Biochemical, toxicological and genomic investigation of a high-biomass producing *Limnothrix* strain isolated from Italian shallow drinking water reservoir

Mara Simonazzi[a*], Tânia K. Shishido[b,c], Endrews Delbaje[c,d], Matti Wahlsten[c], David P. Fewer[c], Kaarina Sivonen[c], Laura Pezzolesi[a], Rossella Pistocchi[a]

[a] Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Via S'Alberto 163, 48123 Ravenna, Italy

[b] Institute of Biotechnology, Helsinki Institute of Life Science, Viikinkaari 5, University of Helsinki, Helsinki, Finland

[c] Department of Microbiology, Faculty of Agriculture and Forestry, Viikinkaari 9, University of Helsinki, Helsinki, Finland

[d] Center for Nuclear Energy in Agriculture (CENA), University of São Paulo (USP), Brazil

* Corresponding author: M.S.

E-mail address: mara.simonazzi2@unibo.it

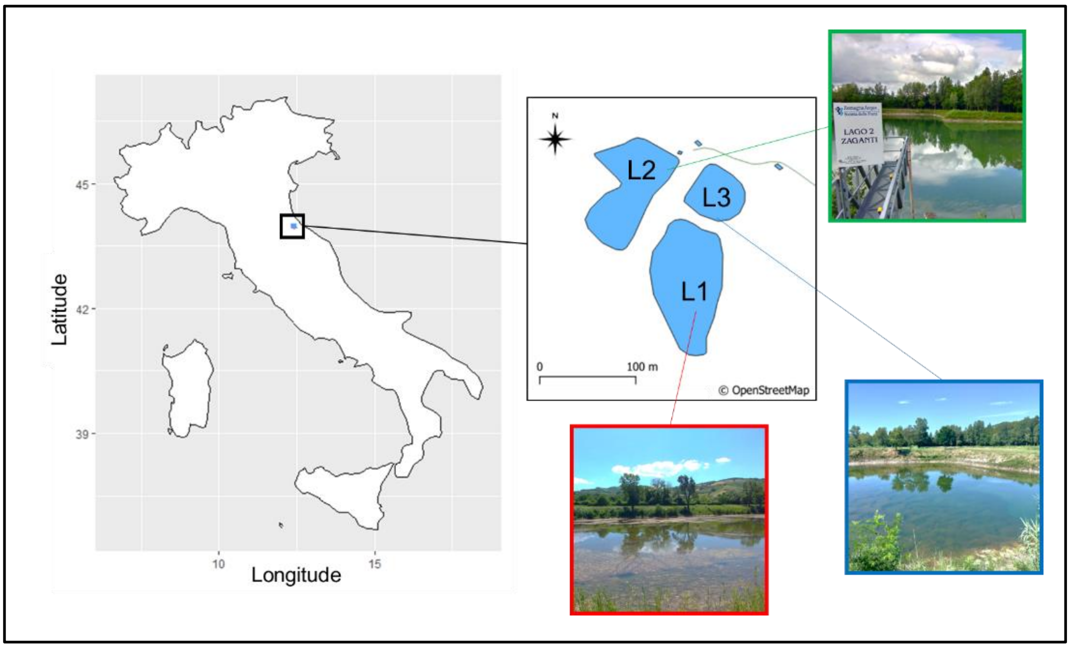

Fig. S1 Sampling sites in Lakes Zaganti, Italy (i.e., L1, L2, and L3) and their location

