Assembly Summary p533: Plant Root Bacteria Genomes

Marco Kreuzer, IBU

03/07/2020

Contents

Summary

The sequence reads produced in this experiment are of good quality, but with high levels of duplication (see Table 1). The genome assembly was performed with SPAdes (version 3.14.0) and the quality of the assembly was assessed with Quast, Confindr and Busco4. Quast provides quantitative metrics of the assemlby, such as the N50 or the L75 metrics (Table 2). The results from confindr indicate that samples **LMX9**, **LMR1**, **LAR21** and **F12** are potentially contaminated (see Table 3). In order to assess whether core genes are assembled correctly, the tool BUSCO4 was applied. Overall, the assemblies show very low fractions of missing, fragmented or duplicated genes. However, samples **F12**, **LBA3** and **LMX9** show a high numbers of duplicated genes (see Figure 1). This could either be due to contamination or actual gene/genome duplication events. Given that confindr found **LMX9** and **F12** to be contaminated, these results point towards a technical issue rather than being biologically meaningful.

  The annotation was performed with Prokka (see Table 4). The taxonomic assignment of the genomes was calculated with GTDB-TK. Samples **LBA3**, **F12** and **LMX9** were assigned to the domain Archaea (see Table 5).

To do

- Meta-assembly of **LBA3**, **F12** and **LMX9** in order to try to separate contamination from contigs.

- Adding meta-data

Read Quality

FastQC

Table 1: Results from FastQC.

Sample	Deduplicated	Adapter	Reads
F01	54.17	pass	5401516
F02	49.98	pass	7785750
F07	47.91	pass	10023062
F119	46.30	pass	7209510
F12	62.60	pass	9895476
F123	45.08	pass	13425742
F125	34.25	pass	13525122
F201	56.82	pass	7993116
F213	54.39	pass	8104574
F214	58.79	pass	5676244
F216	53.45	pass	7650916
F220	54.34	pass	6901386
F230	53.66	pass	8379336
F233	48.44	pass	10086586
F240	42.45	pass	6568674
F242	47.50	pass	12283760
F244	38.45	pass	7833838
F247	42.12	pass	13568344
F49	38.03	pass	16709350
F85	24.11	pass	37087178
LAR12	45.40	pass	9310330
LAR21	44.70	pass	9717056
LBA112	51.90	pass	6881728
LBA21	47.91	pass	8468990
LBA3	57.24	pass	7503264
LBA71	47.78	pass	8416254
LM3X	39.22	pass	10027672
LMA1	37.19	pass	12914838
LMB2	37.48	pass	10824712
LMC3	42.60	pass	9854230
LMD1	48.13	pass	7770724
LME3	49.14	pass	7415878
LMF1	47.69	pass	18189356
LMG1	55.14	pass	10136144
LMI11	38.06	pass	10897694
LMI12	38.01	pass	14987326
LMI13	33.75	pass	14352842
LMI1x	34.99	pass	10090718
LMN1	55.31	pass	8567238
LMO1	40.29	pass	8254038

Table 1: Results from FastQC. *(continued)*

Sample	Deduplicated	Adapter	Reads
LMR1	54.06	pass	7566302
LMS1	48.61	pass	9548274
LMU1	53.71	pass	9637064
LMX7	36.94	pass	11502424
LMX9	52.27	pass	10445622
LMZ1	47.94	pass	7446756
LRH13	48.10	pass	9252620
LSP13	38.74	pass	11842650
LST12	35.32	pass	14938974
LST15	31.18	pass	18498630

Assembly Quality

Quast

Table 2: Results from quast.

