J. (2016). "Small Multidrug Resistance Efflux Pumps" in *Efflux-mediated antimicrobial resistance in bacteria: Mechanisms, regulation and clinical implications.* eds. X.-Z. Li, C. A. Elkins and H. I. Zgurskaya (Cham: Springer International Publishing)

Beauchamp, C. J., Lévesque, G., Prévost, D., and Chalifour, F.-P. (2006). Isolation of free-living dinitrogen-fixing bacteria and their activity in compost containing de-inking paper sludge. *Bioresour. Technol.* 97, 1002–1011. doi: 10.1016/j.biortech.2005.04.041

Benjamin, R. C., Voss, J. A., and Kunz, D. A. (1991). Nucleotide sequence of *xylE* from the TOL pDK1 plasmid and structural comparison with isofunctional catechol-2,3-dioxygenase genes from TOL, pWW0 and NAH7. *J. Bacteriol.* 173, 2724–2728. doi: 10.1128/jb.173.8.2724-2728.1991

Bennasar, A., Mulet, M., Lalucat, J., and García-Valdés, E. (2010). PseudoMLSA: a database for multigenic sequence analysis of *Pseudomonas* species. *BMC Microbiol.* 10:118. doi: 10.1186/1471-2180-10-118

Bennasar, A., Rosselló-Mora, R., Lalucat, J., and Moore, E. R. (1996). 16S rRNA gene sequence analysis relative to genomovars of *Pseudomonas stutzeri* and proposal of *Pseudomonas balearica* sp. nov. *Int. J. Syst. Bacteriol.* 46, 200–205. doi: 10.1099/00207713-46-1-200

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., et al. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* 49, D344–D354. doi: 10.1093/nar/gkaa977

Bosch, R., García-Valdés, E., and Moore, E. R. (1999). Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. *Gene* 236, 149–157. doi: 10.1016/S0378-1119(99)00241-3

Bosch, R., García-Valdés, E., and Moore, E. R. (2000). Complete nucleotide sequence and evolutionary significance of a chromosomally encoded naphthalene-degradation lower pathway from *Pseudomonas stutzeri* AN10. *Gene* 245, 65–74. doi: 10.1016/S0378-1119(00)00038-X

Bravakos, P., Mandalakis, M., Nomikou, P., Anastasiou, T. I., Kristoffersen, J. B., Stavroulaki, M., et al. (2021). Genomic adaptation of *Pseudomonas* strains to acidity and antibiotics in hydrothermal vents at Kolumbo submarine volcano Greece. *Sci. Rep.* 11:1336. doi: 10.1038/s41598-020-79359-y

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., et al. (2016). Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems. *Nat. Commun.* 7:10613. doi: 10.1038/ncomms10613

Busquets, A., Peña, A., Gomila, M., Bosch, R., Nogales, B., García-Valdés, E., et al. (2012). Genome sequence of *Pseudomonas stutzeri* strain JM300 (DSM 10701), a soil isolate and model organism for natural transformation. *J. Bacteriol.* 194, 5477–5478. doi: 10.1128/JB.01257-12

Carlson, C. A., Pierson, L. S., Rosen, J. J., and Ingraham, J. L. (1983). *Pseudomonas stutzeri* and related species undergo natural transformation. *J. Bacteriol.* 153, 93–99. doi: 10.1128/jb.153.1.93-99.1983

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2011). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28, 464–469. doi: 10.1093/bioinformatics/btr703

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A. M., and Banfield, J. F. (2020). Accurate and complete genomes from metagenomes. *Genome Res.* 30, 315–333. doi: 10.1101/gr.258640.119

Cladera, A. M., Garcia-Valdés, E., and Lalucat, J. (2006). Genotype versus phenotype in the circumscription of bacterial species: the case of *Pseudomonas stutzeri* and *Pseudomonas chloritidismutans*. *Arch. Microbiol.* 184, 353–361. doi: 10.1007/s00203-005-0052-x

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610

Contreras-Moreira, B., and Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 79, 7696–7701. doi: 10.1128/AEM.02411-13

