

# Typing Environmental Microorganisms To Genomic Databases Using MALDI Mass Spectrometry

Kenneth C. Parker Virgin Instruments Corporation, Marlborough, MA. Email: Kenneth.parker@simultof.com

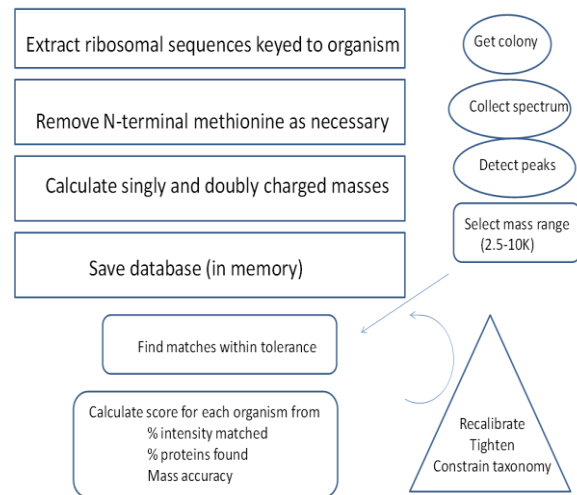


ThP  
523

## Introduction

MALDI has been used to identify clinical microorganisms for decades, usually by comparing spectra to reference spectra obtained from known strains of bacteria. Meanwhile, other scientists have been sequencing new genomes of bacteria at a prodigious rate. Thirdly, environmental microbiologists have been studying environmental samples by sequencing ribosomal RNA encoding DNA leading to tens of thousands of operational taxonomic units (OTUs) per sample. The differences between these methods lead to some taxonomic confusion. We describe methodology for identifying environmental bacterial colonies by MALDI by direct comparison to deposited genomic information. Each identification organizes the NCBI genomes centered upon the spectrum obtained, and indicates how similar the isolate is to these genomes, regardless of whether the genomes are consistently assigned to species.

## Methods



12 *Pseudomonas* isolates from surface water studied here: **Clades defined by shared ribosomal protein sequences.** (see Tables 6 and 7). Isolates selected to represent 12 different clades.

Table 1. Matching statistics for 12 *Pseudomonas* isolates Figure 1. Spectra of *Pseudomonas* isolates; matches to ribosomal proteins within 500 ppm (internal calibration)

N	OID	Clade	Peaks	Proteins	Matches	Score	Org
1	1827	1507	134	100	33	3592	lini
2	1525	1525	134	104	53	9802	fluorescens 1
3	1071	1715	90	103	39	4666	brenneri
4	1717	1717	115	102	48	6530	gingeri 1
5	1537	1791	113	104	37	4773	koreensis 1
6	3715	1861	119	104	45	5942	lurida
7	4337	1819	69	101	24	2749	libanensis
8	2871	2087	84	99	29	2405	poae 1
9	3217	3217	106	105	40	6383	putida 1
10	1539	3117	138	104	38	4795	fluorescens 2
11	1441	3981	108	105	32	3717	asplenii 1
12	4319	4319	133	98	47	10900	stutzeri 1

Note inconsistency in peak number: Based on inconsistency of peak detection parameters (unintended): Robust ID -> ~independent of # of peaks.

Within 500 ppm

Table 2. Matches to 129 peaks from 'brenneri' spectrum (2<sup>nd</sup> above).