Sample	Largest_contig	Total_length	GC_(%)	N50	N75	L50	L75	coverage_N50
F01	881245	4593112	37.59	327456	146534	4	9	20.36
F02	688616	5912340	35.08	303881	124160	7	16	25.01
F07	873518	6452612	61.54	506583	415433	6	9	33.17
F119	419546	4664190	41.64	95043	53069	14	30	30.03
F12	1830551	13785733	61.96	493203	243225	9	18	15.86
F123	1013952	6742308	60.42	671342	331020	4	9	42.57
F125	1640584	3894522	66.79	1452588	1452588	2	2	74.45
F201	1077358	7446743	61.21	304307	191100	6	13	21.25
F213	1387661	7334645	62.33	492167	261449	4	9	21.49
F214	858385	6183327	58.87	275088	141245	7	15	18.90
F216	864846	6568338	59.16	346667	239078	7	13	24.43
F220	414431	6323612	58.15	127096	80477	14	30	22.46
F230	2119226	6584332	59.64	602563	280469	3	8	24.95
F233	1311753	5531954	71.75	945335	378968	3	5	34.97
F240	128060	3407187	42.88	40440	21494	25	53	38.53
F242	2595692	7390886	63.05	825116	662717	3	5	37.19
F244	127756	3421114	42.86	38668	21342	27	55	47.14
F247	670068	6323408	60.91	397225	178437	6	13	41.77
F49	873518	6453746	61.54	526362	415402	5	9	55.65
F85	2606033	5818639	65.71	864313	289436	2	4	124.60
LAR12	877449	5145182	66.35	537289	117184	4	10	38.37
LAR21	728948	4985945	61.91	566766	260716	4	8	42.30
LBA112	940277	5958394	35.14	174209	79783	9	21	22.28
LBA21	790636	5437732	40.48	369740	187715	5	10	27.60
LBA3	832763	10399687	48.79	322064	158269	11	23	14.08
LBA71	690767	5847248	35.20	501680	158776	5	11	28.49
LM3X	2707111	3785152	69.35	2707111	975594	1	2	55.53
LMA1	1047191	4695253	63.31	572320	191481	3	6	54.80
LMB2	2707111	3784416	69.35	2707111	1040343	1	2	60.12
LMC3	1184260	4784280	63.23	256007	185239	4	9	41.61
LMD1	1143117	4689934	66.08	682763	460558	3	5	33.46

Table 2: Results from quast. *(continued)*

Sample	Largest_contig	Total_length	GC_(%)	N50	N75	L50	L75	coverage_N50
LME3	1146938	4893580	54.49	627549	336356	3	6	30.11
LMF1	1747085	10467927	71.39	509957	302651	6	13	34.85
LMG1	650459	8220706	72.33	155407	103252	14	30	23.84
LMI11	1384000	3783591	67.92	1365337	598712	2	3	60.55
LMI12	1548014	4884442	70.06	1217161	452683	2	4	62.72
LMI13	1384000	3783375	67.92	1365337	598712	2	3	79.79
LMI1x	1558175	2959162	69.84	1558175	1256048	1	2	75.65
LMN1	2135740	7874263	46.70	904160	897057	3	5	21.14
LMO1	856084	3699876	41.31	453585	271470	3	5	43.35
LMR1	1614850	6274515	58.87	464579	173181	3	9	25.32
LMS1	1934391	6550235	62.53	835294	614269	3	5	30.28
LMU1	589365	7538343	64.42	374034	204031	8	15	23.93
LMX7	2707111	3784925	69.35	2707111	1040343	1	2	62.33
LMX9	837298	11626669	58.15	325104	194123	11	22	15.16
LMZ1	1648363	4749811	54.53	786934	495464	2	4	30.63
LRH13	1882123	5421971	59.35	841127	417380	2	5	34.76
LSP13	674912	4670527	63.19	346939	134625	5	11	51.88
LST12	1501533	4646124	66.12	1217203	958160	2	3	64.89
LST15	1456640	4558127	66.23	509226	425672	3	5	79.69

confindr

Table 3: Results from confindr. This tool analyses genes that are known to be single-copy and conserved accross all bacteria and flags assemblies, where there are more than one allele present in some genes.

Sample	Genus	NumContamSNVs	PercentContam	ContamStatus
F01_1	ND	1	0	FALSE
F01_2	ND	6	5.66	TRUE
F02_1	Bacillus	0	0	FALSE
F02_2	Bacillus	1	0	FALSE
F07_1	ND	0	0	FALSE
F07_2	ND	1	0	FALSE
F119_1	ND	0	0	FALSE
F119_2	ND	1	0	FALSE
F12_1	ND	11	8.54	TRUE
F12_2	ND	13	11.89	TRUE
F123_1	Pseudomonas	0	0	FALSE
F123_2	Pseudomonas	0	0	FALSE
F125_1	ND	0	0	FALSE
F125_2	ND	1	0	FALSE
F201_1	ND	1	0	FALSE
F201_2	ND	1	0	FALSE
F213_1	ND	2	0	FALSE
F213_2	ND	0	0	FALSE
F214_1	ND	1	0	FALSE
F214_2	ND	1	0	FALSE
F216_1	ND	1	0	FALSE