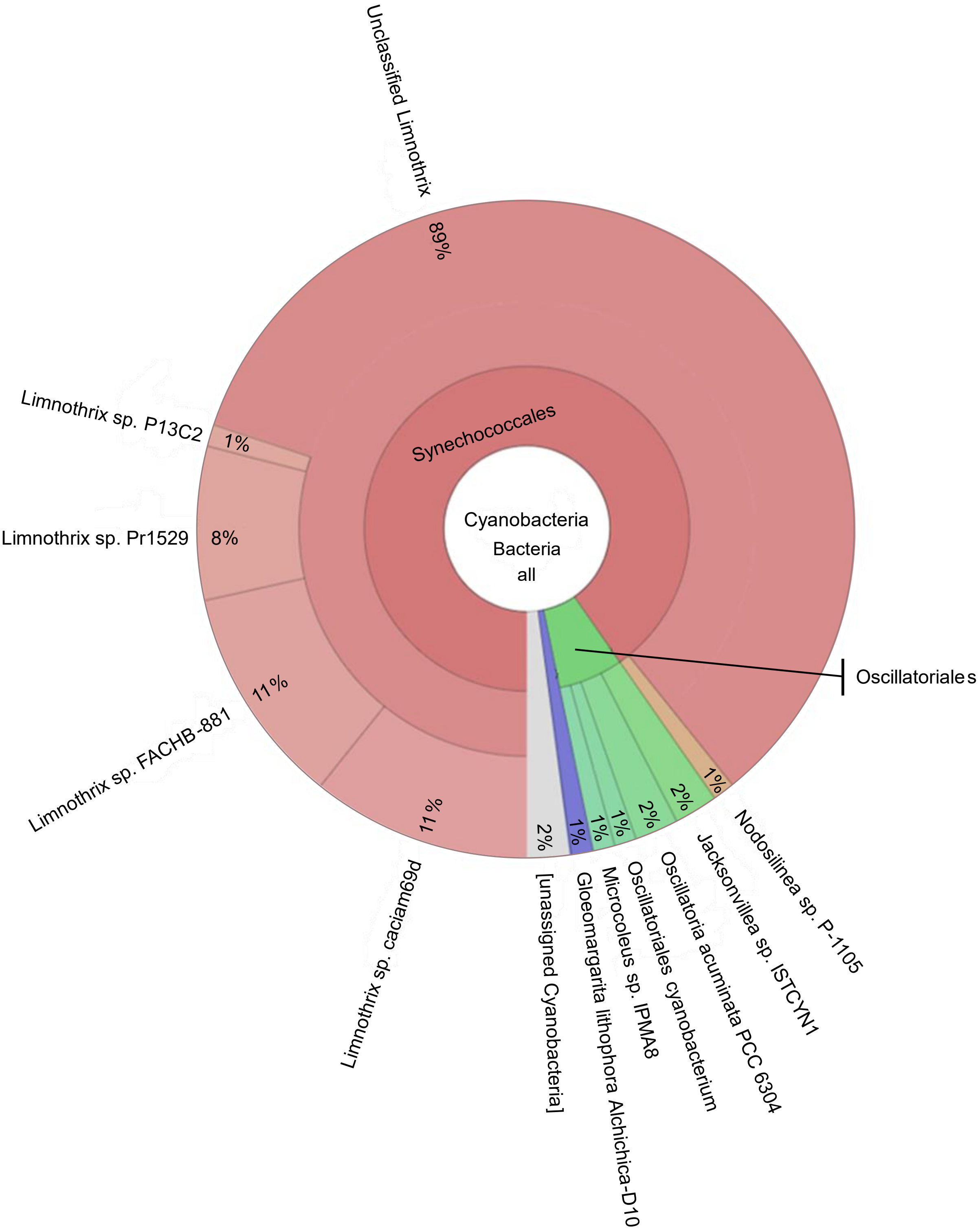

Fig. S2 Genome-based taxonomic assignation of the cyanobacterial isolate, with reads assigned in percentage. The krona chart was generated by Kaiju using the nr Bacteria, Archaea and Viruses database

