Supplementary Material

Genomic and physiological characterization of *Novosphingobium terrae* sp. nov., an alphaproteobacterium isolated from Cerrado soil containing a megasized chromid

Aline Belmok, Felipe Marques de Almeida, Rodrigo Theodoro Rocha, Carla Simone Vizzotto, Marcos Rogério Tótola, Marcelo Henrique Soller Ramada, Ricardo Henrique Krüger, Cynthia Maria Kyaw, Georgios J. Pappas Jr

Supplementary Table S1. Number of predicted proteins in plasmids pGeG2a and pGeG2b of strain GeG2^T classified in each COG functional category.

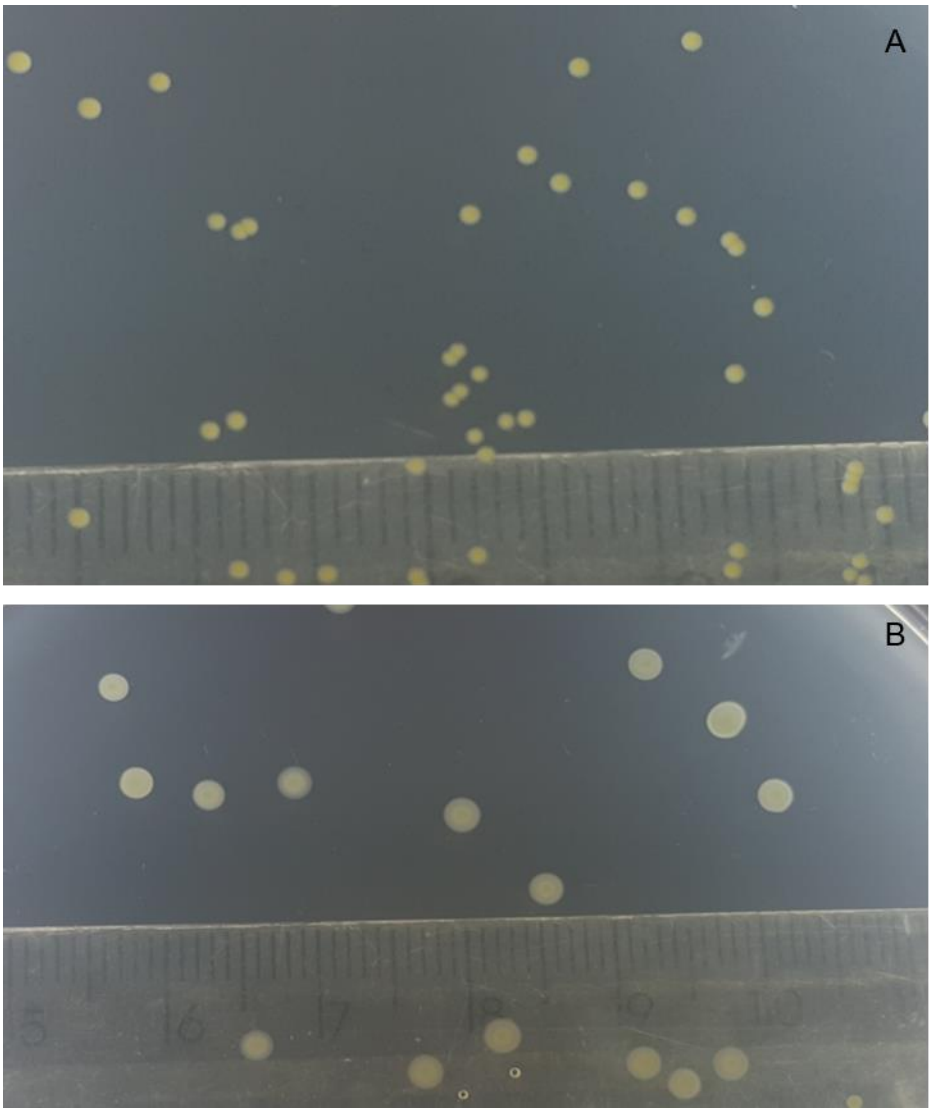
COG Category	Description	pGeG2a		pGeG2b	
		Total[a]	%[b]	Total[a]	%[b]
D	Cell cycle control, cell division, chromosome part.	1	0,70	2	4,55
M	Cell wall/membrane/envelope biogenesis	5	3,50	2	4,55
N	Cell motility	1	0,70	6	13,64
O	Post-translational modification, protein turnover	6	4,20	1	2,27
T	Signal transduction mechanisms	7	4,90	1	2,27
U	Trafficking, secretion, and vesicular transport	8	5,59	17	38,64
V	Defense mechanisms	1	0,70	0	0,00
K	Transcription	24	16,78	5	11,36
L	Replication, recombination, and repair	13	9,09	6	13,64
C	Energy production and conversion	7	4,90	0	0,00
E	Amino acid transport and metabolism	15	10,49	4	9,09
F	Nucleotide transport and metabolism	2	1,40	0	0,00
G	Carbohydrate transport and metabolism	16	11,19	2	4,55
H	Coenzyme transport and metabolism	3	2,10	0	0,00
I	Lipid transport and metabolism	8	5,59	1	2,27
P	Inorganic ion transport and metabolism	7	4,90	3	6,82
Q	Second. metabolites biosynthesis/transport/catabolism	8	5,59	0	0,00
S	Function unknown	37	25,87	6	13,64

[a] Predicted proteins attributed to more than one category were accounted in each category.