Crone, S., Vives-Flórez, M., Kvich, L., Saunders, A. M., Malone, M., Nicolaisen, M. H., et al. (2019). The environmental occurrence of *Pseudomonas aeruginosa*. *APMIS* 128, 220–231. doi: 10.1111/apm.13010

Dutta, J. (2001). Isolation and characterization of polycyclic aromatic hydrocarbon degrading bacteria from the rhizosphere of salt-marsh plants. MSc thesis, Rutgers, The State University of New Jersey.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Edgar, R. (2018). Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ* 6:e5030. doi: 10.7717/peerj.5030

Fernández Juárez, V. (2017). Clonación del sistema CRISPR-Cas de *Pseudomonas balearica* DSM 6083$^T$ en *Escherichia coli*. MSc., University of the Balearic Islands.

Ferrero, M., Llobet-Brossa, E., Lalucat, J., García-Valdés, E., Rosselló-Mora, R., and Bosch, R. (2002). Coexistence of two distinct copies of naphthalene degradation genes in *Pseudomonas* strains isolated from the western Mediterranean region. *Appl. Environ. Microbiol.* 68, 957–962. doi: 10.1128/AEM.68.2.957-962.2002

Filloux, A. (2011). Protein secretion systems in *Pseudomonas aeruginosa*: an essay on diversity, evolution, and function. *Front. Microbiol.* 2:e00155. doi: 10.3389/fmicb.2011.00155

Freschi, L., Vincent, A. T., Jeukens, J., Emond-Rheault, J.-G., Kukavica-Ibrulj, I., Dupont, M.-J., et al. (2019). The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol. Evol.* 11, 109–120. doi: 10.1093/gbe/evy259

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548. doi: 10.1038/21119

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

García-Valdés, E., Mulet, M., and Lalucat, J. (2010). "Insights into the life styles of *Pseudomonas stutzeri*" in *Pseudomonas: Volume 6: Molecular microbiology, infection and biodiversity*. eds. J. L. Ramos and A. Filloux (Dordrecht: Springer Netherlands)

Girard, L., Lood, C., Höfte, M., Vandamme, P., Rokni-Zadeh, H., van Noort, V., et al. (2021). The ever-expanding *Pseudomonas genus*: description of 43 new species and partition of the *Pseudomonas putida* group. *Microorganisms* 9:1766. doi: 10.3390/microorganisms9081766

Gomila, M., Busquets, A., Mulet, M., García-Valdés, E., and Lalucat, J. (2017). Clarification of taxonomic status within the *Pseudomonas syringae* species group based on a phylogenomic analysis. *Front. Microbiol.* 8:e02422. doi: 10.3389/fmicb.2017.02422

Gomila, M., Mulet, M., García-Valdés, E., and Lalucat, J. (2022). Genome-based taxonomy of the genus *Stutzerimonas* and proposal of *S. frequens* sp. nov. and S. *degradans* sp. nov. and emended descriptions of *S. perfectomarina* and *S. chloritidismutans*. *Microorganisms* 10, 10:1363. doi: 10.3390/microorganisms10071363

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Graupner, S., Frey, V., Hashemi, R., Lorenz, M. G., Brandes, G., and Wackernagel, W. (2000). Type IV pilus genes *pilA* and *pilC* of *Pseudomonas stutzeri* are required for natural genetic transformation, and *pilA* can be replaced by corresponding genes from nontransformable species. *J. Bacteriol.* 182, 2184–2190. doi: 10.1128/JB.182.8.2184-2190.2000

Graupner, S., and Wackernagel, W. (2001). Identification and characterization of novel competence genes *comA* and *exbB* involved in natural genetic transformation of *Pseudomonas stutzeri*. *Res. Microbiol.* 152, 451–460. doi: 10.1016/S0923-2508(01)01218-9

Graupner, S., Weger, N., Sohni, M., and Wackernagel, W. (2001). Requirement of novel competence genes *pilT* and *pilU* of *Pseudomonas stutzeri* for natural transformation and suppression of *pilT* deficiency by a hexahistidine tag on the type IV pilus protein PilAI. *J. Bacteriol.* 183, 4694–4701. doi: 10.1128/JB.183.16.4694-4701.2001

