
1
2
3
4
5
6
7
8

A FASTQ
B TSV
C BIOM
D

11
1
1
1
2
3
3
6
6
8
9
10
12
13

—

GB/T 1.1—2020

1

GB/T 35890
GB/T 40226-2021

DNA

DNA RNA
16S rDNA

18S rDNA ITS Internal Transcribed Spacer

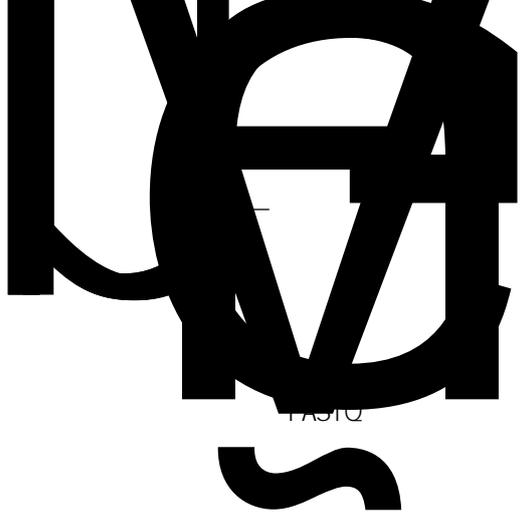
BI OM

OTU

OTU

FASTQ

DNA RNA



D t

FASTQ

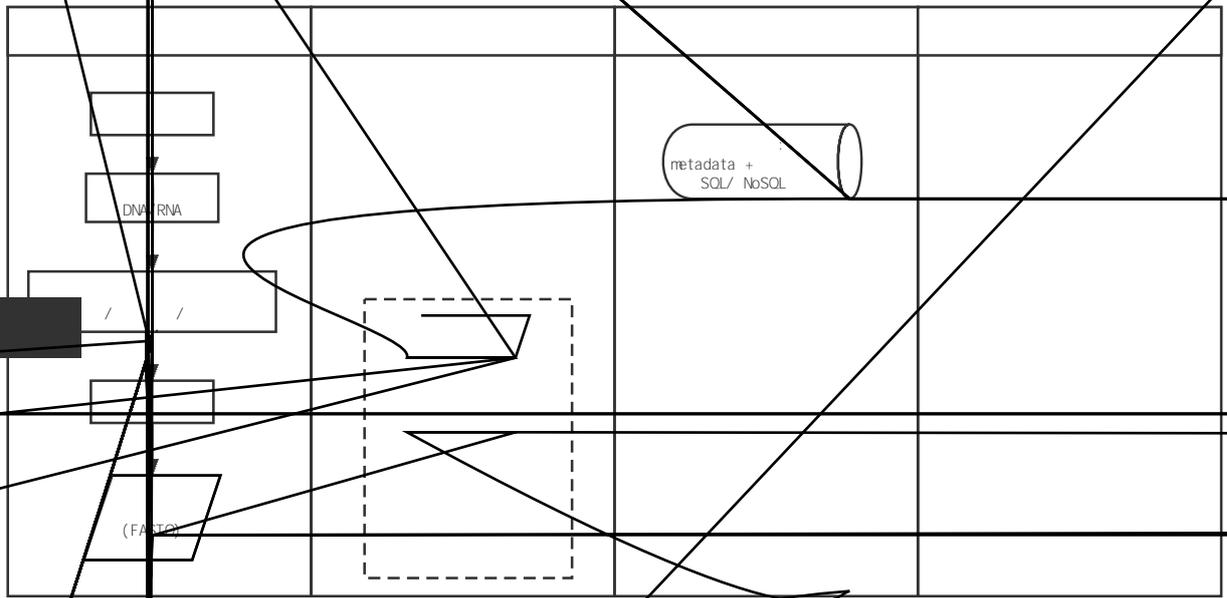
TSV

Uni Ref
API
OTU
TSV

Uni Prot Reference Clusters
Application Programming Interface
Operational Taxonomic Units
Tab-Separated Values

1

RESTful API



bi osampl e_si te				gut, oral
col l ecti on_date				2021-01-01 T00:00:00+0800
bi osampl e_store_temp		:		-80
extracti on_method				Power Soil [®] DNA Isolati on Ki t
extracti on_ti me				2021-01-02 T00:00:00+0800
DNA_mass	DNA			1.0 μ g
DNA concentrati on	DNA			10 ng/μ L
DNA compl eteness	DNA			
DNA_qual i ty	DNA			A
Li brary_i d				HWBAYTGAA170328, HWBAYTGAA170329
DNA " " GB/T 40226-2021 10.1.2				