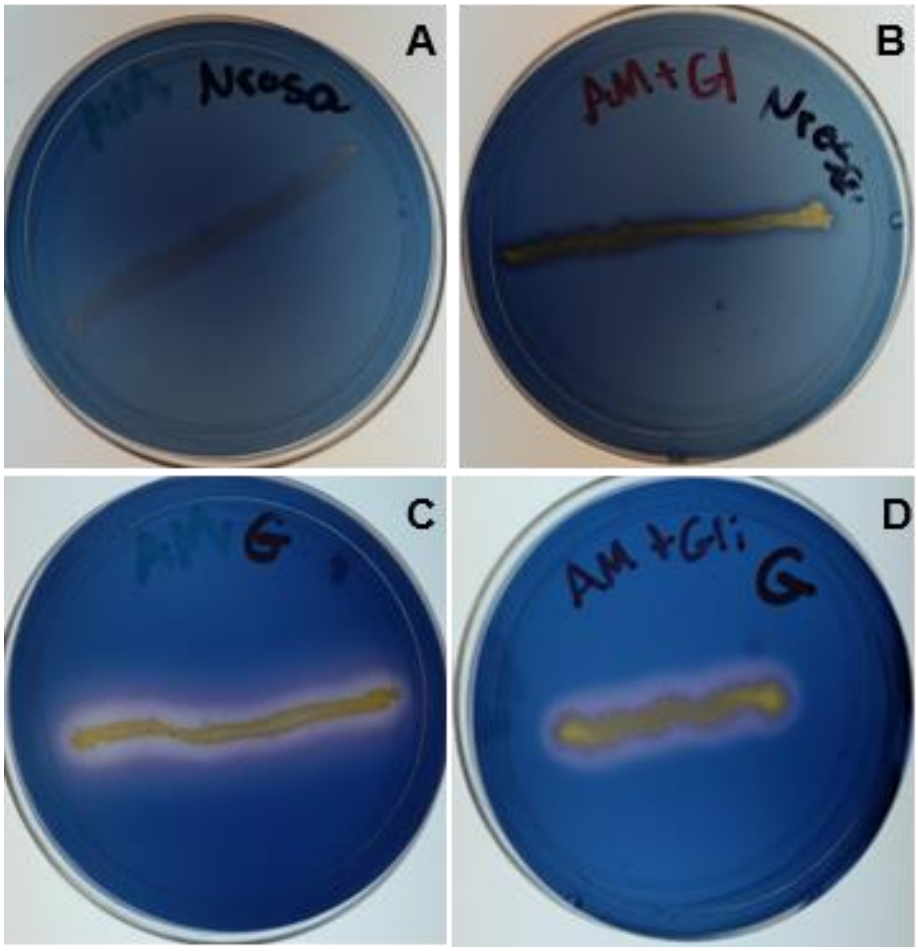
Supplementary Table S2. Mono and dioxygenases found in the genome of strain GeG2^T. Genes for enzymes also detected in genomes of *Novosphingobium* species able to degrade different aromatic compounds (1, 2) are indicated with *

Monooxygenases			
KO	**Name**	**Chromosome**	**Chromid**
K00459*	NMO; nitronate monooxygenase [EC 1.13.12.16]	2	0
K00481*	pobA; 4-hydroxybenzoate 3-monooxygenase [EC 1.14.13.2]	1	0
K03380	phenol 2-monooxygenase (NADPH) [EC 1.14.13.7]	0	1
K03862*	vanA; vanillate monooxygenase [EC:1.14.13.82]	0	8
K14974	nicC; 6-hydroxynicotinate 3-monooxygenase [EC:1.14.13.114]	0	1
K22027	iacA; indole-3-acetate monooxygenase [EC:1.14.13.235]	0	1
K04091*	ssuD; alkanesulfonate monooxygenase [EC:1.14.14.5]	0	2
K16047	hsaA; 3-hydroxy-9,10-secoandrosta-1,3,5(10)-triene-9,17-dione monooxygenase [EC:1.14.14.12]	0	4
K00500	phhA; phenylalanine 4-monooxygenase [EC:1.14.16.1]	0	1
K22553	4-methoxybenzoate monooxygenase (O-demethylating) [EC:1.14.99.15]	0	2
K20942	graA; resorcinol 4-hydroxylase (FADH2) [EC:1.14.14.27]	0	1
K05712	mhpA; 3-(3-hydroxy-phenyl)propionate hydroxylase [EC:1.14.13.127]	1	1
K18242	nagG; salicylate 5-hydroxylase large subunit [EC:1.14.13.172]	0	1
K18243	nagH; salicylate 5-hydroxylase small subunit [EC:1.14.13.172]	0	1
K03185	ubiH; 2-octaprenyl-6-methoxyphenol hydroxylase [EC:1.14.13.-]	1	0
K23464	tgnB; flavin-dependent trigonelline monooxygenase [EC:1.14.14.-]	0	1
K06134*	COQ7; 3-demethoxyubiquinol 3-hydroxylase [EC:1.14.99.60]	1	0
	Total	6	25

