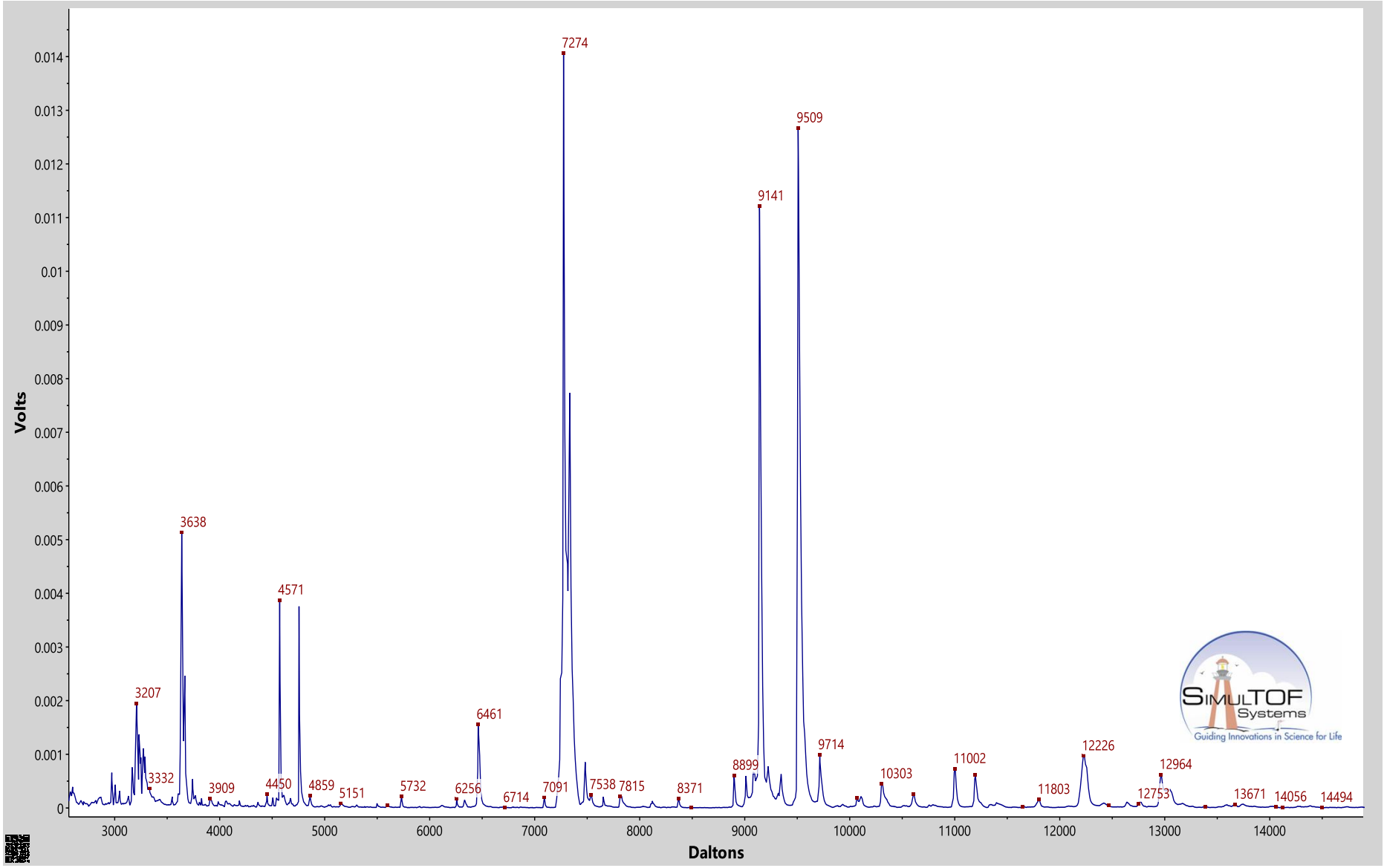
Bacterial ID by MALDI using DNA Databases

