Microbiology Resource Announcements

| Microbial Ecology | Announcement

Genome sequences of four novel *Endozoicomonas* strains associated with a tropical octocoral in a long-term aquarium facility

Matilde Marques,[1,2] Daniela M.G. da Silva,[1,2] Elsa Santos,[3] Núria Baylina,[3] Raquel Peixoto,[4] Nikos C. Kyrpides,[5] Tanja Woyke,[6] William B. Whitman,[6] Tina Keller-Costa,[1,2] Rodrigo Costa[1,2]

AUTHOR AFFILIATIONS See affiliation list on p. 4.

ABSTRACT We report the genome sequences of four *Endozoicomonas* sp. strains isolated from the octocoral *Litophyton* maintained long term at an aquarium facility. Our analysis reveals the coding potential for versatile polysaccharide metabolism; Type II, III, IV, and VI secretion systems; and the biosynthesis of novel ribosomally synthesized and post-translationally modified peptides.

KEYWORDS Chitinases, *Endozoicomonadaceae*, host-microbe interactions, coral holobiont, symbiosis, bacteria

The bacterial genus *Endozoicomonas* (*Pseudomonadota*, *Endozoicomonadaceae*) is a subject of increasing research interest owing to its widespread association with marine animals, particularly corals (1–4). However, *Endozoicomonas* spp. are typically difficult to cultivate and maintain in the laboratory (3, 4).

We report the genomes of four *Endozoicomonas* strains isolated from two *Litophyton* sp. specimens kept in a 19-m^3 tropical exhibition aquarium at the Oceanário de Lisboa, Portugal. Host-derived microbial cell suspensions were retrieved as described previously (2). One gram of coral tissue was homogenized in 9 mL of sterile Ca^{2+}- and Mg^{2+}-free artificial seawater (2). The homogenate was serially diluted, plated separately on marine agar diluted 1:2 and R2A diluted 1:10 media, and incubated at 21°C for 4 weeks. Genomic DNA of single colonies was extracted from cultures freshly grown in 1:2 marine broth using the Wizard Genomic DNA Purification kit (Promega, USA). Purity was confirmed by Sanger sequencing of 16S rRNA genes amplified from genomic DNA using universal primers (F27 and R1492). Taxonomy assignment was performed with the SILVA Alignment, Classification, and Tree Service (v1.2.12) and database (v138.1). The same genomic DNA samples were used for genome sequencing at the DOE Joint Genome Institute (JGI) using PacBio sequencing technology (5). For each sample, genomic DNA was sheared to 6–10 kb, treated using SMRTbell Express Template Prep Kit 3.0, and purified with SMRTbell cleanup beads (PacBio). The purified product was enriched using barcoded amplification oligos (IDT) and SMRTbell gDNA Sample Amplification Kit (PacBio). A 10-kb PacBio SMRTbell library was constructed and sequenced on the PacBio Revio system using HiFi chemistry. Raw reads were quality-filtered as per the JGI standard operating practice (SOP) protocol 1061 using BBTools v.38.86 (http://bbtools.jgi.doe.gov). Filtered reads >5 kb were assembled using Flye v2.8.3 (6). Organism and project metadata were deposited in the Genomes OnLine database (7). Contigs were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.6.7) (8) and the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4) (9) coupled to the Integrated Microbial Genomes and Microbiomes system v7 (IMG/M) for comparative analyses (10). Genome completeness and contamination were assessed with CheckM

Editor Elinne Becket, California State University San Marcos, San Marcos, California, USA

Address correspondence to Rodrigo Costa, rodrigoscosta@tecnico.ulisboa.pt, or Tina Keller-Costa, tinakellercosta@tecnico.ulisboa.pt.

Matilde Marques and Daniela M.G. da Silva contributed equally to this article. Author order was determined in order of increasing seniority.

The authors declare no conflict of interest.

See the funding table on p. 5.