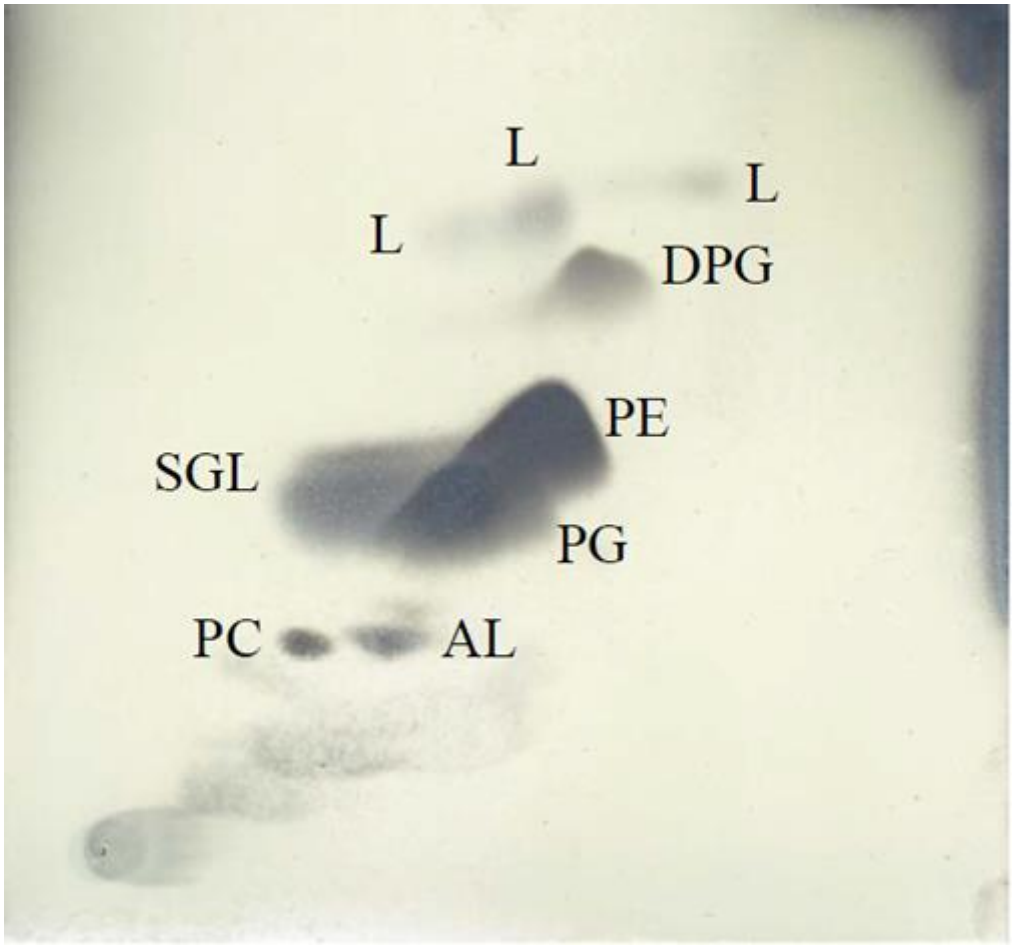
Dioxygenases			
KO	**Name**	**Chromosome**	**Chromid**
K05549*	benA-xylX; benzoate/toluate 1,2-dioxygenase subunit alpha [EC:1.14.12.10 1.14.12.-]	1	1
K05550*	benB-xilY; benzoate/toluate 1,2-dioxygenase subunit beta [EC:1.14.12.10 1.14.12.-]	1	0
K03381*	catA; catechol 1,2-dioxygenase [EC:1.13.11.1]	1	0
K07104	catE; catechol 2,3-dioxygenase [EC:1.13.11.2]	0	1
K00450*	gtdA; gentisate 1,2-dioxygenase [EC:1.13.11.4]	0	2
K00451*	hmgA; homogentisate 1,2-dioxygenase [EC:1.13.11.5]	0	2
K04100	ligA; protocatechuate 4,5-dioxygenase, alpha chain [EC:1.13.11.8]	2	1
K04101	ligB; protocatechuate 4,5-dioxygenase, beta chain [EC:1.13.11.8]	2	1
K06911	PIR; quercetin 2,3-dioxygenase [EC:1.13.11.24]	3	2
K00457*	hppD; 4-hydroxyphenylpyruvate dioxygenase [EC:1.13.11.27]	1	2
K08967*	mtnD/Z; acireductone dioxygenase (Ni2+-requiring) [EC:1.13.11.53]	1	0
K08967*	mtnD/Z ;acireductone dioxygenase [iron(II)-requiring][EC:1.13.11.54]	1	0
K21822	8'-apo-carotenoid 13,14-cleaving dioxygenase [EC:1.13.11.82]	1	0
K00472	P4HA; procollagen-proline 4-dioxygenase	1	0
K18068	pht3; phthalate 4,5-dioxygenase [EC:1.14.12.7]	0	1
K05916*	hmp; nitric oxide dioxygenase [EC:1.14.12.17]	1	0
K15777*	DOPA; 4,5-DOPA dioxygenase extradiol [EC:1.13.11.-]	1	0
	Total	17	13