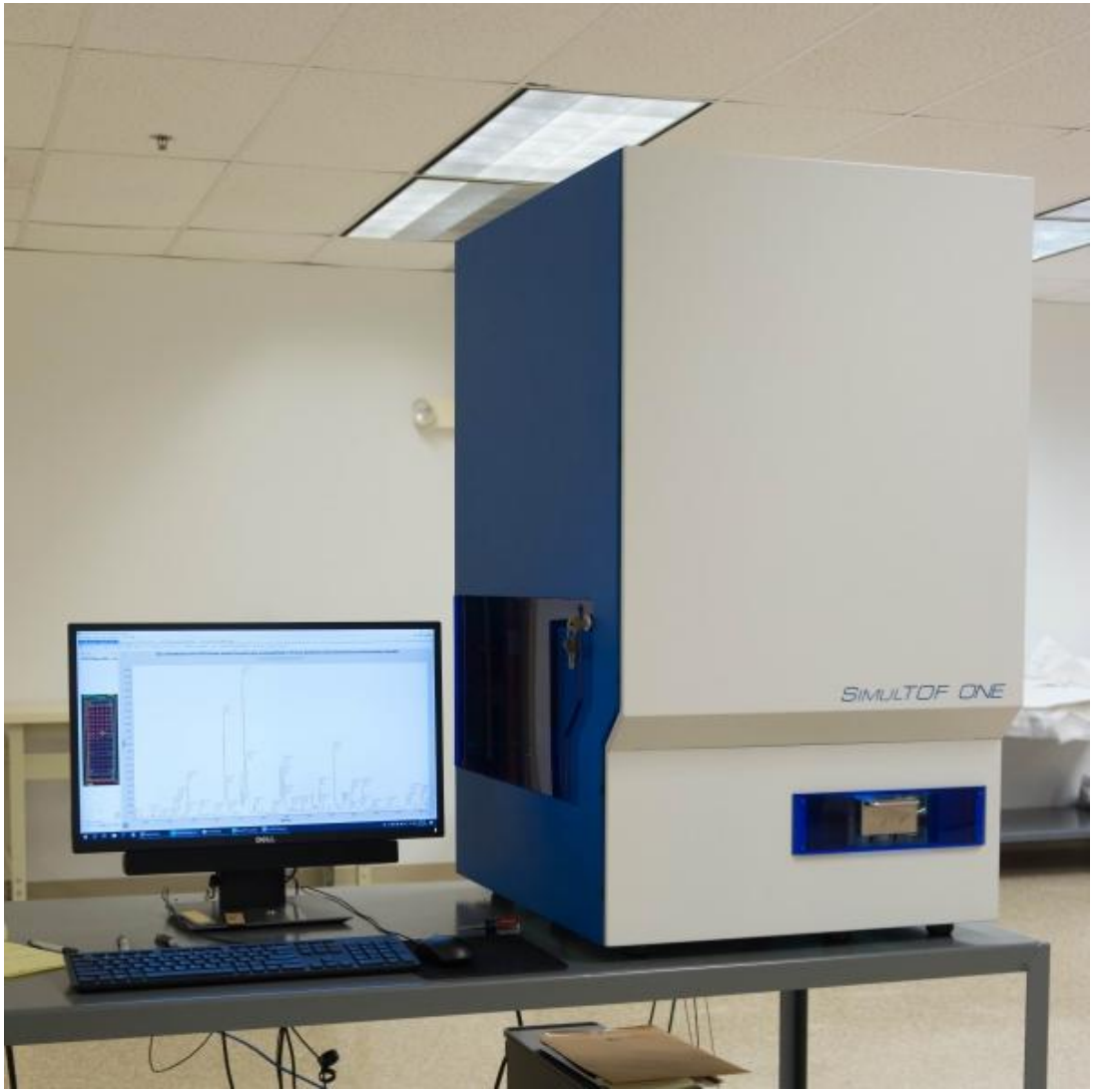
Virgin Instruments / SimulTOF Systems
Marlborough, MA

Ken Parker
Kenneth.parker@simultof.com

SimulTOF Systems
Guiding Innovations in Science for Life

MALDI spectrum of colony from Long Pond, Harwich Well D2; 10200 shots; mz 3000-15000 shown.