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Harayama, S., Rekik, M., Wasserfallen, A., and Bairoch, A. (1987). Evolutionary relationships between catabolic pathways for aromatics: conservation of gene order and nucleotide sequences of catechol oxidation genes of pWW0 and NAH7 plasmids. *Mol. Gen. Genet.* 210, 241–247. doi: 10.1007/BF00325689

Herrera, M. C., and Ramos, J.-L. (2007). Catabolism of phenylalanine by *Pseudomonas putida*: the NtrC-family PhhR regulator binds to two sites upstream from the *phhA* gene and stimulates transcription with σ$^{70}$. *J. Mol. Biol.* 366, 1374–1386. doi: 10.1016/j.jmb.2006.12.008

Hesse, C., Schulz, F., Bull, C. T., Shaffer, B. T., Yan, Q., Shapiro, N., et al. (2018). Genome-based evolutionary history of *Pseudomonas* spp. *Environ. Microbiol.* 20, 2142–2159. doi: 10.1111/1462-2920.14130

Hirose, J., Watanabe, T., Futagami, T., Fujihara, H., Kimura, N., Suenaga, H., et al. (2021). A new ICE*clc* subfamily integrative and conjugative element responsible for

horizontal transfer of biphenyl and salicylic acid catabolic pathway in the PCB-degrading strain *Pseudomonas stutzeri* KF716. *Microorganisms* 9:2462. doi: 10.3390/microorganisms9122462

Hreha, T. N., Foreman, S., Duran-Pinedo, A., Morris, A. R., Diaz-Rodriguez, P., Jones, J. A., et al. (2021). The three NADH dehydrogenases of *Pseudomonas aeruginosa*: their roles in energy metabolism and links to virulence. *PLoS One* 16:e0244142. doi: 10.1371/journal.pone.0244142

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47, D309–D314. doi: 10.1093/nar/gky1085

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Joshi, N.A., and Fass, J.N. (2011). Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files Version 1.33.

Jude, F., Arpin, C., Brachet-Castang, C., Capdepuy, M., Caumette, P., and Quentin, C. (2004). TbtABM, a multidrug efflux pump associated with tributyltin resistance in *Pseudomonas stutzeri*. *FEMS Microbiol. Lett.* 232, 7–14. doi: 10.1016/S0378-1097(04)00012-6

Jukes, T. H., and Cantor, C. R. (1969). "CHAPTER 24 - evolution of protein molecules A2 - MUNRO, H.N" in *Mammalian Protein Metabolism* (Cambridge: Academic Press)

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2020). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. doi: 10.1093/nar/gkaa970

Kivisaar, M., Kasak, L., and Nurk, A. (1991). Sequence of the plasmid-encoded catechol 1,2-dioxygenase-expressing gene, *pheB*, of phenol-degrading *Pseudomonas* sp. strain EST1001. *Gene* 98, 15–20. doi: 10.1016/0378-1119(91)90098-V

Kizhakedathil, M. P. J., and Subathra, D. C. (2021). Acid stable α-amylase from *Pseudomonas balearica* VITPS19—production, purification and characterization. *Biotechnol. Rep.* 30:e00603. doi: 10.1016/j.btre.2021.e00603

Köhler, T., Michéa-Hamzehpour, M., Henze, U., Gotoh, N., Kocjancic Curty, L., and Pechère, J.-C. (1997). Characterization of MexE–MexF–OprN, a positively regulated multidrug efflux system of *Pseudomonas aeruginosa*. *Mol. Microbiol.* 23, 345–354. doi: 10.1046/j.1365-2958.1997.2281594.x

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* 36, 6688–6719. doi: 10.1093/nar/gkn668

Köse, M., Oztürk, M., Kuyucu, T., Günes, T., Akcakus, M., and Sümerkan, B. (2004). Community-acquired pneumonia and empyema caused by *Pseudomonas stutzeri*: a case report. *Turk. J. Pediatr.* 46, 177–178.