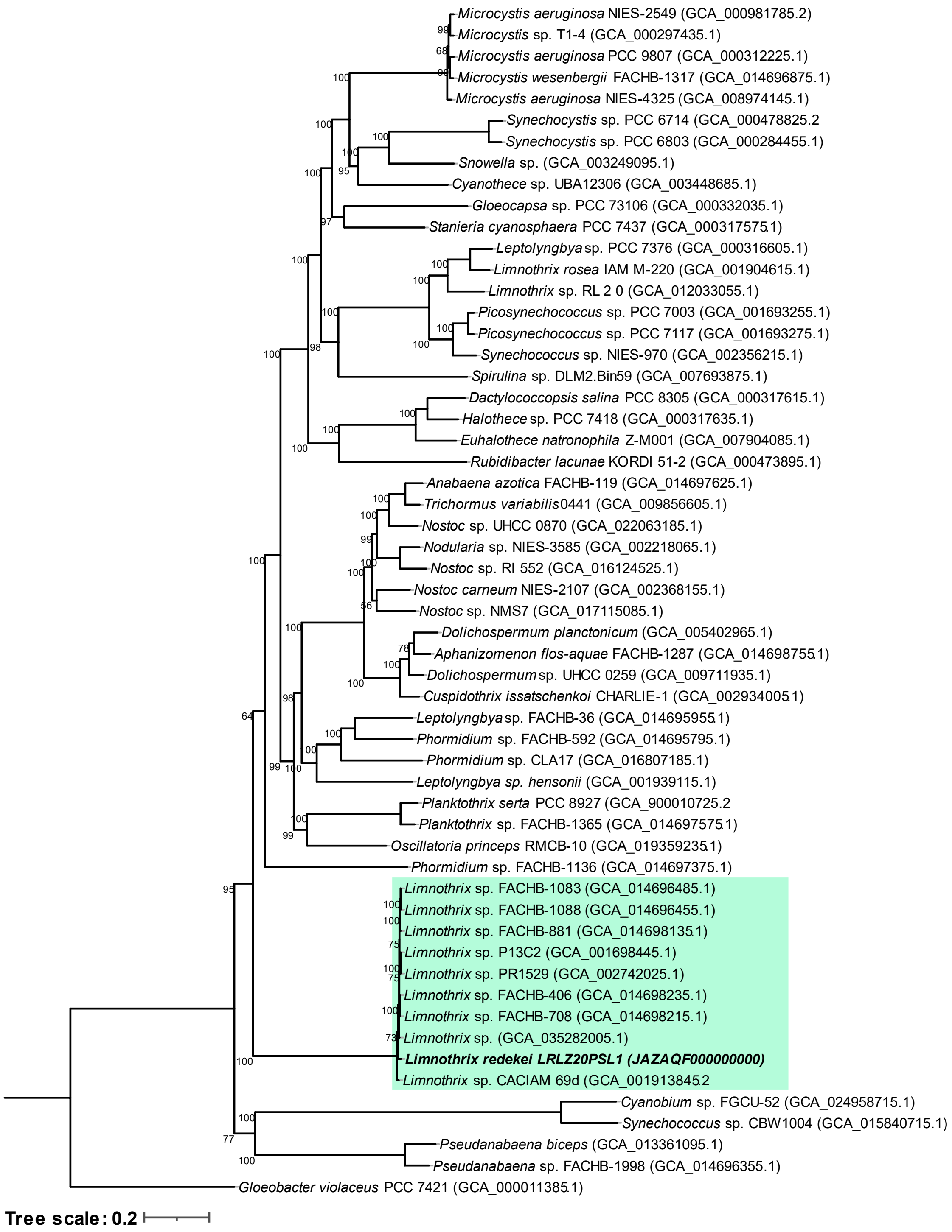

Fig. S3 Phylogenomic tree constructed based on genomic data of LRLZ20PSL1, utilizing 164 concatenated single-copy proteins common to cyanobacterial genomic sequences.