2

l i brary_i d				HWBAYTGAA170328-18
l i brary_strategy				WGA, AMPLI CON
l i brary_source				METAGENOMIC
l i brary_sel ecti on		/		unspeci fi ed
l i brary_l ayout				2
i nsert_si ze	(bp)			250
nomi nal_si ze	(bp)			10
pl atform				DNBSEQ-G400, DNBSEQ-G400
i nstrument_model				DNBSEQ-G400, DNBSEQ-G400
read_type				PE100, PE100
sl i de_i d				CL100031417, CL100031417
l ane_i d				L02, L02
barcode		,		525, 525

sequencing_time				2017-07-01T00:00:00+0800, 2017-07-01T00:00:00+0800
raw_data_readnum	FASTQ			2000000000, 2000000000
raw_data_basenum	FASTQ			2000000000, 2000000000

raw_data_gc_content	FASTQ	GC		0.412012, 0.412039
raw_data_n_count	FASTQ	N		0.056686, 0.056752
raw_data_q20	FASTQ	Phred	20	0.9680, 0.9521
raw_data_q30	FASTQ	Phred	30	0.8861, 0.8512
clean_data_readnum	FASTQ			2000000, 2000000
clean_data_basenum	FASTQ			200000000, 200000000
clean_data_gc_content	FASTQ	GC		0.412012, 0.412039
clean_data_n_count	FASTQ	N		0.056686, 0.056752
clean_data_q20	FASTQ	Phred	20	0.9680, 0.9521
clean_data_q30	FASTQ	Phred	30	0.8861, 0.8512
host_rate				0.95
" "				

3

pipeline / MetaPhlan2
 pipeline_version / 2.7.1
 data_q3_eta

—

Integrated Gene Catalog, IGC
Unified Human Gastrointestinal Protein Catalog

Uni Ref

MetaPhlan2 mOTU Kraken2
QIIME2

HUMAN2

(CARD)

KEGG eggNOG metaCyc
(CAZy)

FASTQ FASTQ

GB/T 35890

A

—

D JSON API

—

@CL100122427L1C001R001_16
CCGCTTTGAGACCTTTGCTGGAAATGGGAATATCT
+
FFFFFFFF>FCFFFFFFFFFAFFEFFEDDFFFD
@CL100122427L1C001R001_17
CCGCTTTGAGACCTTTGCTGGAAATGGGAATATCT
+
FFFFFFFF>FCFFFFFFFFFAFFEFFEDDFFFD