Kristensen, D. M., Kannan, L., Coleman, M. K., Wolf, Y. I., Sorokin, A., Koonin, E. V., et al. (2010). A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* 26, 1481–1487. doi: 10.1093/bioinformatics/btq229

Kumar, A., Saini, H. S., and Kumar, S. (2018). Bioleaching of gold and silver from waste printed circuit boards by *Pseudomonas balearica* SAE1 isolated from an e-waste recycling facility. *Curr. Microbiol.* 75, 194–201. doi: 10.1007/s00284-017-1365-0

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., and Ussery, D. W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35, 3100–3108. doi: 10.1093/nar/gkm160

Lalucat, J., Bennasar, A., Bosch, R., García-Valdés, E., and Palleroni, N. J. (2006). Biology of *Pseudomonas stutzeri*. *Microbiol. Mol. Biol. Rev.* 70, 510–547. doi: 10.1128/MMBR.00047-05

Lalucat, J., Gomila, M., Mulet, M., Zaruma, A., and García-Valdés, E. (2022). Past, present and future of the boundaries of the *Pseudomonas* genus: proposal of *Stutzerimonas* gen. nov. *Syst. Appl. Microbiol.* 45, 126289. doi: 10.1016/j.syapm.2021.126289

Lalucat, J., Mulet, M., Gomila, M., and García-Valdés, E. (2020). Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Genes* 11:139. doi: 10.3390/genes11020139

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30, 3276–3278. doi: 10.1093/bioinformatics/btu531

Lee, S. E., Chung, J. W., Won, H. S., Lee, D. S., and Lee, Y.-W. (2012). Removal of methylmercury and tributyltin (TBT) using marine microorganisms. *Bull. Environ. Contam. Toxicol.* 88, 239–244. doi: 10.1007/s00128-011-0501-y

Leinonen, R., Sugawara, H., and Shumway, M. International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019

Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Li, X., Yang, Z., Wang, Z., Li, W., Zhang, G., and Yan, H. (2022). Comparative genomics of *Pseudomonas stutzeri* complex: taxonomic assignments and genetic diversity. *Front. Microbiol.* 12, 755874. doi: 10.3389/fmicb.2021.755874

Li, X.-Z., Zhang, L., and Poole, K. (1998). Role of the multidrug efflux systems of *Pseudomonas aeruginosa* in organic solvent tolerance. *J. Bacteriol.* 180, 2987–2991. doi: 10.1128/JB.180.11.2987-2991.1998

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2018). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 47, D660–D665. doi: 10.1093/nar/gky1123

Liu, B., Zheng, D., Zhou, S., Chen, L., and Yang, J. (2022). VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res.* 50, D912–D917. doi: 10.1093/nar/gkab1107

Loman, N. J., and Pallen, M. J. (2015). Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* 13, 787–794. doi: 10.1038/nrmicro3565

Lorenz, M. G., and Sikorski, J. (2000). The potential for intraspecific horizontal gene exchange by natural genetic transformation: sexual isolation among genomovars of *Pseudomonas stutzeri*. *Microbiology* 146, 3081–3090. doi: 10.1099/00221287-146-12-3081

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2019). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., et al. (2011). Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* 9, 467–477. doi: 10.1038/nrmicro2577

Mandalakis, M., Gavriilidou, A., Polymenakou, P. N., Christakis, C. A., Nomikou, P., Medvecký, M., et al. (2019). Microbial strains isolated from $CO_2$-venting Kolumbo submarine volcano show enhanced CO-tolerance to acidity and antibiotics. *Mar. Environ. Res.* 144, 102–110. doi: 10.1016/j.marenvres.2019.01.002

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* 14:60. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902

Michael, E., Gomila, M., Lalucat, J., Nitzan, Y., Pechatnikov, I., and Cahan, R. (2017). Proteomic assessment of the expression of genes related to toluene catabolism and porin synthesis in *Pseudomonas stutzeri* ST-9. *J. Proteome Res.* 16, 1683–1692. doi: 10.1021/acs.jproteome.6b01044

Mulet, M., Gomila, M., Gruffaz, C., Meyer, J. M., Palleroni, N. J., Lalucat, J., et al. (2008). Phylogenetic analysis and siderotyping as useful tools in the taxonomy of *Pseudomonas stutzeri*: description of a novel genomovar. *Int. J. Syst. Evol. Microbiol.* 58, 2309–2315. doi: 10.1099/ijs.0.65797-0