Received 25 July 2024
Accepted 26 October 2024
Published 9 December 2024

TABLE 1 General sequencing statistics and genome features of the *Endozoicomonas* sp. reported in this study

Strain[a]	Genome size (Mb)	GC content (%)	Genome coverage (x)	Number of contigs	Contig N50 (Mb)	Number of reads[b]	Average read length (bp)[b]	Estimated Completeness (%)	Estimated Contamination (%)	Counts[c]								GenBank accession number	SRA accession number	Bioproject accession number	Biosample accession number
										Genes	CDs	RNA	rRNA	tRNA	ncRNA	COG[d]	Pfam[d]				
NE35	5.5	49.0	187.0	4	5	1,788,957| 3,242	10,318 ± 3,193.2| 9,971 ± 3,275.8	99.08	4.41	4,955*| 4,861†	4,828*| 4,667†	137†	25*| 25†	97*| 107†	5†	3,458*	4,933*	JBEWTA000000	SRR28058472 SRR28058473	PRJNA1075803	SAMN39945177
NE40	5.5	49.0	202.0	3	5.1	7,826,899| 10,098	9,477 ± 2,410.7| 9,290 ± 2,556.6	99.14	4.19	4,947*| 4,849†	4,820*| 4,657†	137†	25*| 25†	97*| 107†	5†	3,458*	4,933*	JBEWTB000000	SRR28058719 SRR28058720	PRJNA1075804	SAMN39945184
NE41	5.5	49.0	195.0	6	5	3,594,929| 5,045	10,617 ± 2,883.3| 10,003 ± 2,860.4	99.03	4.08	4,981*| 4,888†	4,856*| 4,699†	137†	25*| 25†	97*| 107†	5†	3,449*	4,929*	JBEWTC000000	SRR28058712 SRR28058713	PRJNA1075805	SAMN39945181
NE43	5.5	49.0	196.0	3	5.1	4,034,152| 6,017	10,290 ± 2,785.9| 9,884 ± 2,807.1	99.21	4.41	4,941*| 4,855†	4,814*| 4,667†	137†	25*| 25†	122*| 107†	5†	3,463*	4,939*	JBEWTD000000	SRR28058717 SRR28058718	PRJNA1075806	SAMN39945185

[a]All strains reported in this study have been isolated from the octocoral host *Litophyton* sp. Strains NE35, NE41, and NE43 were isolated from the same specimen on MA 1:2, whereas strain NE40 was isolated from a second specimen on R2A 1:10 medium.
[b]Values per run on two different SMRT cells. SRA accessions are provided per run.
[c]Annotation was performed using the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4) (*) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP v.6.7) (†).
[d]Annotation files are publicly accessible on Zenodo (https://doi.org/10.5281/zenodo.13863125).