Clade	Proteins	Matches	Score	%	%Ch	ppm	Organism
1	1827	104	49	6479	44	60	78 Pseudomonas sp. FH4
2	1525	103	48	6408	44	60	76 Pseudomonas brenneri BS2771
3	1071	102	47	6304	44	60	77 Pseudomonas sp. 25 R 14
4	1715	100	45	5878	42	59	80 Pseudomonas sp. 06C 126
5	1537	103	45	5394	41	58	80 Pseudomonas proteolytica BS2985
6	3715	104	43	4710	39	55	83 Pseudomonas fluorescens Pf08
7	4337	98	39	4661	41	54	86 Pseudomonas sp. CF150
8	2871	90	39	4419	34	57	89 Pseudomonas fluorescens NCTC3756
9	3217	99	39	4378	38	54	82 Pseudomonas gessardii DSM 17152
10	1539	98	37	4286	34	51	89 Pseudomonas fluorescens FHS
11	1441	103	39	4113	39	52	83 Pseudomonas gessardii BS2982
12	4319	102	38	4088	39	51	73 Pseudomonas sp. CFT9
13	1715	100	39	3953	34	54	88 Pseudomonas sp. ICMF 8385
14	1715	100	40	3939	33	54	83 Pseudomonas fluorescens UK4
15	1715	93	37	3890	33	54	90 Pseudomonas brenneri RGC8 108
16	4007	101	37	3468	33	51	101 Pseudomonas fluorescens SS101
17	3501	101	39	3431	33	51	102 Pseudomonas sp. HMWFO11
18	3089	104	38	3402	34	51	101 Pseudomonas sp. FDUAR605 380
19	3847	104	39	3370	34	51	106 Pseudomonas fluorescens TDH5
20	2883	102	36	3339	34	49	87 Pseudomonas sp. 2995-1
21	1811	104	38	3333	34	51	103 Pseudomonas poae PpR24
22	3653	104	35	3288	36	48	72 Pseudomonas fluorescens AH4
23	1007	100	33	3283	36	48	87 Pseudomonas sp. 2588-5
24	4007	100	36	3276	33	51	103 Pseudomonas sp. Root9
25	3629	106	39	3272	34	51	106 Pseudomonas sp. MF6394
26	1811	102	36	3269	33	50	101 Pseudomonas symantha FinPuedo5
27	1715	90	37	3265	27	55	97 Pseudomonas fluorescens NCTC10392
28	3653	100	35	3262	33	50	95 Pseudomonas sp. AP42
29	3509	102	37	3255	33	50	97 Pseudomonas fluorescens PL5.4
30	4007	100	35	3206	33	50	102 Pseudomonas lactis DSM 29167
31	1525	104	37	3178	34	48	71 Pseudomonas fluorescens E24
32	3509	104	38	3176	33	49	102 Pseudomonas sp. HMWFO34
33	1411	97	35	3142	31	50	78 Pseudomonas fluorescens AS06
34	1961	102	38	3103	40	51	136 Pseudomonas fluorescens BWWK6
35	1525	101	35	3087	33	48	72 Pseudomonas fluorescens E24
36	3509	98	34	3083	32	49	95 Pseudomonas fluorescens Pn14
37	1529	104	36	3082	33	48	80 Pseudomonas fluorescens A3422A
38	1529	104	36	3082	33	48	80 Pseudomonas fluorescens G2Y
39	3653	104	36	3082	33	48	80 Pseudomonas fluorescens TDH40
40	3585	103	35	3074	31	47	77 Pseudomonas sp. LBUJ020
41	1007	104	36	3040	33	48	96 Pseudomonas sp. MYb193
42	1007	104	38	3011	40	50	136 Pseudomonas azotoformans SC 14
43	1961	104	38	3011	40	50	136 Pseudomonas fluorescens 74475
44	1525	104	38	3011	40	50	136 Pseudomonas fluorescens 74954
45	3653	102	36	3001	31	49	81 Pseudomonas fluorescens 80446/1
46	3653	102	36	3001	31	49	81 Pseudomonas fluorescens ML11A
47	3653	102	36	3001	31	49	81 Pseudomonas sp. FW300-E2
48	3653	102	36	3001	31	49	81 Pseudomonas sp. FW305-124
49	3237	102	36	3001	31	49	81 Pseudomonas sp. GL93
50	3653	102	36	3001	31	49	81 Pseudomonas sp. GW247-3R2A

Organisms colored by clade

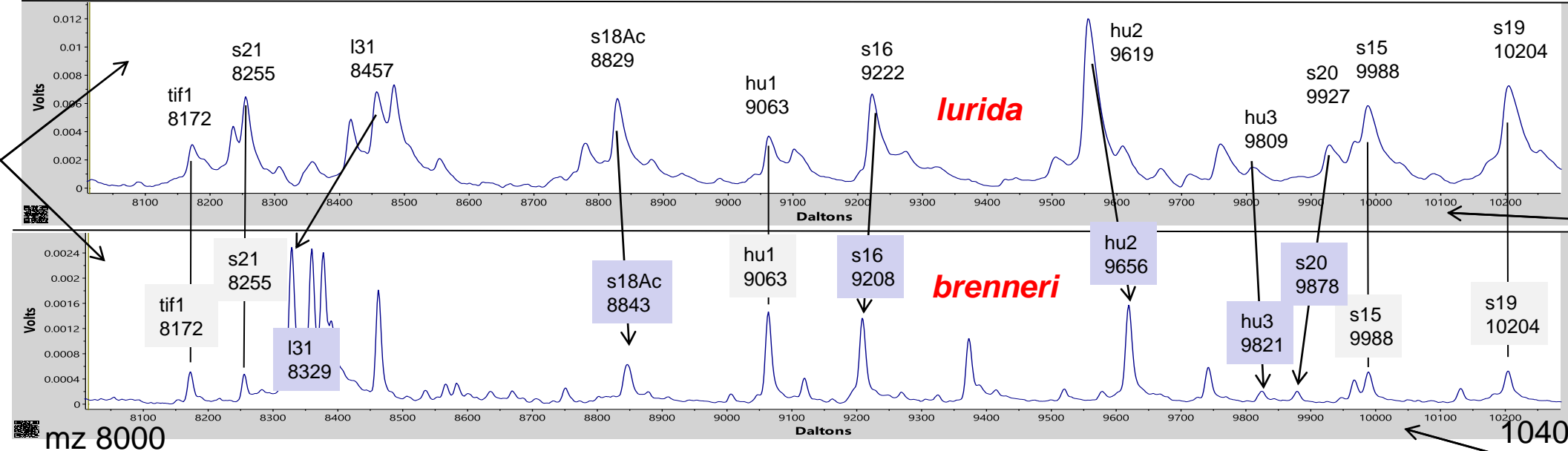


Table 3. Ribosomal Xtab for 12 isolates. Ribosomal masses and intensity ranks shown. Proteins in green seen in >=8 of 12 isolates. Colored numbers indicate identical masses. Black numbers are invariant among these 12. Colored cells are from 2<sup>nd</sup> clade of organisms: Examples of potential 'hybrids'.