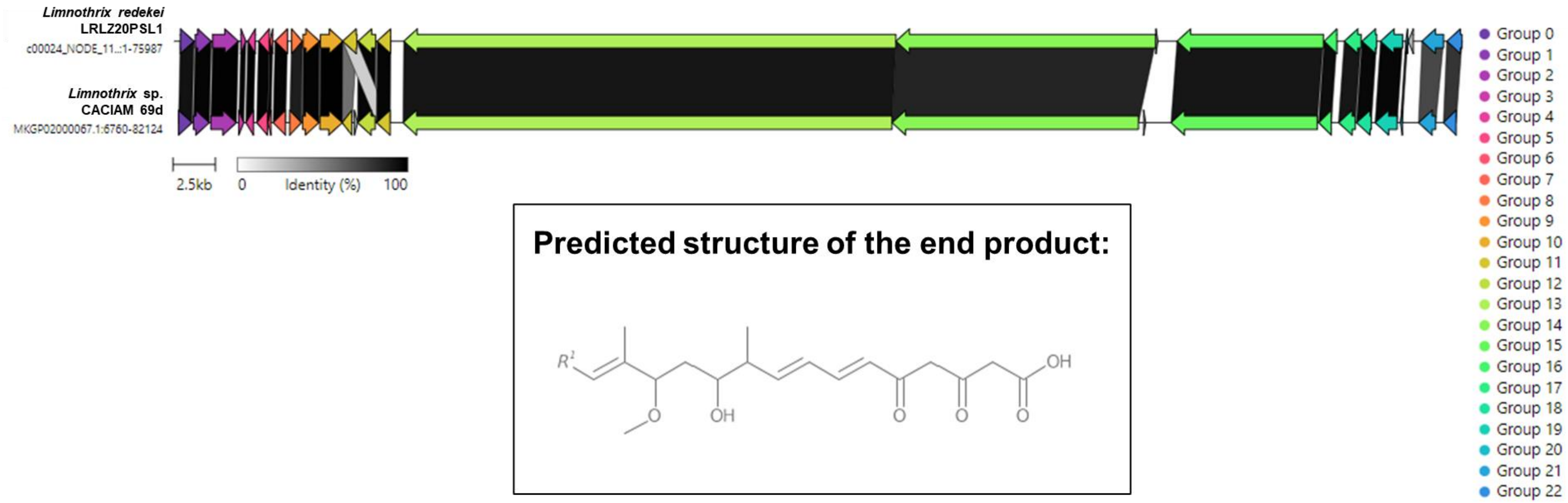

Fig. S4 Comparison of the *trans*-AT PKS region found in the studied genome of LRLZ20PSL1 (present work) and from *Limnothrix* sp. CACIAM 69d generated using Clinker [1]. A tentative structure of the end product was predicted by transATor [2] and it is shown in the black rectangle

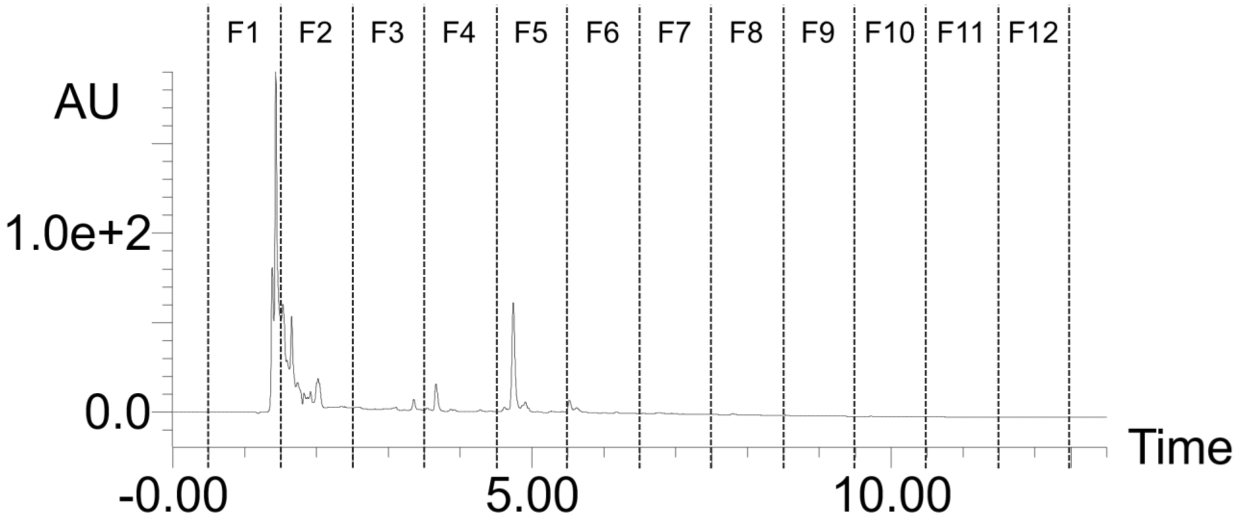

Fig. S5 UV spectra of the 12 fractions (F1-12) obtained from LRLZ20PSL1 extracts following the separation with HILIC chromatography