Mulet, M., Gomila, M., Lalucat, J., Bosch, R., Rossello-Mora, R., and García-Valdés, E. (2023). *Stutzerimonas decontaminans* sp. nov. isolated from marine polluted sediments. *Syst. Appl. Microbiol.* 46:126400. doi: 10.1016/j.syapm.2023.126400

Mulet, M., Gomila, M., Ramírez, A., Cardew, S., Moore, E. R., Lalucat, J., et al. (2017). Uncommonly isolated clinical *Pseudomonas*: identification and phylogenetic assignation. *Eur. J. Clin. Microbiol. Infect. Dis.* 36, 351–359. doi: 10.1007/s10096-016-2808-4

Mulet, M., Lalucat, J., and Garcia-Valdes, E. (2010). DNA sequence-based analysis of the *Pseudomonas* species. *Environ. Microbiol.* 12, 1513–1530. doi: 10.1111/j.1462-2920.2010.02181.x

Murakami, S. (2016). "Structures and transport mechanisms of RND efflux pumps" in *Efflux-mediated antimicrobial resistance in bacteria: Mechanisms, regulation and clinical implications.* eds. X.-Z. Li, C. A. Elkins and H. I. Zgurskaya (Cham: Springer International Publishing)

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., et al. (2021). A genomic catalog of Earth's microbiomes. *Nat. Biotechnol.* 39, 499–509. doi: 10.1038/s41587-020-0718-6

Neidle, E. L., Hartnett, C., Bonitz, S., and Ornston, L. N. (1988). DNA sequence of the *Acinetobacter calcoaceticus* catechol 1,2-dioxygenase I structural gene *catA*: evidence for evolutionary divergence of intradiol dioxygenases by acquisition of DNA sequence repetitions. *J. Bacteriol.* 170, 4874–4880. doi: 10.1128/jb.170.10.4874-4880.1988

Nivens, D. E., Ohman, D. E., Williams, J., and Franklin, M. J. (2001). Role of alginate and its O acetylation in formation of *Pseudomonas aeruginosa* microcolonies and biofilms. *J. Bacteriol.* 183, 1047–1057. doi: 10.1128/JB.183.3.1047-1057.2001

Nordlund, I., Powlowski, J., and Shingler, V. (1990). Complete nucleotide sequence and polypeptide analysis of multicomponent phenol hydroxylase from *Pseudomonas* sp. strain CF600. *J. Bacteriol.* 172, 6826–6833. doi: 10.1128/jb.172.12.6826-6833.1990

Okonechnikov, K., Golosova, O., Fursov, M., and Team, U. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Olvera, C., Goldberg, J. B., Sánchez, R., and Soberón-Chávez, G. (1999). The *Pseudomonas aeruginosa algC* gene product participates in rhamnolipid biosynthesis. *FEMS Microbiol. Lett.* 179, 85–90. doi: 10.1111/j.1574-6968.1999.tb08712.x

O'Toole, G. A., and Kolter, R. (1998). Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol. Microbiol.* 30, 295–304. doi: 10.1046/j.1365-2958.1998.01062.x

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. G. J. (2013). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, D737–D743. doi: 10.1093/nar/gkt1252

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., et al. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* 36, 996–1004. doi: 10.1038/nbt.4229

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Parreira, A. G., Tótola, M. R., Jham, G. N., Da Silva, S. L., and Borges, A. C. (2011). Microbial biodegradation of aromatic compounds in a soil contaminated with gasohol. *Br. Biotechnol. J.* 1, 18–28. doi: 10.9734/BBJ/2011/295

Parte, A. C. (2014). LPSN--list of prokaryotic names with standing in nomenclature. *Nucleic Acids Res.* 42, D613–D616. doi: 10.1093/nar/gkt1111

Paulsen, I. T., Littlejohn, T. G., Rådström, P., Sundström, L., Sköld, O., Swedberg, G., et al. (1993). The 3′ conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. *Antimicrob. Agents Chemother.* 37, 761–768. doi: 10.1128/AAC.37.4.761

Pérez-Pantoja, D., De la Iglesia, R., Pieper, D. H., and González, B. (2008). Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol. Rev.* 32, 736–794. doi: 10.1111/j.1574-6976.2008.00122.x