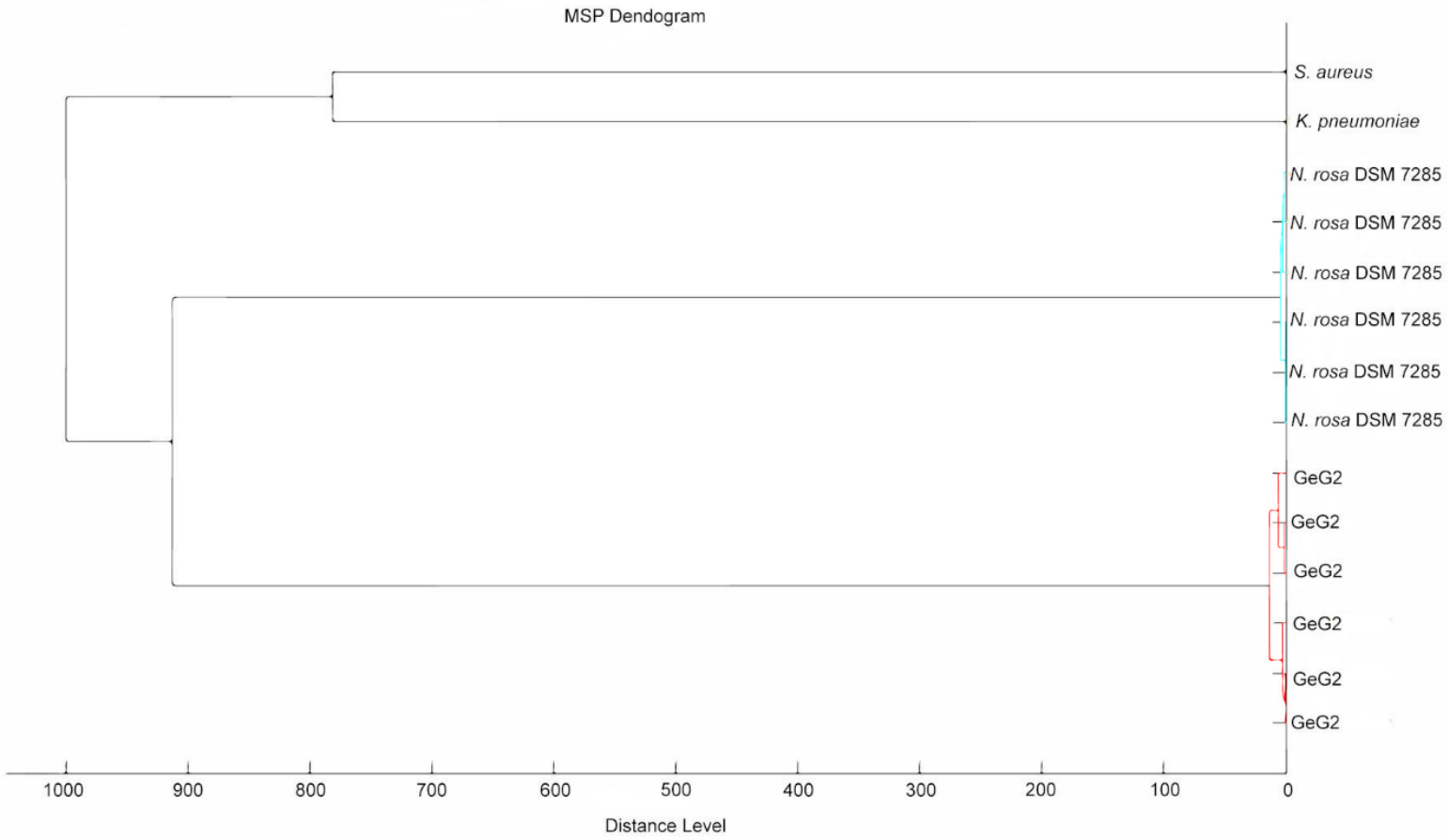
Supplementary Figure S1. Morphological differences observed in colonies of *Novosphingobium rosa* DSM7285^{T} (**A**) and strain GeG2^{T} (**B**) grown in MM for 48 hours.