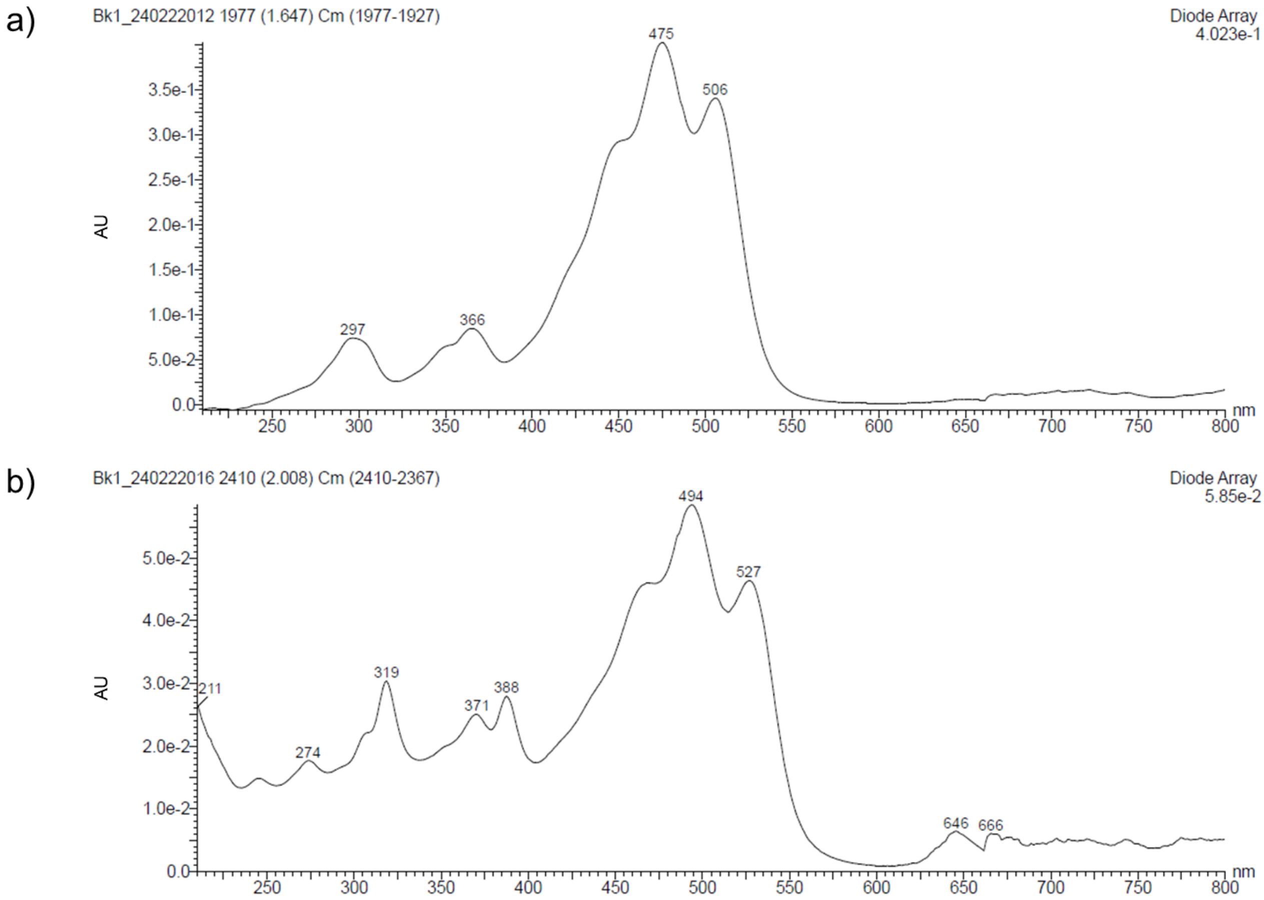

Fig. S6 Absorbance peaks at a) 1.65 and b) 2.01 from fraction F2 derived from HILIC fractionation

Table S1

Fraction Number	Start Time (min)	End Time (min)
F1	0.502	1.502
F2	1.502	2.502
F3	2.502	3.502
F4	3.502	4.502
F5	4.502	5.502
F6	5.502	6.502
F7	6.503	7.503
F8	7.503	8.503
F9	8.505	9.505
F10	9.505	10.505
F11	10.505	11.505
F12	11.505	12.501