Table 3: Results from confindr. This tool analyses genes that are known to be single-copy and conserved accross all bacteria and flags assemblies, where there are more than one allele present in some genes. *(continued)*

Sample	Genus	NumContamSNVs	PercentContam	ContamStatus
F216_2	ND	1	0	FALSE
F220_1	ND	1	0	FALSE
F220_2	ND	0	0	FALSE
F230_1	ND	0	0	FALSE
F230_2	ND	0	0	FALSE
F233_1	ND	2	0	FALSE
F233_2	ND	2	0	FALSE
F240_1	ND	0	0	FALSE
F240_2	ND	0	0	FALSE
F242_1	ND	0	0	FALSE
F242_2	ND	1	0	FALSE
F244_1	ND	0	0	FALSE
F244_2	ND	0	0	FALSE
F247_1	ND	0	0	FALSE
F247_2	Error processing sample	0	ND	FALSE
F49_1	ND	0	0	FALSE
F49_2	ND	1	0	FALSE
F85_1	Error processing sample	0	ND	FALSE
F85_2	Error processing sample	0	ND	FALSE
LAR12_1	Error processing sample	0	ND	FALSE
LAR12_2	Error processing sample	0	ND	FALSE
LAR21_1	ND	27	38.67	TRUE
LAR21_2	ND	29	37.40	TRUE
LBA112_1	Error processing sample	0	ND	FALSE
LBA112_2	Error processing sample	0	ND	FALSE
LBA21_1	Bacillus	0	0	FALSE
LBA21_2	Bacillus	1	0	FALSE
LBA3_1	Bacillus	0	0	FALSE
LBA3_2	Bacillus	1	0	FALSE
LBA71_1	Bacillus	0	0	FALSE
LBA71_2	Bacillus	0	0	FALSE
LM3X_1	Error processing sample	0	ND	FALSE
LM3X_2	Error processing sample	0	ND	FALSE
LMA1_1	ND	0	0	FALSE
LMA1_2	ND	0	0	FALSE
LMB2_1	ND	0	0	FALSE
LMB2_2	ND	0	0	FALSE
LMC3_1	Error processing sample	0	ND	FALSE
LMC3_2	Error processing sample	0	ND	FALSE
LMD1_1	ND	0	0	FALSE
LMD1_2	ND	0	0	FALSE
LME3_1	Enterobacter	0	0	FALSE
LME3_2	Enterobacter	0	0	FALSE

Table 3: Results from confindr. This tool analyses genes that are known to be single-copy and conserved accross all bacteria and flags assemblies, where there are more than one allele present in some genes. *(continued)*

Sample	Genus	NumContamSNVs	PercentContam	ContamStatus
LMF1_1	ND	1	0	FALSE
LMF1_2	ND	0	0	FALSE
LMG1_1	ND	0	0	FALSE
LMG1_2	ND	1	0	FALSE
LMI11_1	ND	0	0	FALSE
LMI11_2	ND	0	0	FALSE
LMI12_1	ND	0	0	FALSE
LMI12_2	ND	2	0	FALSE
LMI13_1	ND	0	0	FALSE
LMI13_2	ND	0	0	FALSE
LMI1x_1	ND	0	0	FALSE
LMI1x_2	ND	0	0	FALSE
LMN1_1	ND	0	0	FALSE
LMN1_2	ND	0	0	FALSE
LMO1_1	ND	0	0	FALSE
LMO1_2	Bacillus	0	0	FALSE
LMR1_1	ND	8	7.31	TRUE
LMR1_2	ND	10	13.00	TRUE
LMS1_1	ND	0	0	FALSE
LMS1_2	ND	1	0	FALSE
LMU1_1	ND	0	0	FALSE
LMU1_2	ND	0	0	FALSE
LMX7_1	ND	1	0	FALSE
LMX7_2	ND	0	0	FALSE
LMX9_1	Enterobacter	173	20.60	TRUE
LMX9_2	Enterobacter	165	22.85	TRUE
LMZ1_1	Enterobacter	0	0	FALSE
LMZ1_2	Enterobacter	0	0	FALSE
LRH13_1	ND	2	0	FALSE
LRH13_2	ND	0	0	FALSE
LSP13_1	ND	0	0	FALSE
LSP13_2	ND	0	0	FALSE
LST12_1	ND	0	0	FALSE
LST12_2	ND	0	0	FALSE
LST15_1	ND	1	0	FALSE
LST15_2	ND	0	0	FALSE