Pfam ID	Description	NE35	NE40	NE41	NE43
pfam00728	Glycosyl hydrolase family 20 (beta-N-acetylhexosaminidase)	1	2	1	1
pfam01915	Glycosyl hydrolase family 3 (glycoside hydrolase)	5	5	5	5
pfam17167	Glycosyl hydrolase 36 superfamily (chitobiose phosphorylase)	2	2	2	2
pfam03644	Glycosyl hydrolase family 85 (GH18 chitinase-like)	1	1	1	2
pfam00703	Glycosyl hydrolases family 2 (beta-galactosidase)	4	4	4	4
pfam04616	Glycosyl hydrolases family 43 (arabinase)	2	2	2	2
pfam00182	Chitinase class I (GH19 chitinase)	1	1	1	1
pfam01832	Mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase	1	1	1	1
pfam01522	Polysaccharide deacetylase	4	4	4	4
pfam02302	PTS system, Lactose/Cellobiose specific IIB subunit	2	2	2	2
pfam09614	CRISPR-associated protein (Cas_Csy2)	1	1	1	1
pfam09615	CRISPR-associated protein (Cas_Csy3)	1	1	1	1
pfam09618	CRISPR-associated protein (Cas_Csy4)	1	1	1	1
pfam01527	Transposase	2	2	2	2
pfam13007	Transposase C of IS166 homeodomain	7	7	7	7
pfam01609	Transposase DDE domain	12	12	12	12
pfam05157	Type II secretion system (T2SS), protein E, N-terminal domain	2	2	2	2
pfam00482	Type II secretion system (T2SS), protein F	3	3	3	3
pfam00263	Bacterial type II and III secretion system protein	4	4	4	4
pfam00437	Type II/IV secretion system protein	6	6	6	6
pfam08988	Type III secretion system, cytoplasmic E component of needle	1	1	1	1
pfam18269	T3SS EscN ATPase C-terminal domain	2	2	2	2
pfam11104	Type IV pilus assembly protein PilM	1	1	1	1
pfam05638	Type VI secretion system effector, Hcp	4	4	4	4
pfam04717	Type VI secretion system/phage-baseplate injector OB domain	4	4	4	4
pfam00812	Ephrin	1	1	1	1
COG ID	**Description**				
COG0666	Ankyrin repeat	6	6	6	6
COG0457	Tetratricopeptide (TPR) repeat	11	11	11	11
COG0790	TPR repeat	2	2	2	2
COG2319	WD40 repeat	1	1	1	1
COG2356	Endonuclease I	1	1	1	1
COG0648	Endonuclease IV	2	2	2	2
COG0778	Nitroreductase	4	4	4	4
COG1566	Multidrug resistance efflux pump	1	1	1	1
SM-BGCs	**Description**				
RiPP-like	Ribosomally synthesised and post-translationally modified peptides	3	3	3	3
Arylpolyene	Aryl polyene	1	1	1	1
Terpene	Terpene cluster	1	1	1	1

Number of Pfam, COG and SM-BGCs entries:

0 | 1 | 2-3 | 4-5 | 6-9 | 10-12

FIG 1 Presence and abundance of select functional features of the *Endozoicomonas* sp. genomes described in this study. Values of each entry represent the numbers of coding sequences assigned to Pfam (top) and COG (middle) functions per genome (https://doi.org/10.5281/zenodo.13863125), and the number of SM-BGCs (bottom) coding for major compound classes identified with antiSMASH v.7.0 (https://doi.org/10.5281/zenodo.13683288).

(v1.2.3) (11). AntiSMASH v7.1 (12) was used to identify secondary metabolite biosynthetic gene clusters (SM-BGCs). Default parameters were used for all software, unless otherwise specified.

Sequencing statistics and genome features are shown in Table 1. Average nucleotide identities (ANIs), calculated with FastANI v0.1.3 on KBase (13, 14), among strains NE35, NE40, NE41, and NE43, were above 99.9% in all pairwise comparisons. All four strains shared approximately 89.3% ANI with their closest relative, as determined by phylogenomics, including all *Endozoicomonas*-type strains with a publicly available

genome: *Endozoicomonas gorgoniicola* PS125^T (GCA_025562715), also isolated from an octocoral (15).

All four genomes encode several glycoside hydrolases, featuring chitinase, polysaccharide deacetylase, N-acetylglucosaminidase, and beta-galactosidase-encoding genes, congruent with the emerging view of complex carbon metabolism among *Endozoicomonadaceae* spp. associated with marine invertebrates (16–18). Multiple protein domains underlying Type II, III, IV, and VI secretion systems were predicted to be encoded in all genomes. Additionally, three CRISPR–Cas antiviral defense systems, several eukaryotic-like repeat protein motifs, and the potential to synthesize putatively novel ribosomally synthesized and post-translationally modified peptides, among other natural products, were encoded (Fig. 1).