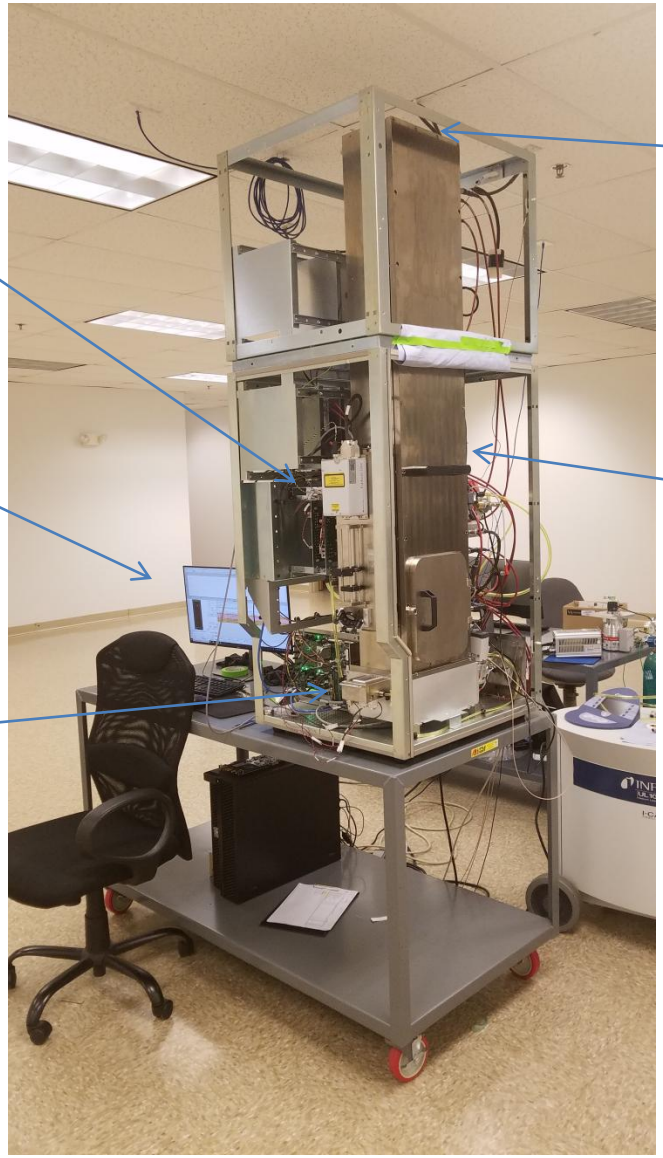




SimulTOF One with Skins

www.simultof.com

SimulTOF Systems
Guiding Innovations in Science for Life



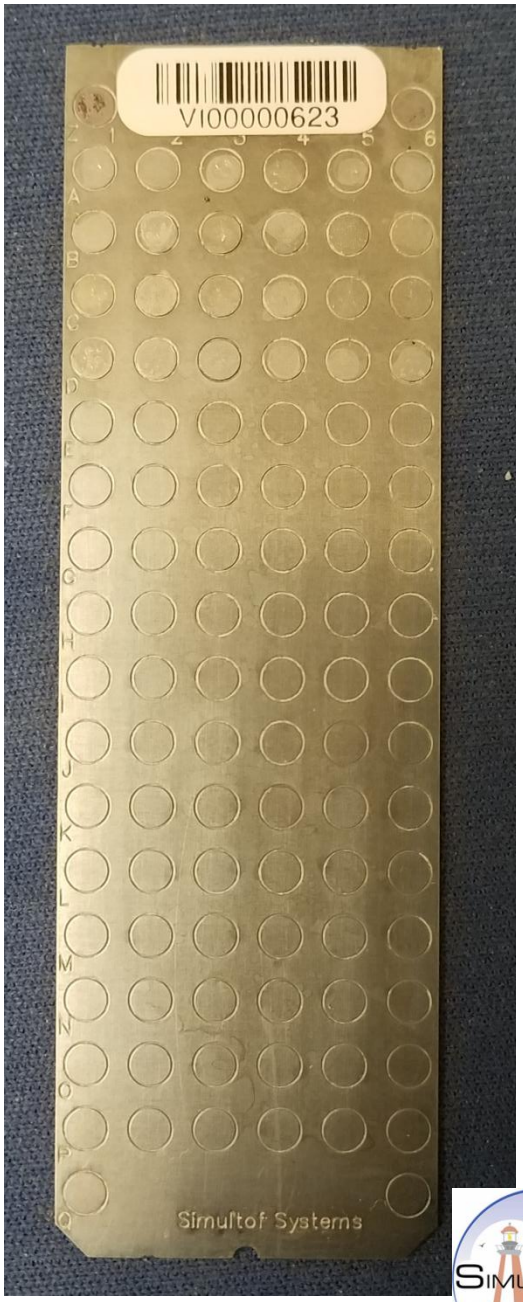