Start and end times (min) of each collected fraction (F1-F12) of LRLZ20PSL1 biomass fractionantion with HILIC chromatography

Table S2

Region in the genome	Type of product	From (nt)	To (nt)	Total length (nt)	Well-defined Cluster	Most similar in MIBiG	Similarity (%)
3.1	Saccharide	35,353	98,067	62,714	No	–	–
4.1	Saccharide	1	38,963	38,962	No	–	–
6.1	Saccharide	20,517	44,049	23,532	No	–	–
8.1	Saccharide	7,801	28,352	20,551	No	O-antigen	15%
9.1	Saccharide	16,798	41,440	24,642	No	–	–
12.1	Saccharide	65,765	86,718	20,953	No	–	–
13.1	Saccharide	14,282	42,420	28,138	No	–	–
14.1	Saccharide	7,698	39,540	31,842	No	–	–
16.1	Saccharide	129,815	150,876	21,061	No	–	–
18.1	Saccharide	3,118	51,574	48,456	No	Pseudaminic acid	33%
21.1	Saccharide	1	53,814	53,813	No	–	–
24.1	***trans*-AT PKS**	**1**	**90,954**	**90,953**	**Yes**	**–**	**–**
26.1	Fatty acid	1	13,625	13,624	No	TP-1161	8%
29.1	Saccharide	102,497	131,899	29,402	No	–	–
33.1	Saccharide	236	21,441	21,205	No	–	–
33.2	Fatty acid	92,872	113,894	21,022	No	–	–
33.3	Saccharide	145,307	164,354	19,047	No	–	–
33.4	Saccharide	171,010	183,668	12,658	No	–	–
34.1	Saccharide	153,598	172,953	19,355	No	–	–
34.2	Saccharide	174,876	198,267	23,391	No	–	–
43.1	Fatty acid	25,114	46,424	21,310	No	–	–
46.1	Saccharide	105,630	128,267	22,637	No	–	–
46.2	Saccharide	188,125	252,773	64,648	No	–	–
46.3	**Terpene**	**266,688**	**281,144**	**14,456**	**Yes**	**–**	**–**
48.1	Saccharide	36,664	60,485	23,821	No	O-antigen	10%
49.1	Saccharide	1	25,416	25,415	No	–	–
51.1	Saccharide	17,757	37,514	19,757	No	–	–
51.2	**Cyanobactin**	**103,483**	**130,291**	**26,808**	**Yes**	**Limnothamide**	**71%**
51.3	Saccharide	162,928	193,100	30,172	No	–	–
51.4	Saccharide	326,302	353,583	27,281	No	–	–
51.5	Saccharide	400,189	423,035	22,846	No	–	–
52.1	Saccharide	51,329	73,806	22,477	No	–	–
56.1	**Terpene**	**36,604**	**57,728**	**21,124**	**Yes**	**–**	**–**

List of biosynthetic gene clusters predicted by antiSMASH 6.0 [3]. Well-defined clusters are in bold and highlighted in colors: yellow = Trans-acyltransferase polyketide synthases (*trans*-AT PKS), violet = terpenes, light-blue = cyanobactin. Most similar in MIBiG = the closest compound found in MIBiG data repository (Minimum Information about a Biosynthetic Gene cluster); Similarity (%) = percentage of genes within the closest known compound that have a significant BLAST hit to genes of the query region

References

[1] C.L.M. Gilchrist, Y.H. Chooi, Clinker & clustermap.js: Automatic generation of gene cluster comparison figures, Bioinformatics. 37 (2021) 2473–2475. https://doi.org/10.1093/bioinformatics/btab007.

[2] E.J.N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R.A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, Automated structure prediction of trans-acyltransferase polyketide synthase products, Nat. Chem. Biol. 15 (2019) 813–821. https://doi.org/10.1038/s41589-019-0313-7.

[3] K. Blin, S. Shaw, A.M. Kloosterman, Z. Charlop-Powers, G.P. Van Wezel, M.H. Medema, T. Weber, AntiSMASH 6.0: Improving cluster detection and comparison capabilities, Nucleic Acids Res. 49 (2021) W29–W35. https://doi.org/10.1093/nar/gkab335.