BUSCO Assessment Results

Complete (C) and single-copy (S)   **Complete (C) and duplicated (D)**

Fragmented (F)   **Missing (M)**

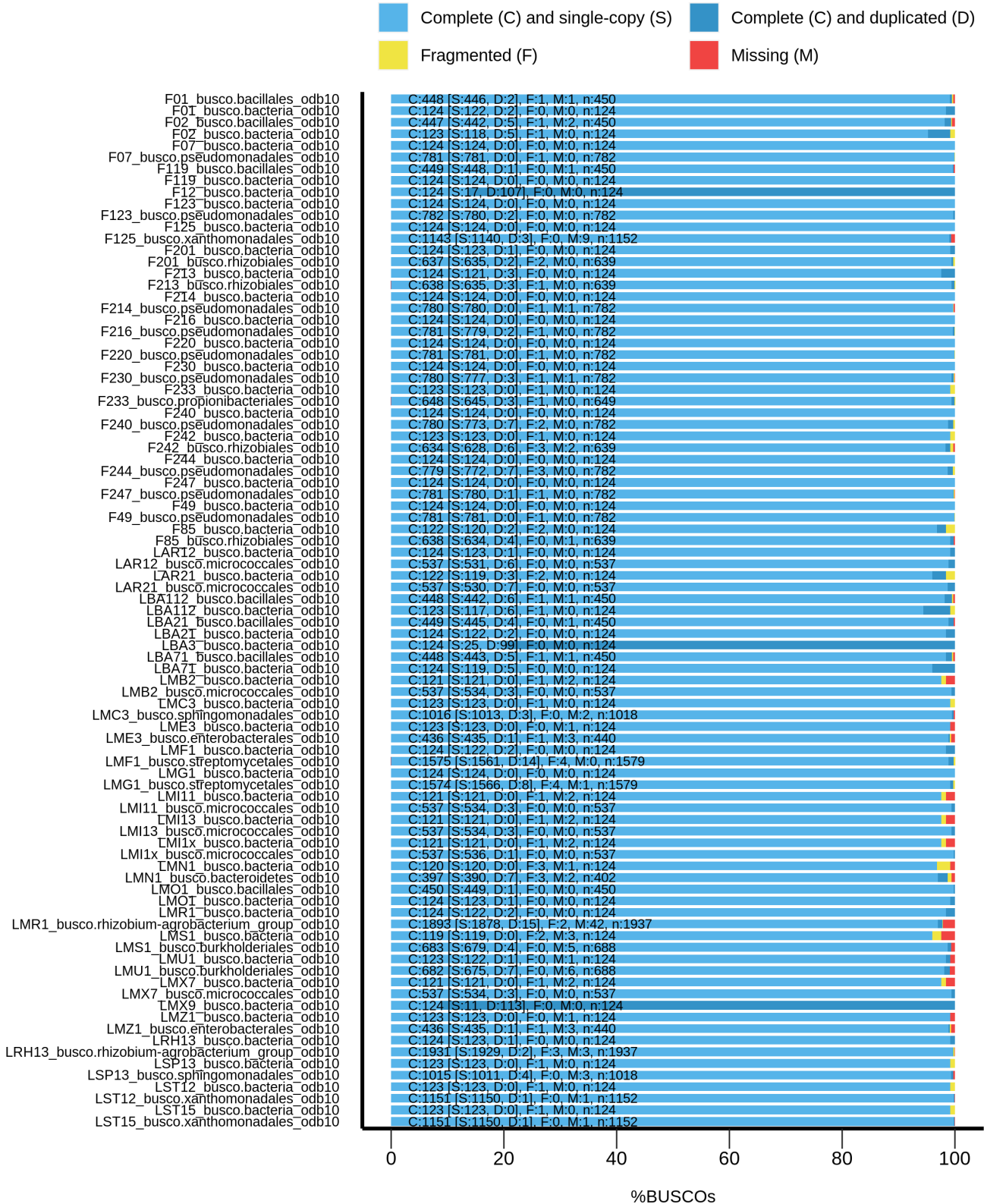| Sample	BUSCO Result
F01_busco.bacillales_odb10	C:448 [S:446, D:2], F:1, M:1, n:450
F01_busco.bacteria_odb10	C:124 [S:122, D:2], F:0, M:0, n:124
F02_busco.bacillales_odb10	C:447 [S:442, D:5], F:1, M:2, n:450
F02_busco.bacteria_odb10	C:123 [S:118, D:5], F:1, M:0, n:124
F07_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F07_busco.pseudomonadales_odb10	C:781 [S:781, D:0], F:1, M:0, n:782
F119_busco.bacillales_odb10	C:449 [S:448, D:1], F:0, M:1, n:450
F119_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F12_busco.bacteria_odb10	C:124 [S:17, D:107], F:0, M:0, n:124
F123_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F123_busco.pseudomonadales_odb10	C:782 [S:780, D:2], F:0, M:0, n:782
F125_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F125_busco.xanthomonadales_odb10	C:1143 [S:1140, D:3], F:0, M:9, n:1152
F201_busco.bacteria_odb10	C:124 [S:123, D:1], F:0, M:0, n:124
F201_busco.rhizobiales_odb10	C:637 [S:635, D:2], F:2, M:0, n:639
F213_busco.bacteria_odb10	C:124 [S:121, D:3], F:0, M:0, n:124
F213_busco.rhizobiales_odb10	C:638 [S:635, D:3], F:1, M:0, n:639
F214_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F214_busco.pseudomonadales_odb10	C:780 [S:780, D:0], F:1, M:1, n:782
F216_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F216_busco.pseudomonadales_odb10	C:781 [S:779, D:2], F:1, M:0, n:782
F220_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F220_busco.pseudomonadales_odb10	C:781 [S:781, D:0], F:1, M:0, n:782
F230_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F230_busco.pseudomonadales_odb10	C:780 [S:777, D:3], F:1, M:1, n:782
F233_busco.bacteria_odb10	C:123 [S:123, D:0], F:1, M:0, n:124
F233_busco.propionibacteriales_odb10	C:648 [S:645, D:3], F:1, M:0, n:649
F240_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F240_busco.pseudomonadales_odb10	C:780 [S:773, D:7], F:2, M:0, n:782