SimulTOF-200; partially functional

SimulTOF Systems
Guiding Innovations in Science for Life

A MALDI plate

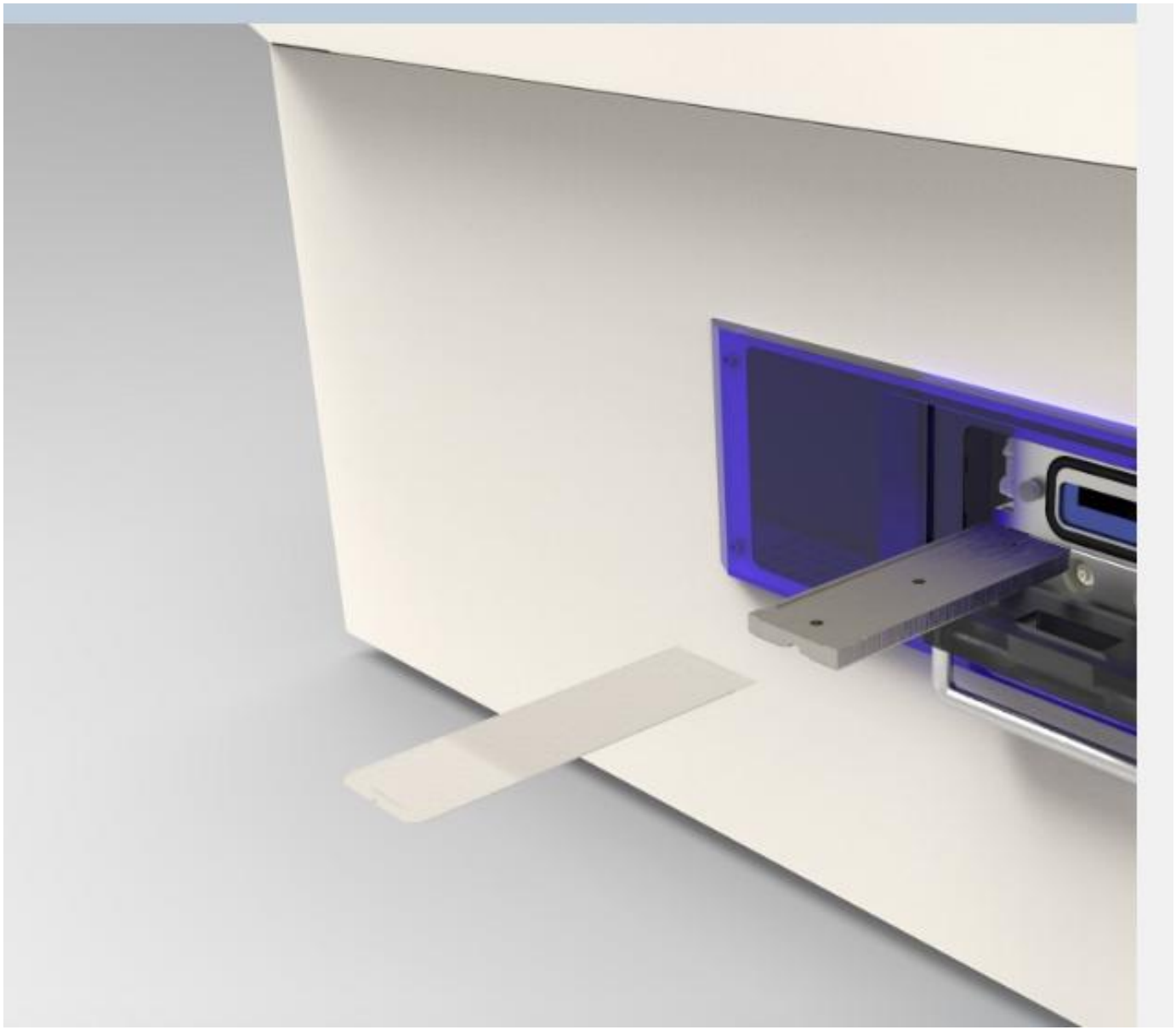
Room for
6 x16 samples
4 standards
at the corners