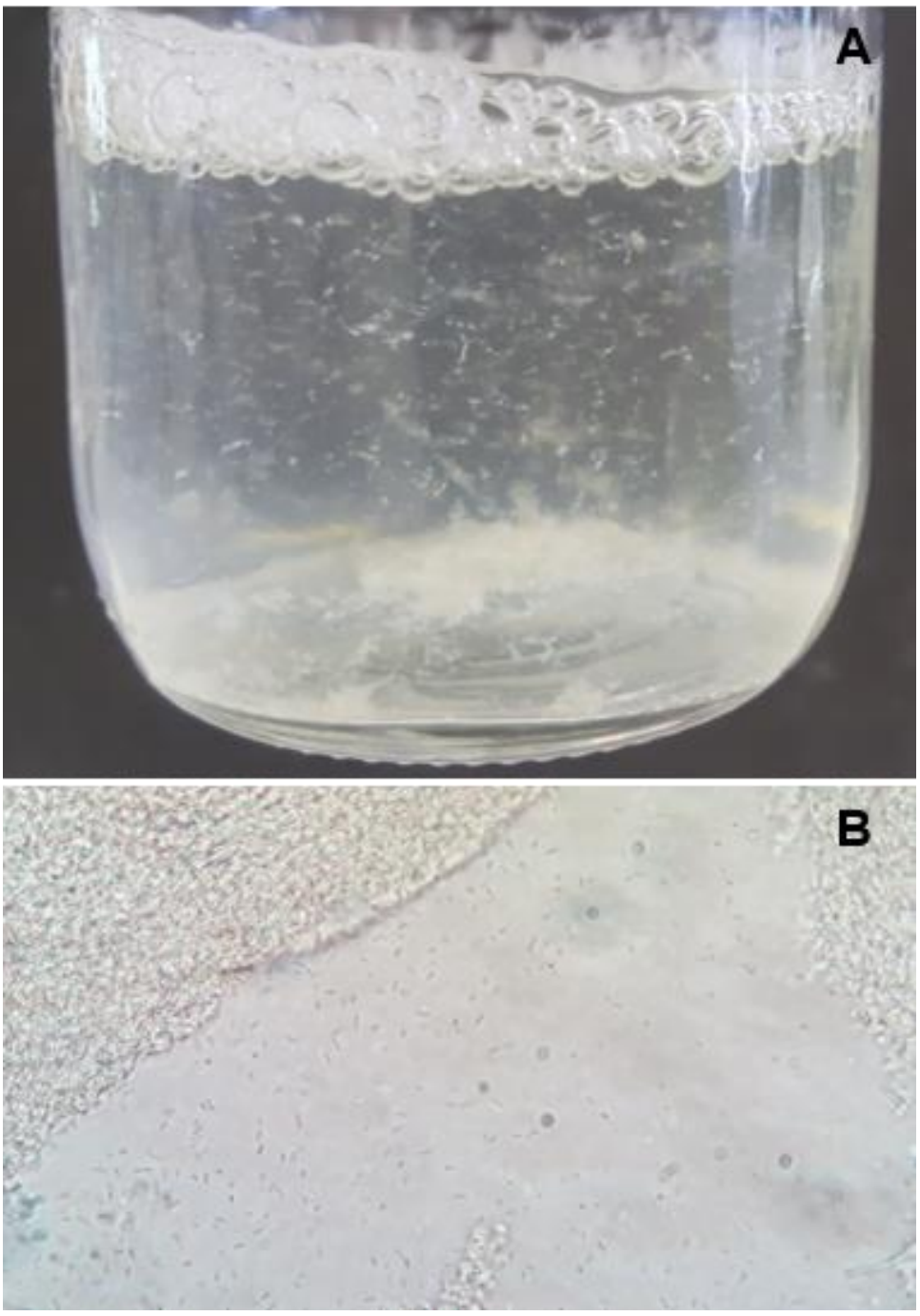
Supplementary Figure S2. Starch hydrolysis test results for strains *N. rosa* DSM7285^{T} (**A** and **B**) and GeG2^{T} (**C** and **D**) grown for 24 days in MM plates containing both starch and glucose (**B** and **D**) or starch as sole carbon source (**A** and **C**).