F242_busco.bacteria_odb10	C:123 [S:123, D:0], F:1, M:0, n:124
F242_busco.rhizobiales_odb10	C:634 [S:628, D:6], F:3, M:2, n:639
F244_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F244_busco.pseudomonadales_odb10	C:779 [S:772, D:7], F:3, M:0, n:782
F247_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F247_busco.pseudomonadales_odb10	C:781 [S:780, D:1], F:1, M:0, n:782
F49_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
F49_busco.pseudomonadales_odb10	C:781 [S:781, D:0], F:1, M:0, n:782
F85_busco.bacteria_odb10	C:122 [S:120, D:2], F:2, M:0, n:124
F85_busco.rhizobiales_odb10	C:638 [S:634, D:4], F:0, M:1, n:639
LAR12_busco.bacteria_odb10	C:124 [S:123, D:1], F:0, M:0, n:124
LAR12_busco.micrococcales_odb10	C:537 [S:531, D:6], F:0, M:0, n:537
LAR21_busco.bacteria_odb10	C:122 [S:119, D:3], F:2, M:0, n:124
LAR21_busco.micrococcales_odb10	C:537 [S:530, D:7], F:0, M:0, n:537
LBA112_busco.bacillales_odb10	C:448 [S:442, D:6], F:1, M:1, n:450
LBA112_busco.bacteria_odb10	C:123 [S:117, D:6], F:1, M:0, n:124
LBA21_busco.bacillales_odb10	C:449 [S:445, D:4], F:0, M:1, n:450
LBA21_busco.bacteria_odb10	C:124 [S:122, D:2], F:0, M:0, n:124
LBA3_busco.bacteria_odb10	C:124 [S:25, D:99], F:0, M:0, n:124
LBA71_busco.bacillales_odb10	C:448 [S:443, D:5], F:1, M:1, n:450
LBA71_busco.bacteria_odb10	C:124 [S:119, D:5], F:0, M:0, n:124
LMB2_busco.bacteria_odb10	C:121 [S:121, D:0], F:1, M:2, n:124
LMB2_busco.micrococcales_odb10	C:537 [S:534, D:3], F:0, M:0, n:537
LMC3_busco.bacteria_odb10	C:123 [S:123, D:0], F:1, M:0, n:124
LMC3_busco.sphingomonadales_odb10	C:1016 [S:1013, D:3], F:0, M:2, n:1018
LME3_busco.bacteria_odb10	C:123 [S:123, D:0], F:0, M:1, n:124
LME3_busco.enterobacterales_odb10	C:436 [S:435, D:1], F:1, M:3, n:440
LMF1_busco.bacteria_odb10	C:124 [S:122, D:2], F:0, M:0, n:124
LMF1_busco.streptomycetales_odb10	C:1575 [S:1561, D:14], F:4, M:0, n:1579
LMG1_busco.bacteria_odb10	C:124 [S:124, D:0], F:0, M:0, n:124