Dimensions:
26 x 75 mm



SimulTOF Systems
Guiding Innovations in Science for Life

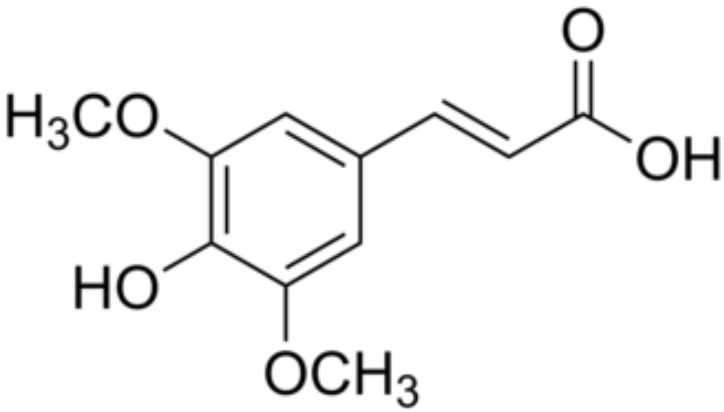
Loading a MALDI Sample Plate

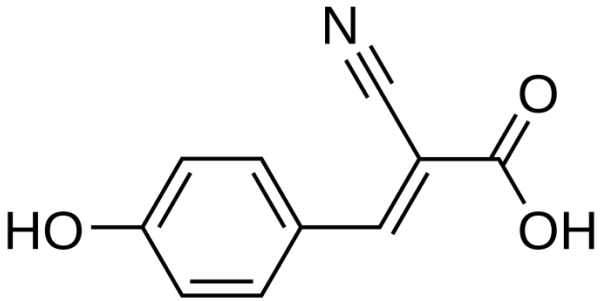


SimulTOF Systems
Guiding Innovations in Science for Life

MALDI –TOF

Matrix-assisted Laser Desorption Ionization Time-of Flight





Sinapinic acid **α-Cyano-4-hydroxycinnamic acid**

Two common matrix molecules-
that absorb laser light at 349 nm

SimulTOF Systems
Guiding Innovations in Science for Life

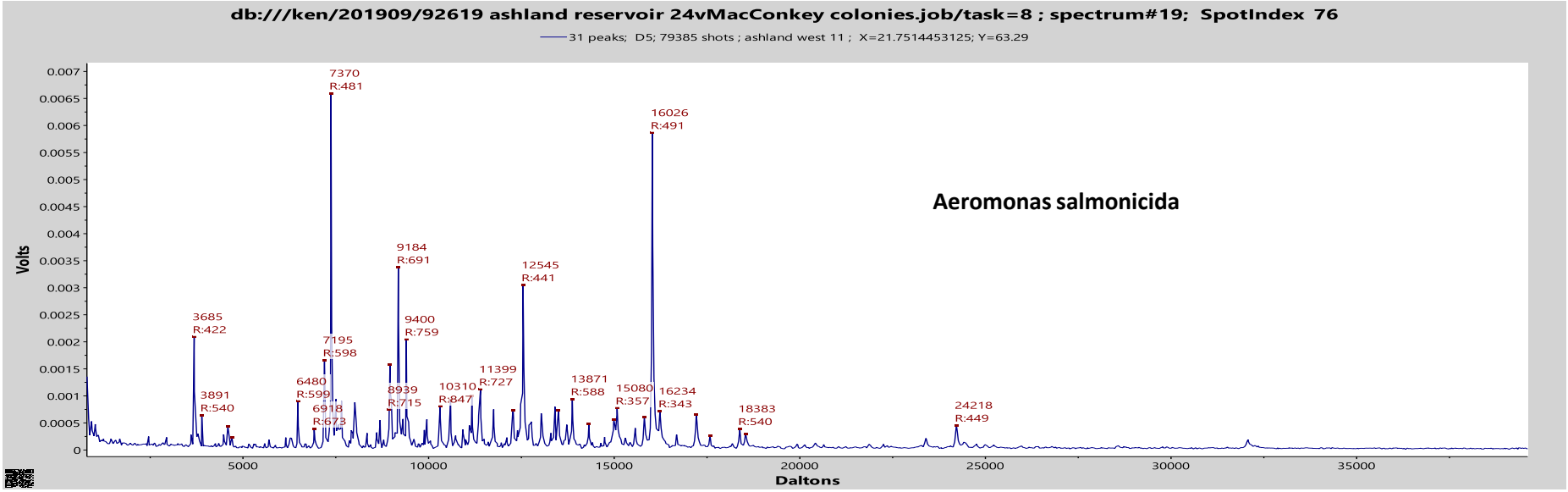
Molecules are like golf balls, baseballs, and basketballs.