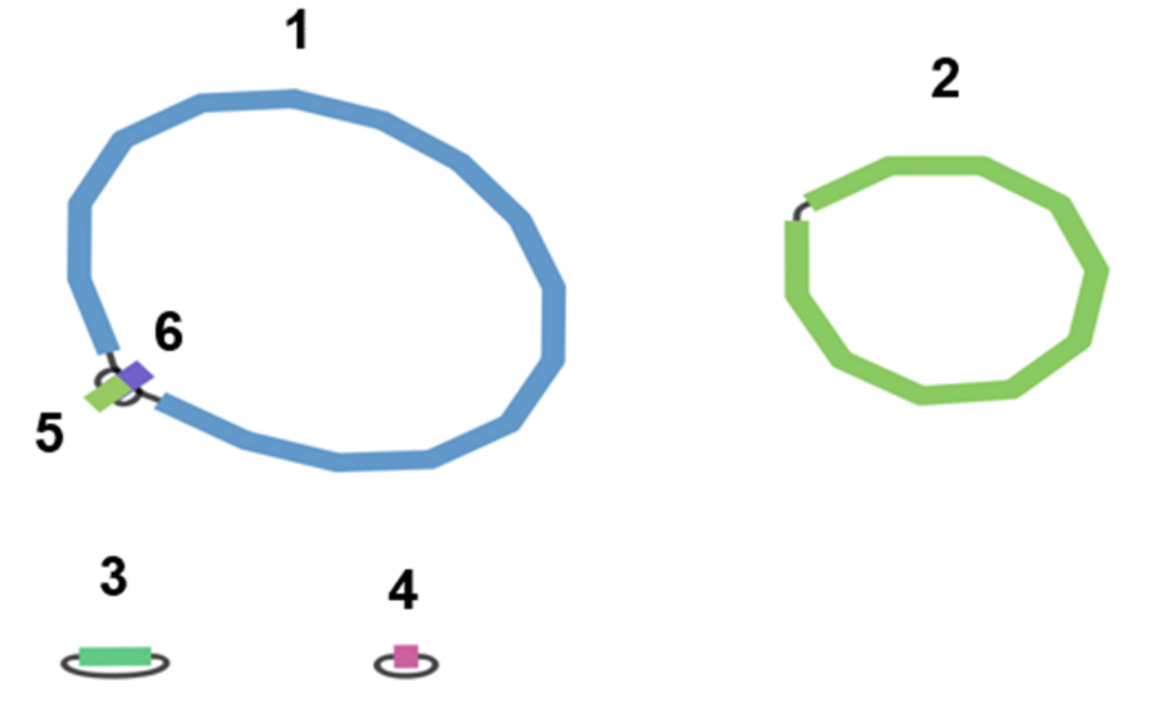
Supplementary Figure S3. Total polar lipid profile of strain GeG2^{T} after two-dimensional thin layer chromatography and staining with molybdatophosphoric acid; PG, Phosphatidylglycerol; PE, Phosphatidylethanolamine; SGL, Sphingoglycolipid; DPG, Diphosphatidylglycerol; PC, Phosphatidylcholine; AL, unidentified aminolipid; L, unidentified lipids.