LMG1_busco.streptomycetales_odb10	C:1574 [S:1566, D:8], F:4, M:1, n:1579
LMI11_busco.bacteria_odb10	C:121 [S:121, D:0], F:1, M:2, n:124
LMI11_busco.micrococcales_odb10	C:537 [S:534, D:3], F:0, M:0, n:537
LMI13_busco.bacteria_odb10	C:121 [S:121, D:0], F:1, M:2, n:124
LMI13_busco.micrococcales_odb10	C:537 [S:534, D:3], F:0, M:0, n:537
LMI1x_busco.bacteria_odb10	C:121 [S:121, D:0], F:1, M:2, n:124
LMI1x_busco.micrococcales_odb10	C:537 [S:536, D:1], F:0, M:0, n:537
LMN1_busco.bacteria_odb10	C:120 [S:120, D:0], F:3, M:1, n:124
LMN1_busco.bacteroidetes_odb10	C:397 [S:390, D:7], F:3, M:2, n:402
LMO1_busco.bacillales_odb10	C:450 [S:449, D:1], F:0, M:0, n:450
LMO1_busco.bacteria_odb10	C:124 [S:123, D:1], F:0, M:0, n:124
LMR1_busco.bacteria_odb10	C:124 [S:122, D:2], F:0, M:0, n:124
LMR1_busco.rhizobium-agrobacterium_group_odb10	C:1893 [S:1878, D:15], F:2, M:42, n:1937
LMS1_busco.bacteria_odb10	C:119 [S:119, D:0], F:2, M:3, n:124
LMS1_busco.burkholderiales_odb10	C:683 [S:679, D:4], F:0, M:5, n:688
LMU1_busco.bacteria_odb10	C:123 [S:122, D:1], F:0, M:1, n:124
LMU1_busco.burkholderiales_odb10	C:682 [S:675, D:7], F:0, M:6, n:688
LMX7_busco.bacteria_odb10	C:121 [S:121, D:0], F:1, M:2, n:124
LMX7_busco.micrococcales_odb10	C:537 [S:534, D:3], F:0, M:0, n:537
LMX9_busco.bacteria_odb10	C:124 [S:11, D:113], F:0, M:0, n:124
LMZ1_busco.bacteria_odb10	C:123 [S:123, D:0], F:0, M:1, n:124
LMZ1_busco.enterobacterales_odb10	C:436 [S:435, D:1], F:1, M:3, n:440
LRH13_busco.bacteria_odb10	C:124 [S:123, D:1], F:0, M:0, n:124
LRH13_busco.rhizobium-agrobacterium_group_odb10	C:1931 [S:1929, D:2], F:3, M:3, n:1937
LSP13_busco.bacteria_odb10	C:123 [S:123, D:0], F:1, M:0, n:124
LSP13_busco.sphingomonadales_odb10	C:1015 [S:1011, D:4], F:0, M:3, n:1018
LST12_busco.bacteria_odb10	C:123 [S:123, D:0], F:1, M:0, n:124
LST12_busco.xanthomonadales_odb10	C:1151 [S:1150, D:1], F:0, M:1, n:1152
LST15_busco.bacteria_odb10	C:123 [S:123, D:0], F:1, M:0, n:124
LST15_busco.xanthomonadales_odb10	C:1151 [S:1150, D:1], F:0, M:1, n:1152