Which one could David Ortiz hit the farthest?

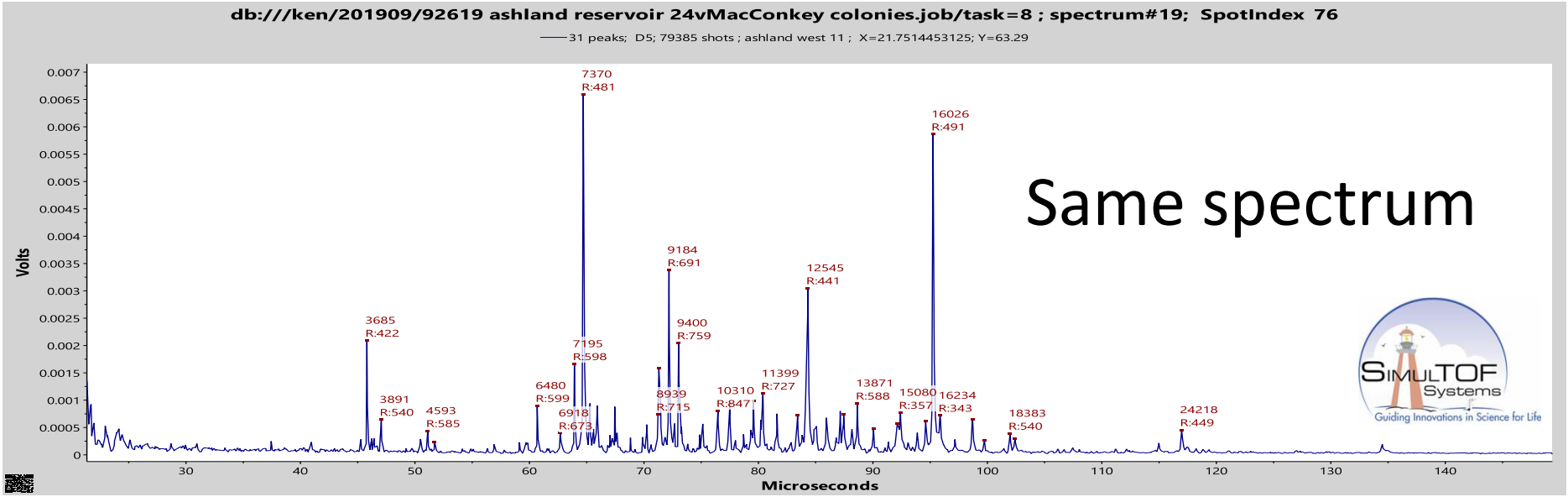
Supposing he was on the moon?

SimulTOF Systems

Guiding Innovations in Science for Life



0 mz 40,000



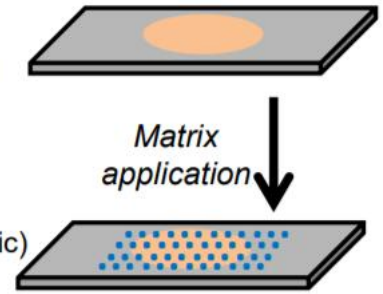
0 Microseconds 150

Imaging Mass Spectrometry

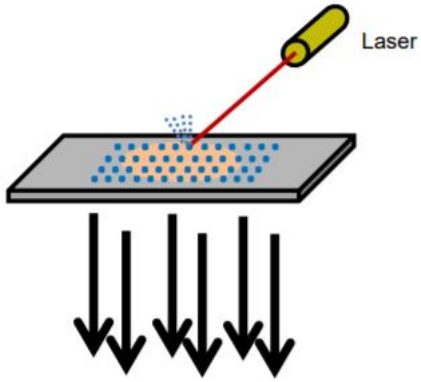
Sample Preparation

Matrix
SA – proteins/peptides
DHB –proteins/peptides
CHCA- lipids/peptides
DAN– lipids