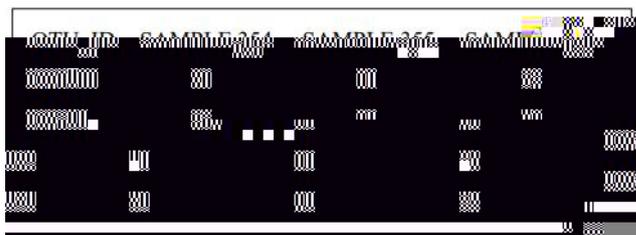
TSV

B.1 B.2

90%

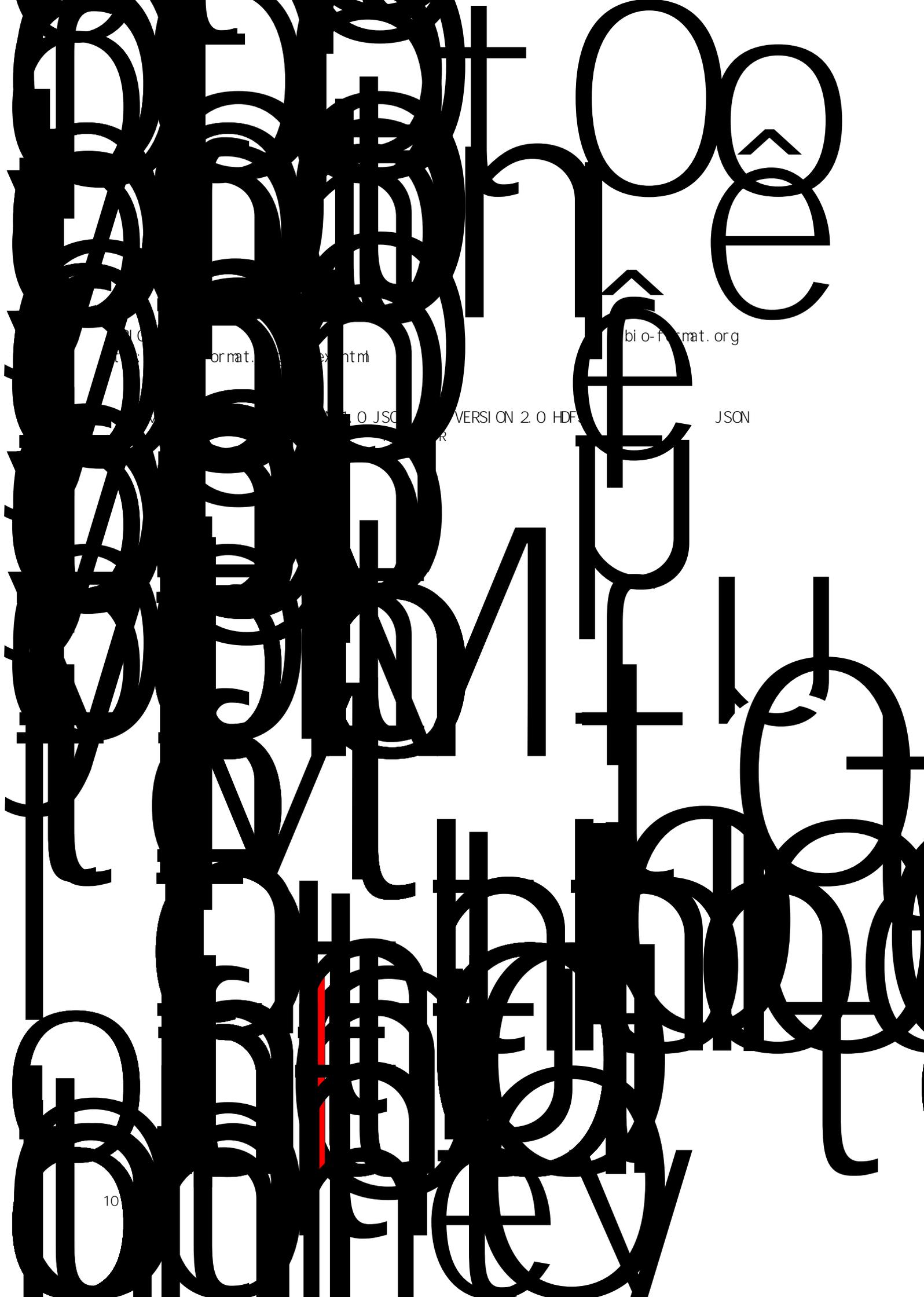
0

OTU_ID	SAMPLE.354
OTU0	1
OTU1	5
OTU2	1
OTU3	2
OTU4	1



0

SAMPLE.354	OTU1	6
SAMPLE.354	OTU2	1
SAMPLE.356	OTU0	4
SAMPLE.356	OTU2	7
SAMPLE.356	OTU3	3



```
[ 3, 2, 1],
      [ 3, 5, 1],
      [ 4, 1, 1],
      [ 4, 2, 1]
    ]
  }
}
```

JSON

```
{
  "id": null,
  "format": "Biological Observation Matrix 0.9.1-dev",
  "format_url":
"http://biomformat.org/documentation/format_versions/biom1.0.html",
  "type": "OTU table",
  "generated_by": "QIIME revision 1.4.0-dev",
  "date": "2011-12-19T19:00:00",
  "rows": [
    {"id": "GG_OTU_1", "metadata": null},
    {"id": "GG_OTU_2", "metadata": null},
    {"id": "GG_OTU_3", "metadata": null},
    {"id": "GG_OTU_4", "metadata": null},
    {"id": "GG_OTU_5", "metadata": null}
  ],
  "columns": [
    {"id": "Sample1", "metadata": null},
    {"id": "Sample2", "metadata": null},
    {"id": "Sample3", "metadata": null},
    {"id": "Sample4", "metadata": null},
    {"id": "Sample5", "metadata": null},
    {"id": "Sample6", "metadata": null}
  ],
  "matrix_type": "dense",
  "matrix_element_type": "int",
  "shape": [5, 6],
  "data": [[0, 0, 1, 0, 0, 0],
           [5, 1, 0, 2, 3, 1],
           [0, 0, 1, 4, 2, 0],
           [2, 1, 1, 0, 0, 1],
           [0, 1, 1, 0, 0, 0]]
}
```

D. 1

/			
sample_info	-	-	
sample_info	biosample_id	-	
sample_info	sample_time	-	
sample_info	biosample_site	-	
sample_info	specimen	-	
observation	-	-	
observation	unit	-	
observation	type	-	
observation	value	-	value 10.23, unit percent, 10.23%
coding	-	-	/ /
coding	system	-	NCBI : https://www.ncbi.nlm.nih.gov/Taxonomy/ MetaCyc: https://metacyc.org/
coding	id	-	, txid537011 GLUCOSE1PMTAB
coding	name	-	s_Prevotella_copri glucose and glucose-1-phosphate degradation
background	-	-	
background	reference_cohort	-	
background	background	-	, 5% 25% 50% 75% 95%
background	frequency	-	
background	rank_ratio	-	
background	mean	-	

-
- [1] GB/T 7408
- [2] Kottmann R, Gray T, Murphy S, et al. A standard MGS/MMS compliant XML Schema: toward the development of the Genomic Contextual Data Markup Language (GCDML). *OMICS*. 2008; 12(2): 115-121. doi: 10.1089/omi.2008.0A10
- [3] McDonald D, Clemente JC, Kuczynski J, et al. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the one-one. *Genascience*. 2012; 1(1): 7. Published 2012 Jul 12. doi: 10.1186/2047-217X-1-7
-