OID	Clade	Species	1	2	3	4	5	6	7	8	9	10	11	12	
1	1071	1715	brenneri	63	33	27	25	25	25	20	18	18	18	6	2
2	1525	1525	fluorescens 1	33	64	36	31	37	25	21	20	18	21	6	2
3	1715	1861	lurida	27	36	64	37	33	25	20	23	22	21	6	2
4	4337	1819	libanensis	25	31	37	61	34	20	21	21	18	19	4	1
5	2871	2087	poae 1	25	37	33	34	60	21	19	18	16	16	4	1
6	1827	1507	lini	25	25	25	20	21	62	21	19	19	21	6	2
7	1441	3981	asplenii 1	20	21	20	21	19	21	64	34	33	18	5	2
8	1537	1791	koreensis 1	18	20	23	21	18	19	34	64	41	20	6	2
9	1539	3117	fluorescens 2	18	18	22	18	16	19	33	41	64	20	6	2
10	1717	1717	gingeri 1	18	21	21	19	16	21	18	20	20	63	6	2
11	3217	3217	putida 1	6	6	6	4	4	6	5	6	6	6	65	2
12	4319	4319	stutzeri 1	2	2	2	2	1	1	2	2	2	2	2	61

Small shifts in mass between homologous proteins is typical for any pair of related organisms. Black lines are conserved proteins between these two organisms. Tables 6 and 7 tabulate # of sequence identities to the top scoring organism for these spectra.

## Purpose of this case study:

- To determine how well MALDI is able to map environmental isolates to the > 40,000 genomes of sequenced bacteria:
- Given that many species have not been sequenced:
- Each newly sequence genome ought to be placed in context (into clades).

Table 4. Sequences shared between clades

- Ribosomal proteins shared between each of the 12 organisms.
- If all organisms in at least one species belong to the clade:
  - Species name is listed.
- If the species in the clade are polyphyletic:
  - The species is followed by a number.

OID	Clade	Species	1	2	3	4	5	6	7	8	9	10	11	12	
1	1071	1715	brenneri	63	33	27	25	25	25	20	18	18	18	6	2
2	1525	1525	fluorescens 1	33	64	36	31	37	25	21	20	18	21	6	2
3	1715	1861	lurida	27	36	64	37	33	25	20	23	22	21	6	2
4	4337	1819	libanensis	25	31	37	61	34	20	21	21	18	19	4	1
5	2871	2087	poae 1	25	37	33	34	60	21	19	18	16	16	4	1
6	1827	1507	lini	25	25	25	20	21	62	21	19	19	21	6	2
7	1441	3981	asplenii 1	20	21	20	21	19	21	64	34	33	18	5	2
8	1537	1791	koreensis 1	18	20	23	21	18	19	34	64	41	20	6	2
9	1539	3117	fluorescens 2	18	18	22	18	16	19	33	41	64	20	6	2
10	1717	1717	gingeri 1	18	21	21	19	16	21	18	20	20	63	6	2
11	3217	3217	putida 1	6	6	6	4	4	6	5	6	6	6	65	2
12	4319	4319	stutzeri 1	2	2	2	2	1	1	2	2	2	2	2	61

Table 6. # of Sequences shared between top scoring organisms, sorted by 1st, 2nd, 3rd ... columns.

OID	Clade	Score	Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
1	1855	1861	7239 P. lurida LMG 21995	63	61	61	61	61	61	61	61	61	61	61	53	53	53	50	50	50	50	50
2	1557	1861	6601 P. fluorescens MYb12	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
3	1555	1861	6601 P. fluorescens MYb17	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
4	1553	1861	6601 P. fluorescens MYb22	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
5	1861	1861	6601 P. lurida MYb1	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
6	1857	1861	6601 P. lurida MYb11	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
7	1859	1861	6601 P. lurida MYb16	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
8	3715	1861	6601 P. sp. MYb13	61	64	64	64	64	64	64	64	64	64	64	54	54	54	51	51	51	51	51
9	1513	1861	5742 P. fluorescens L228	60	60	60	60	60	60	60	60	60	60	62	53	53	53	51	50	50	50	48
10	1005	1005	4650 P. azotoformans F77	53	54	54	54	54	54	54	54	53	64	63	63	52	52	52	52	52	52	50
11	4013	1005	5237 P. sp. S09F 262	53	54	54	54	54	54	54	54	53	63	64	64	52	52	52	52	52	52	50
12	4017	1005	5237 P. sp. S10E 269	53	54	54	54	54	54	54	54	53	63	64	64	52	52	52	52	52		