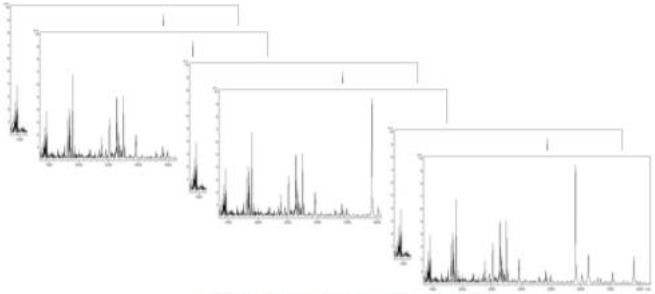
Application Method
Spraying (manual, robotic)
Sublimation
Spotting (robotic)



Matrix-Assisted Laser Desorption/Ionization

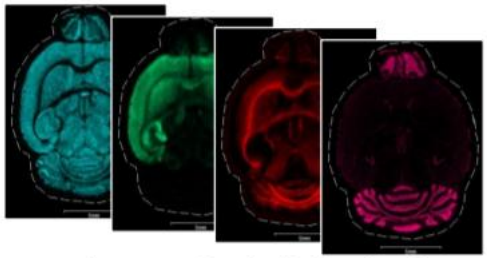


Data Processing



MS spectrum for each x, y coordinate (pixel)

Lipid & Protein Ion Images



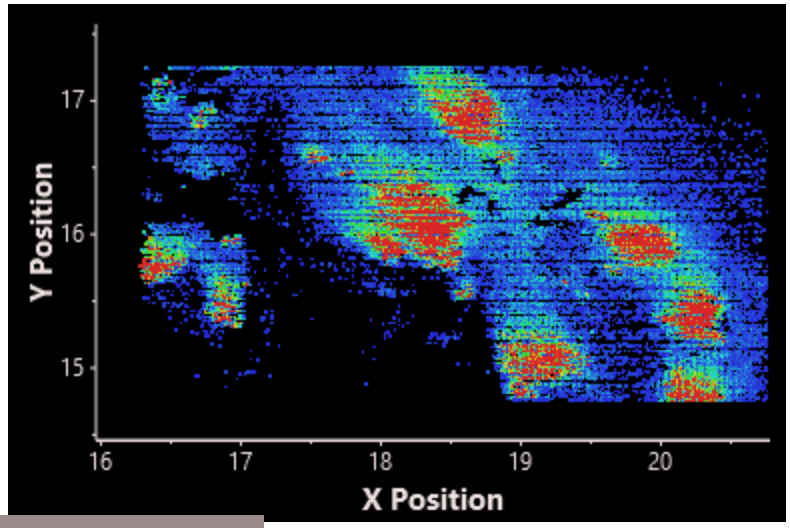
Images for individual m/z values integrated over all pixels

Boone M. Prentice[1], Kevin M. Hayden[2], Marvin L. Vestal[2], Richard M. Caprioli[1]
[1]Mass Spectrometry Research Center, Department of Biochemistry, Vanderbilt University, Nashville, TN 37235
[2]SimulTOF Systems, Marlborough, MA 01752
MSACL 2014 US March 1 - 5, 2014, San Diego, CA

SimulTOF Systems
Guiding Innovations in Science for Life

Pancreas Islet of Langerhans tissue imaging

3 x 5 mm
10 um resolution
86,700 pixels
10 laser shots/pixel
5 minutes to acquire