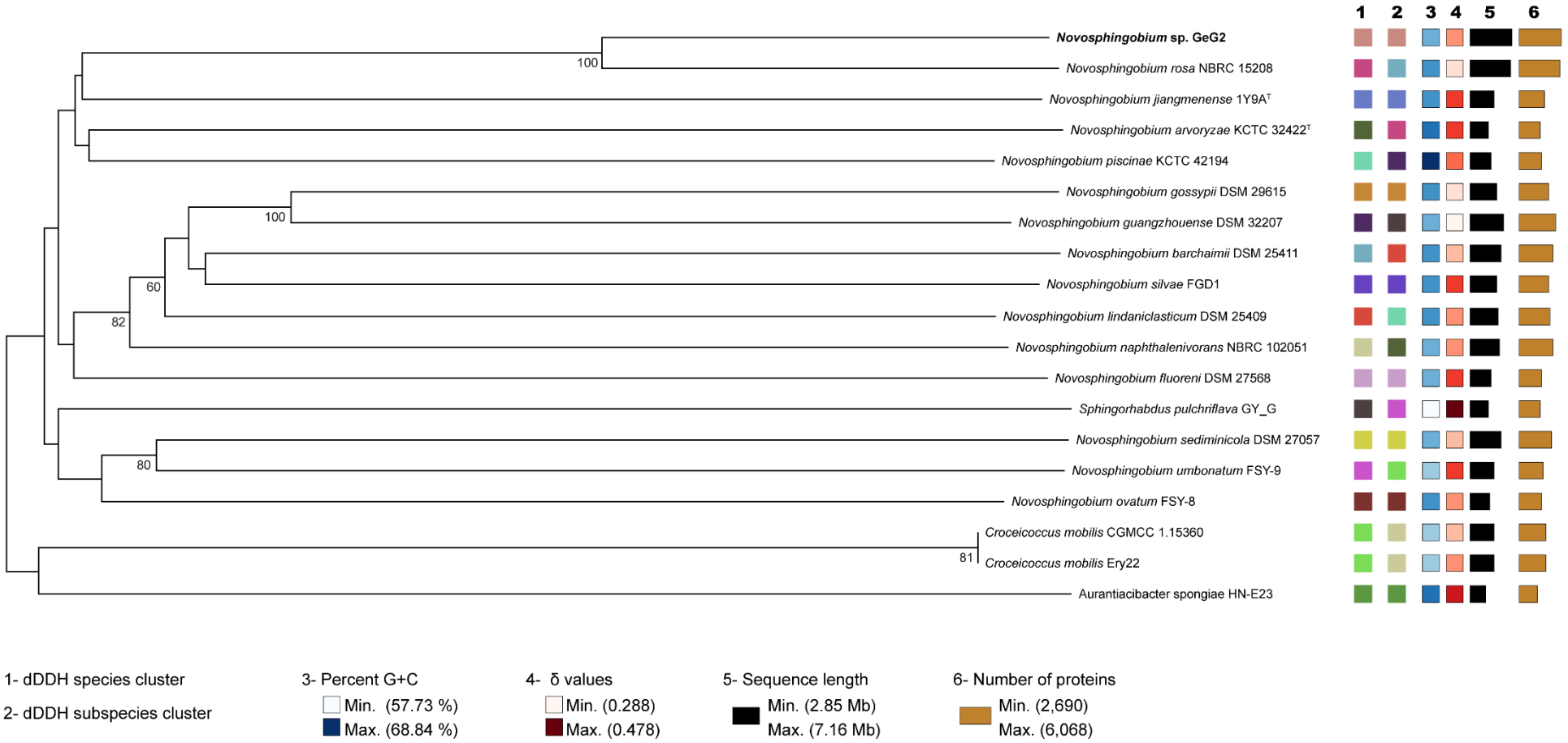
Supplementary Figure S4. Main spectra profiles (MSP) dendrogram obtained from strain GeG2^{T} and DSM 7285^{T} (*Novosphingobium rosa*)**.** Each MSP is composed of six independent mass spectra obtained from 2,000-20,000 m/z. The dendrogram was constructed using MSP comparisons of six biological replicates. *Staphylococcus aureus* and *Klebsiella pneumoniae* MSP retrieved from the MALDI Biotyper® (Bruker) database were used as outgroups. GeG2 strain replicates are grouped in a different clade than DSM 7285^{T}, indicating different species. This was also confirmed by a score identification lower than 2.0 when comparing MSPs in MALDI Biotyper 3.0 software.