ACKNOWLEDGMENTS

This study was financed by the "Blue Bioeconomy Pact" (Project N°. C644915664-00000026), co-funded by Next Generation EU European Fund, under the incentive line "Agendas for Business Innovation" within Funding Scheme 5-Capitalization and Business Innovation of the Portuguese Recovery and Resilience Plan (RRP). Further support was provided by the Portuguese Foundation for Science and Technology (FCT) through the projects UIDB/04565/2020 and UIDP/04565/2020 of iBB and the project LA/P/0140/2020 of i4HB. Sequencing, assembly, and annotation of the four *Endozoicomonas* sp. genomes were performed at the Joint Genome Institute (JGI) as part of the Genomic Encyclopedia of Type Strains, Phase V (KMG-V): Genome sequencing to study the core and pangenomes of soil and plant-associated prokaryotes. The work (proposal: 10.46936/10.25585/60001079) conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. M.M. is the recipient of a PhD scholarship conceded by FCT through the MIT Portugal program (10.54499/SFRH/BD/151376/2021). D.M.G.D.S. is the recipient of a MSc grant conceded by the "Blue Bioeconomy Pact" project. T.K.-C. is the recipient of a Research Scientist contract conceded by FCT (CEECIND/00788/2017).

AUTHOR AFFILIATIONS

[1]Institute for Bioengineering and Biosciences and i4HB-Institute for Health and Bioeconomy, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[2]Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[3]Oceanário de Lisboa, Esplanada D. Carlos I, Lisbon, Portugal
[4]King Abdullah University of Science and Technology, Red Sea Research Center, Thuwal, Saudi Arabia
[5]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[6]Department of Microbiology, University of Georgia, Athens, Georgia, USA

AUTHOR ORCIDs

Matilde Marques http://orcid.org/0000-0001-9443-0893
Raquel Peixoto http://orcid.org/0000-0002-9536-3132
Nikos C. Kyrpides http://orcid.org/0000-0002-6131-0462
Tanja Woyke http://orcid.org/0000-0002-9485-5637
William B. Whitman http://orcid.org/0000-0003-1229-0423
Tina Keller-Costa http://orcid.org/0000-0003-3702-9192
Rodrigo Costa http://orcid.org/0000-0002-5932-4101

FUNDING

Funder	Grant(s)	Author(s)
Next Generation EU European Fund	C644915664-00000026	Rodrigo Costa
MEC | Fundação para a Ciência e a Tecnologia (FCT)	UIDB/04565/2020	Rodrigo Costa
MEC | Fundação para a Ciência e a Tecnologia (FCT)	UIDP/04565/2020	Rodrigo Costa
MEC | Fundação para a Ciência e a Tecnologia (FCT)	LA/P/0140/2020	Rodrigo Costa
U.S. Department of Energy (DOE)	10.46936/10.25585/60001079	William B. Whitman

AUTHOR CONTRIBUTIONS

Matilde Marques, Formal analysis, Investigation, Writing – original draft, Writing – review and editing | Daniela M.G. da Silva, Formal analysis, Investigation, Writing – review and editing | Elsa Santos, Data curation, Resources, Visualization, Writing – review and editing | Núria Baylina, Data curation, Resources, Writing – review and editing | Raquel Peixoto, Conceptualization, Supervision, Writing – review and editing | Nikos C. Kyrpides, Funding acquisition, Resources, Writing – review and editing | Tanja Woyke, Data curation, Funding acquisition, Resources, Validation, Writing – review and editing | William B. Whitman, Funding acquisition, Resources, Writing – review and editing | Tina Keller-Costa, Conceptualization, Investigation, Supervision, Writing – review and editing | Rodrigo Costa, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The genome sequences of the four *Endozoicomonas* sp. strains have been deposited in GenBank/NCBI. GenBank accession numbers are listed in Table 1. The assemblies of NE35, NE40, NE41, and NE43 are available under the BioProject accession numbers PRJNA1075803, PRJNA1075804, PRJNA1075805, and PRJNA1075806, respectively. The raw reads are available under accession numbers SRR28058472 and SRR28058473 for NE35, SRR28058719 and SRR28058720 for NE40, SRR28058712 and SRR28058713 for NE41, and under SRR28058717 and SRR28058718 for NE43. COG and Pfam annotation results on IMG/M v7 and AntiSMASH results are available under https://doi.org/10.5281/zenodo.13863125 and https://doi.org/10.5281/zenodo.13683288, respectively.