Pérez-Pantoja, D., Donoso, R., Agulló, L., Córdova, M., Seeger, M., Pieper, D. H., et al. (2012). Genomic analysis of the potential for aromatic compounds biodegradation in *Burkholderiales*. *Environ. Microbiol.* 14, 1091–1117. doi: 10.1111/j.1462-2920.2011.02613.x

Phale, P. S., Shah, B. A., and Malhotra, H. (2019). Variability in assembly of degradation operons for naphthalene and its derivative, carbaryl, suggests mobilization through horizontal gene transfer. *Genes* 10:569. doi: 10.3390/genes10080569

Pichinoty, F., Mandel, M., Greenway, B., and Garcia, J. L. (1977). Etude de 14 bactéries dénitrifiantes appartenant au groupe *Pseudomonas stutzeri* isolées du sol par culture d'enrichissement en présence d'oxyde nitreux [Study of 14 denitrifying soil bacteria of the *"Pseudomonas stutzeri"* group isolated by enrichment culture in the presence of nitrous oxide (author's transl)]. Ann Microbiol (Paris). 128, 75–87. French.

Pier, G. B. (2007). *Pseudomonas aeruginosa* lipopolysaccharide: a major virulence factor, initiator of inflammation and target for effective immunity. *Int. J. Med. Microbiol.* 297, 277–295. doi: 10.1016/j.ijmm.2007.03.012

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Rossello, R., Garcia-Valdes, E., Lalucat, J., and Ursing, J. (1991). Genotypic and phenotypic diversity of *Pseudomonas stutzeri*. *Syst. Appl. Microbiol.* 14, 150–157. doi: 10.1016/S0723-2020(11)80294-8

Rosselló, R., Garcia-Valdes, E., Macario, A. J. L., Lalucat, J., and Conway De Macario, E. (1992). Antigenic diversity of *Pseudomonas stutzeri*. *Syst. Appl. Microbiol.* 15, 617–623. doi: 10.1016/S0723-2020(11)80124-4

Rosselló-Mora, R. A., Lalucat, J., Dott, W., and Kämpfer, P. (1994). Biochemical and chemotaxonomic characterization of *Pseudomonas stutzeri* genomovars. *J. Appl. Bacteriol.* 76, 226–233. doi: 10.1111/j.1365-2672.1994.tb01620.x

Sader, H. S., and Jones, R. N. (2005). Antimicrobial susceptibility of uncommonly isolated non-enteric Gram-negative bacilli. *Int. J. Antimicrob. Agents* 25, 95–109. doi: 10.1016/j.ijantimicag.2004.10.002

Salgar-Chaparro, S. J., Castillo-Villamizar, G., Poehlein, A., Daniel, R., Machuca, L. L., and Putonti, C. (2020). Complete genome sequence of *Pseudomonas balearica* strain EC28, an iron-oxidizing bacterium isolated from corroded steel. *Microbiol. Resour. Announc.* 9, e00275–e00220. doi: 10.1128/MRA.00275-20

Salvà-Serra, F., Jakobsson, H. E., Busquets, A., Gomila, M., Jaén-Luchoro, D., Seguí, C., et al. (2017). Genome sequences of two naphthalene-degrading strains of *Pseudomonas balearica*, isolated from polluted marine sediment and from an oil refinery site. *Genome Announc.* 5, e00116-17. doi: 10.1128/genomeA.00116-17

Salvà-Serra, F., Svensson-Stadler, L., Busquets, A., Jaén-Luchoro, D., Karlsson, R., Moore, R. B., et al. (2018). A protocol for extraction and purification of high-quality and quantity bacterial DNA applicable for genome sequencing: a modified version of the Marmur procedure. *Protocol Exch*. doi: 10.1038/protex.2018.084

Sampath, R., Venkatakrishnan, H., Ravichandran, V., and Chaudhury, R. R. (2012). Biochemistry of TBT-degrading marine pseudomonads isolated from Indian coastal waters. *Water Air Soil Pollut.* 223, 99–106. doi: 10.1007/s11270-011-0842-5

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Schoch, C. L., Sherry, S. T., et al. (2022). GenBank. *Nucleic Acids Res.* 50, D161–D164. doi: 10.1093/nar/gkab1135