%BUSCOs

Figure 1: Genome assembly assessment using BUSCO4. The figure shows how many expected Genes are complete, complete & duplicate, fragmented or missing.

Annotation and Taxonomy

Prokka Annotation

Table 4: Annotation results from prokka analysis.

Sample	contigs	bases	CDS	rRNA	tRNA	tmRNA	repeat_region
F01	83	4593112	4414	26	149	1	1
F02	99	5912340	5880	15	100	1	NA
F07	53	6452612	5769	8	70	1	NA
F119	149	4664190	4604	12	86	NA	3
F12	165	13785733	12623	11	131	2	NA
F123	46	6742308	6076	5	61	1	NA
F125	14	3894522	3425	6	55	1	NA
F201	103	7446743	6986	3	61	1	NA
F213	94	7334645	6840	3	61	1	NA
F214	72	6183327	5706	4	63	1	NA
F216	75	6568338	5862	6	64	1	NA
F220	215	6323612	5733	5	63	1	NA
F230	55	6584332	6000	6	65	1	NA
F233	30	5531954	5283	5	56	1	NA
F240	292	3407187	3263	9	82	1	NA
F242	32	7390886	6753	3	52	1	NA
F244	305	3421114	3281	9	82	1	NA
F247	75	6323408	5529	8	66	1	NA
F49	63	6453746	5772	8	71	1	NA
F85	49	5818639	5431	3	53	1	NA
LAR12	81	5145182	4665	7	58	1	NA
LAR21	46	4985945	4576	8	56	1	NA
LBA112	159	5958394	6088	18	103	1	NA
LBA21	60	5437732	5252	16	82	1	NA
LBA3	130	10399687	10172	24	173	2	NA
LBA71	98	5847248	5993	19	101	1	NA
LM3X	10	3785152	3681	6	51	1	NA
LMA1	74	4695253	4336	5	56	1	NA
LMB2	7	3784416	3682	6	51	1	NA
LMC3	96	4784280	4446	6	57	1	NA
LMD1	31	4689934	4303	6	56	1	NA
LME3	48	4893580	4575	9	79	1	NA
LMF1	67	10467927	9039	7	84	1	NA
LMG1	113	8220706	7249	8	80	1	NA
LMI11	14	3783591	3604	5	51	1	NA
LMI12	24	4884442	4381	3	58	1	NA
LMI13	13	3783375	3605	5	51	1	NA
LMI1x	7	2959162	2850	5	54	1	NA
LMN1	67	7874263	6304	9	69	1	NA
LMO1	33	3699876	3747	10	73	1	NA
LMR1	113	6274515	6095	7	62	1	NA
LMS1	68	6550235	5720	4	96	1	NA
LMU1	86	7538343	6261	5	87	1	NA
LMX7	10	3784925	3680	6	51	1	NA
LMX9	99	11626669	10514	15	135	2	NA
LMZ1	39	4749811	4448	13	79	1	NA
LRH13	23	5421971	5145	3	52	1	NA
LSP13	103	4670527	4322	5	56	1	NA

Table 4: Annotation results from prokka analysis. *(continued)*

Sample	contigs	bases	CDS	rRNA	tRNA	tmRNA	repeat_region
LST12	14	4646124	4146	7	69	1	NA
LST15	37	4558127	4091	11	70	1	NA

Taxonomy

Table 5: Results from GTDB-TK. The column placement method indicates whther the taxonomic assignment was done by Average Nucleotide Identity (ANI), or by estimating the placement in a phylogenetic reference tree (placement). "ANI/placement" means that both methods agree on the taxonomic assignment.