REFERENCES

1. Esteves AIS, Hardoim CCP, Xavier JR, Gonçalves JMS, Costa R. 2013. Molecular richness and biotechnological potential of bacteria cultured from irciniidae sponges in the north-east Atlantic. FEMS Microbiol Ecol 85:519–536. https://doi.org/10.1111/1574-6941.12140
2. Keller-Costa T, Eriksson D, Gonçalves JMS, Gomes NCM, Lago-Lestón A, Costa R. 2017. The gorgonian coral *Eunicella labiata* hosts a distinct prokaryotic consortium amenable to cultivation. FEMS Microbiol Ecol 93:1–14. https://doi.org/10.1093/femsec/fix143
3. Sweet M, Villela H, Keller-Costa T, Costa R, Romano S, Bourne DG, Cárdenas A, Huggett MJ, Kerwin AH, Kuek F, Medina M, Meyer JL, Müller M, Pollock FJ, Rappé MS, Sere M, Sharp KH, Voolstra CR, Zaccardi N, Ziegler M, Peixoto R. 2021. Insights into the cultured bacterial fraction of corals. mSystems 6:e0124920. https://doi.org/10.1128/mSystems.01249-20
4. Pogoreutz C, Ziegler M. 2024. Frenemies on the reef? Resolving the coral-*Endozoicomonas* association. Trends Microbiol 32:422–434. https://doi.org/10.1016/j.tim.2023.11.006
5. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, et al. 2009. Real-time DNA sequencing from single polymerase molecules. Science 323:133–138. https://doi.org/10.1126/science.1162986
6. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474
7. Mukherjee S, Stamatis D, Li CT, Ovchinnikova G, Bertsch J, Sundaramurthi JC, Kandimalla M, Nicolopoulos PA, Favognano A, Chen IMA, Kyrpides NC, Reddy TBK. 2023. Twenty-five years of genomes online database (GOLD): data updates and new features in v.9. Nucleic Acids Res 51:D957–D963. https://doi.org/10.1093/nar/gkac974
8. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569
9. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). Stand Genomic Sci 10:86. https://doi.org/10.1186/s40793-015-0077-y

10. Chen IMA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter SJ, Webb C, Wu D, Varghese NJ, Reddy TBK, Mukherjee S, Ovchinnikova G, Nolan M, Seshadri R, Roux S, Visel A, Woyke T, Eloe-Fadrosh EA, Kyrpides NC, Ivanova NN. 2023. The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res 51:D723–D732. https://doi.org/10.1093/nar/gkac976
11. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114
12. Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, van Wezel GP, Medema MH, Weber T. 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 51:W46–W50. https://doi.org/10.1093/nar/gkad344
13. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9
14. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: The United States department of energy systems biology knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163
15. Pike RE, Haltli B, Kerr RG. 2013. Description of *Endozoicomonas euniceicola* sp. nov. and *Endozoicomonas gorgoniicola* sp. nov., bacteria isolated from the octocorals *Eunicea fusca* and *Plexaura* sp., and an emended description of the genus *Endozoicomonas*. Int J Syst Evol Microbiol 63:4294–4302. https://doi.org/10.1099/ijs.0.051490-0
16. Keller-Costa T, Kozma L, Silva SG, Toscan R, Gonçalves J, Lago-Lestón A, Kyrpides NC, Nunes da Rocha U, Costa R. 2022. Metagenomics-resolved genomics provides novel insights into chitin turnover, metabolic specialization, and niche partitioning in the octocoral microbiome. Microbiome 10:151. https://doi.org/10.1186/s40168-022-01343-7
17. da Silva DMG, Pedrosa FR, Ângela Taipa M, Costa R, Keller-Costa T. 2023. Widespread occurrence of chitinase-encoding genes suggests the *Endozoicomonadaceae* family as a key player in chitin processing in the marine benthos. ISME Commun 3:109. https://doi.org/10.1038/s43705-023-00316-7
18. Jensen S, Frank JA, Arntzen MØ, Duperron S, Vaaje-Kolstad G, Hovland M. 2021. *Endozoicomonadaceae* symbiont in gills of *Acesta* clam encodes genes for essential nutrients and polysaccharide degradation. FEMS Microbiol Ecol 97:70. https://doi.org/10.1093/femsec/fiab070