Scotta, C., Mulet, M., Sánchez, D., Gomila, M., Ramírez, A., Bennasar, A., et al. (2012). Identification and genomovar assignation of clinical strains of *Pseudomonas stutzeri*. *Eur. J. Clin. Microbiol. Infect. Dis.* 31, 2133–2139. doi: 10.1007/s10096-012-1547-4

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., et al. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* 19, 823–826. doi: 10.1038/s41592-022-01539-7

Shi, Y.-H., Ren, L., Jia, Y., and Yan, Y.-C. (2015). Genome sequence of organophosphorus pesticide-degrading bacterium *Pseudomonas stutzeri* strain YC-YH1. *Genome Announc.* 3, e00192–e00115. doi: 10.1128/genomeA.00192-15

Shingler, V., Franklin, F. C. H., Tsuda, M., Holroyd, D., and Bagdasarian, M. (1989). Molecular analysis of a plasmid-encoded phenol hydroxylase from *Pseudomonas* CF600. *Microbiology* 135, 1083–1092. doi: 10.1099/00221287-135-5-1083

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539. doi: 10.1038/msb.2011.75

Sikorski, J., Lalucat, J., and Wackernagel, W. (2005). Genomovars 11 to 18 of *Pseudomonas stutzeri*, identified among isolates from soil and marine sediment. *Int. J. Syst. Evol. Microbiol.* 55, 1767–1770. doi: 10.1099/ijs.0.63535-0

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406, 959–964. doi: 10.1038/35023079

Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* 56, 564–577. doi: 10.1080/10635150701472164

Tanizawa, Y., Fujisawa, T., Kaminuma, E., Nakamura, Y., and Arita, M. (2016). DFAST and DAGA: web-based integrated genome annotation tools and resources. *Biosci. Microb. Food Health* 35, 173–184. doi: 10.12938/bmfh.16-003

Tanizawa, Y., Fujisawa, T., and Nakamura, Y. (2017). DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34, 1037–1039. doi: 10.1093/bioinformatics/btx713

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc. Natl. Acad. Sci.* 102, 13950–13955. doi: 10.1073/pnas.0506758102

Udaondo, Z., Molina, L., Segura, A., Duque, E., and Ramos, J. L. (2016). Analysis of the core genome and pangenome of *Pseudomonas putida*. *Environ. Microbiol.* 18, 3268–3283. doi: 10.1111/1462-2920.13015

Uddin, F., Sohail, M., Shaikh, Q. H., Hussain, M. T., Roulston, K., and McHugh, T. D. (2022). Verona integron-encoded metallo-Beta-lactamase (VIM) and Vietnam extended-spectrum Beta-lactamase (VEB) producing *Pseudomonas balearica* from a clinical specimen. *J. Pakis. Med. Assoc.* 72, 761–763. doi: 10.47391/JPMA.3890

Whelan, F. J., Hall, R. J., and McInerney, J. O. (2021). Evidence for selection in the abundant accessory gene content of a prokaryote pangenome. *Mol. Biol. Evol.* 38, 3697–3708. doi: 10.1093/molbev/msab139

Willenbrock, H., Hallin, P. F., Wassenaar, T. M., and Ussery, D. W. (2007). Characterization of probiotic *Escherichia coli* isolates with a novel pan-genome microarray. *Genome Biol.* 8:R267. doi: 10.1186/gb-2007-8-12-r267

Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K.-H., et al. (2008). The all-species living tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst. Appl. Microbiol.* 31, 241–250. doi: 10.1016/j.syapm.2008.07.001

Yoon, S. H., Ha, S. M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Youssef, E. M., Kim, H., Ahmad, I., Brauner, A., Liu, Y., Skurnik, M., et al. (2017). Stand-alone EAL domain proteins form a distinct subclass of EAL proteins involved in regulation of cell motility and biofilm formation in enterobacteria. *J. Bacteriol.* 199, e00179–e00117. doi: 10.1128/JB.00179-17

Zhang, Q., and Ye, Y. (2017). Not all predicted CRISPR-Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC Bioinform.* 18:92. doi: 10.1186/s12859-017-1512-4