Sample	Domain	Phylum	Class	Order	Family	Genus	Species	classification_method
F01	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_A	Bacillus_X		Placement
F02	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_G	Bacillus_A	Bacillus_A toyonensis	ANI/Placement
F07	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E		Placement
F119	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_A	Bacillus_BJ	Bacillus_BJ canaveralius	ANI/Placement
F12	Archaea	Thermoplasmatota	Poseidoniia	Poseidoniales	Poseidoniaceae	MGIIa-K1		Placement
F123	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E	Pseudomonas_E marginalis	ANI/Placement
F125	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Rhodanobacteraceae	Rhodanobacter		Placement
F201	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Ensifer	Ensifer adhaerens_A	ANI/Placement
F213	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Ensifer	Ensifer adhaerens_F	ANI/Placement
F214	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E	Pseudomonas_E fluorescens_E	ANI/Placement
F216	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E	Pseudomonas_E sp001655615	ANI/Placement
F220	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E		Placement
F230	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E	Pseudomonas_E moorei	ANI/Placement
F233	Bacteria	Actinobacteriota	Actinobacteria	Propionibacteriales	Nocardioidaceae	Nocardioides		Placement
F240	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Acinetobacter fasciculus	ANI/Placement
F242	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Labraceae	Labrys		Placement
F244	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Acinetobacter fasciculus	ANI/Placement
F247	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E	Pseudomonas_E brassicacearum_C	ANI
F49	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas_E		Placement
F85	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	Bosea		Placement
LAR12	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Micrococcaceae	Arthrobacter_I		Placement
LAR21	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Micrococcaceae	Paenarthrobacter	Paenarthrobacter nitroguajacolicus	ANI/Placement
LBA112	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_G	Bacillus_A	Bacillus_A cereus_AY	ANI/Placement
LBA21	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_A	Bacillus_X	Bacillus_X frigoritolerans	ANI/Placement
LBA3	Archaea	Nanohaloarchaeota	Nanosalinia					Placement
LBA71	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_G	Bacillus_A	Bacillus_A cereus_P	ANI/Placement
LM3X	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	Microbacterium sp001262495	ANI/Placement
LMA1	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium		Placement
LMB2	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	Microbacterium sp001262495	ANI/Placement
LMC3	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium		Placement
LMD1	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Micrococcaceae	Pseudarthrobacter		Placement
LME3	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Enterobacter	Enterobacter ludwigii	ANI/Placement
LMF1	Bacteria	Actinobacteriota	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces	Streptomyces ossamyceticus	ANI/Placement
LMG1	Bacteria	Actinobacteriota	Actinobacteria	Streptomycetales	Streptomycetaceae	Streptomyces		Placement
LMI11	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	Microbacterium foliorum_A	ANI/Placement
LMI12	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium		Placement
LMI13	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	Microbacterium foliorum_A	ANI/Placement
LMI1x	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	Microbacterium sp000383475	ANI/Placement
LMN1	Bacteria	Bacteroidota	Bacteroidia	Chitinophagales	Chitinophagaceae	Chitinophaga	Chitinophaga ginsengisegetis	ANI/Placement
LMO1	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	Bacillus altitudinis	ANI/Placement
LMR1	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Neorhizobium	Neorhizobium sp900472575	ANI/Placement
LMS1	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	Janthinobacterium	Janthinobacterium lividum_C	ANI/Placement
LMU1	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Burkholderiaceae	Duganella		Placement
LMX7	Bacteria	Actinobacteriota	Actinobacteria	Actinomycetales	Microbacteriaceae	Microbacterium	Microbacterium sp001262495	ANI/Placement

Table 5: Results from GTDB-TK. The column placement method indicates whther the taxonomic assignment was done by Average Nucleotide Identity (ANI), or by estimating the placement in a phylogenetic reference tree (placement). "ANI/placement" means that both methods agree on the taxonomic assignment. *(continued)*

Sample	Domain	Phylum	Class	Order	Family	Genus	Species	classification_method
LMX9	Archaea	Asgardarchaeota	Lokiarchaeia	CR-4				Placement
LMZ1	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Enterobacter	Enterobacter ludwigii	ANI/Placement
LRH13	Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Agrobacterium	Agrobacterium tumefaciens	ANI/Placement
LSP13	Bacteria	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium		Placement
LST12	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	Stenotrophomonas indicatrix	ANI/Placement
LST15	Bacteria	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	Stenotrophomonas lactitubi	ANI/Placement

Software Versions

```
## [1] "fastp_version,0.19.5,"
## [2] "fastqc_version,0.11.7,"
## [3] "spades_version,3.14.0,"
## [4] "quast_version,4.6.0,"
## [5] "BUSCO 4.0.6"
## [6] "ConFindr 0.7.2"
## [7] "gtdbtk: version 0.3.2 Copyright 2017 Pierre Chaumeil"
```