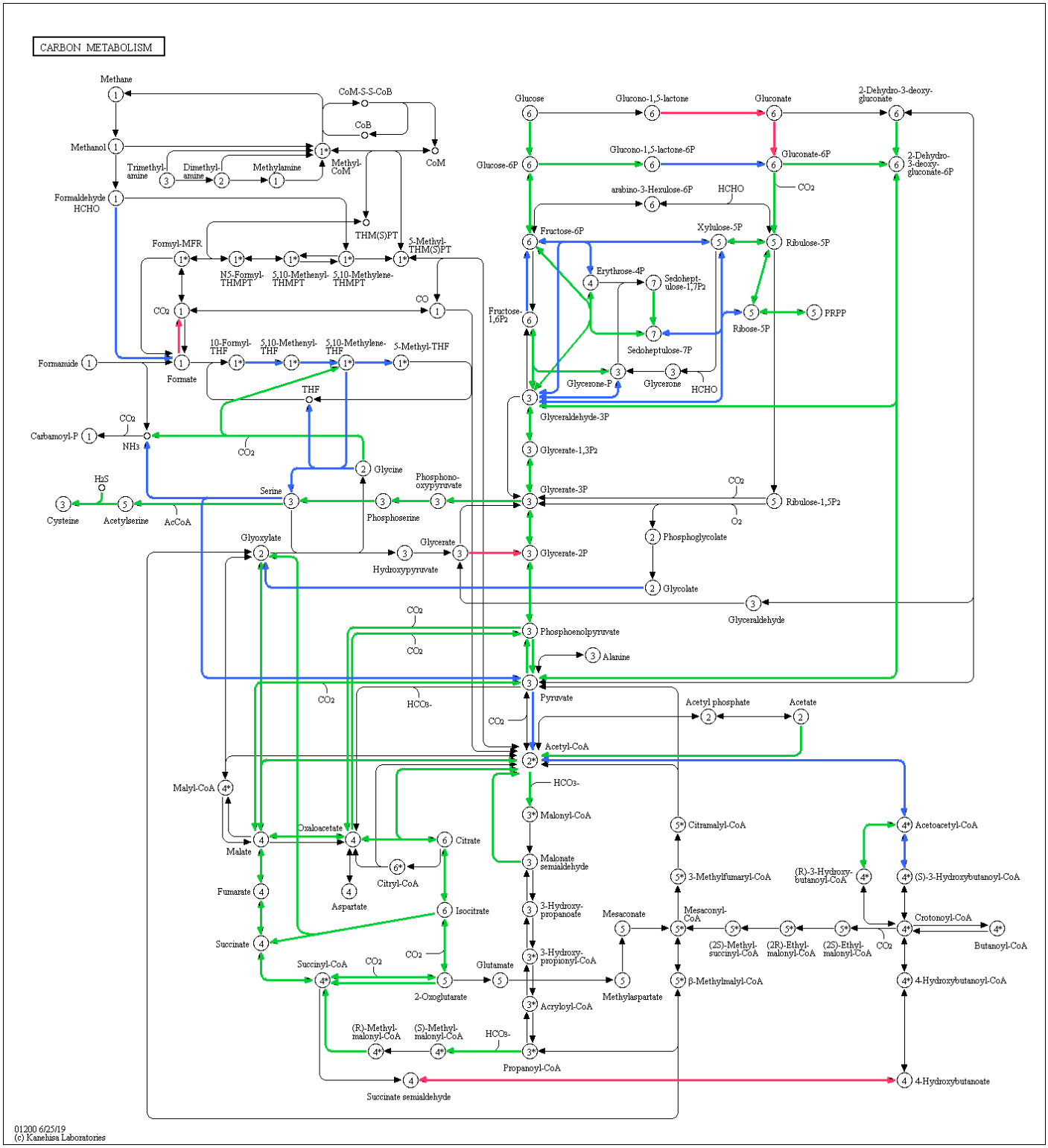
Supplementary Figure S5. Macroscopic flocks observed in cultures of strain GeG2 grown in liquid MM for 5 days at 28ºC under agitation conditions (**A**) and phase contrast microscopy picture showing the border of a flock evidencing non-motile cell aggregates embedded in extracellular matrix and free motile cells (**B**). 1000x magnification.



Supplementary Figure S6. Strain GeG2 whole genome assembly graph from Unicycler hybrid assembly strategy, using both Illumina short reads and PacBio long reads. 1: Chromosome; 2: Chromid; 3: pGeG2a plasmid; 4: pGeG2b plasmid; 5: 5,558 bp contig containing the entire rRNA operon; 6: 359 bp contig containing a single tRNA gene ($tRNA^{MET(CAT)}$). Graph was drawn with Bandage tool (3).



Supplementary Figure S7. Whole genome-sequence tree generated by TYGS server (in October 2021) including genomes of strain GeG2^T (in bold) and type strains available in server´s database. The tree was inferred with FastME 2.1.6.1 (4) from GBDP (Genome BLAST Distance Phylogeny) distance calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5. The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications, with an average branch support of 50.8%. The tree was rooted at the midpoint (5).



Supplementary Figure S8. Pathways associated with the carbon metabolism, obtained with KEGG Mapper tool (6), indicating genes detected in the chromosome (in green), the chromid (in pink) or in both replicons (in blue) of strain GeG2 genome.

References

1. Aylward, F.O., McDonald, B.R., Adams, S.M., Valenzuela, A., Schmidt, R.A., Goodwin, L.A., et al. (2013). Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Appl. Environ. Microbiol*.,79(12):3724-3733. doi: 10.1128/AEM.00518-13
2. Wang, J., Wang, C., Li, J., Bai, P., Li, Q., Shen, M. et al. (2018). Comparative Genomics of Degradative *Novosphingobium* Strains With Special Reference to Microcystin-Degrading *Novosphingobium* sp. THN1. *Front. Microbiol*., 9:2238. doi: 10.3389/fmicb.2018.02238
3. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics. 2015 Oct 15;31(20):3350-2. doi: 10.1093/bioinformatics/btv383.
4. Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol. 2015;32: 2798–2800. DOI: 10.1093/molbev/msv150
5. Farris JS. Estimating phylogenetic trees from distance matrices. Am Nat. 1972;106: 645–667.
6. Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M. et al. (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Res., 36:D480-D484. doi: 10.1093/nar/gkm882