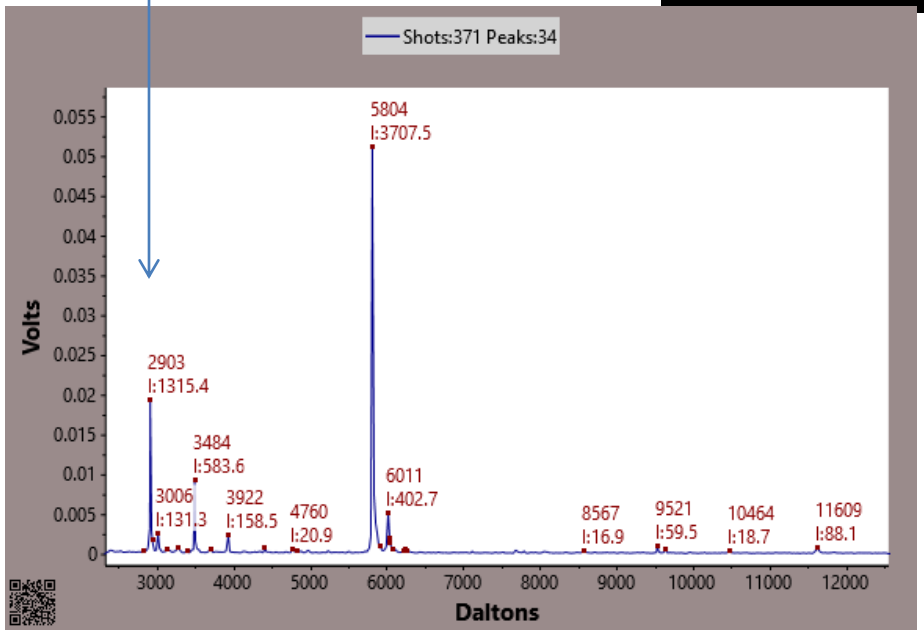


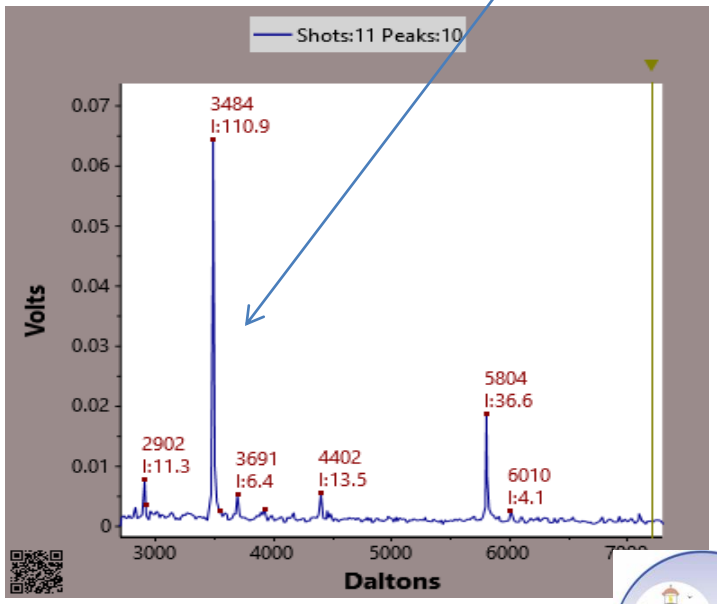
Insulin heat map

Glucagon

Insulin (2+)

Insulin





J Mass Spectrom. 2019 Jul 17. doi: 10.1002/jms.4423; Marvin Vestal et al.

SimulTOF Systems
Guiding Innovations in Science for Life

Protocol for Collecting Spectra

- Grow colony in plate
  - Or liquid culture. Need to wash out free medium
- Add 2 ul of lysing solution to MALDI plate
  - MALDI plate has 96 'wells'
  - Lysing solution -> 25 % formic acid / 25 %ethanol/ 50% water
  - Can load to lysing solution to multiple wells
- Pick colony with 20 ul micropipet tip, and resuspend in lysing solution
  - Distribute lysate as desired across multiple wells
- Wait for solution to evaporate
  - This is the slowest step!
- Overlay 2ul of sinapinic acid matrix solution or HCCA over lysed colony
- Wait for evaporation
  - This takes less time because matrix solution is more volatile
- Collect spectra (1000- to 100000 shots). It takes from a few seconds to a few minutes per well. It takes longer for a less 'ideal' extract.

SimulTOF Systems
Guiding Innovations in Science for Life

History

- Mass spec has been used for bacterial typing for decades
  - starting in the **1970**s, originally via pyrolysis of lipids
- **1996**: MALDI was used to type bacteria in protein mz range
  - ribosomal proteins correspond to many of the strongest masses
- **1997**: Pineda et al. -> matching organisms to ribosomal protein mass lists
- Since then, database look-up has largely been ignored
  - Except for selected clades (see next slide)
- Instead, FDA approved methods to identify bacteria via spectral matching
  - requires careful curation of library spectra
- In this talk, MALDI -> bacterial ID to DNA databases described
  - Want to extend MALDI market beyond pathogen ID
- Can easily search **40, 000** strains / species
  - no limit yet.

SimulTOF Systems
Guiding Innovations in Science for Life

Extract ribosomal sequences keyed to organism

Remove N-terminal methionine as necessary

Calculate singly and doubly charged masses

Save database (in memory)

Get colony

Collect spectrum

Detect peaks

Select mass range
(2.5-30K)

Find matches within tolerance

Calculate score for each organism from
% intensity matched
% proteins found
Mass accuracy

Recalibrate
Tighten
Constrain taxonomy

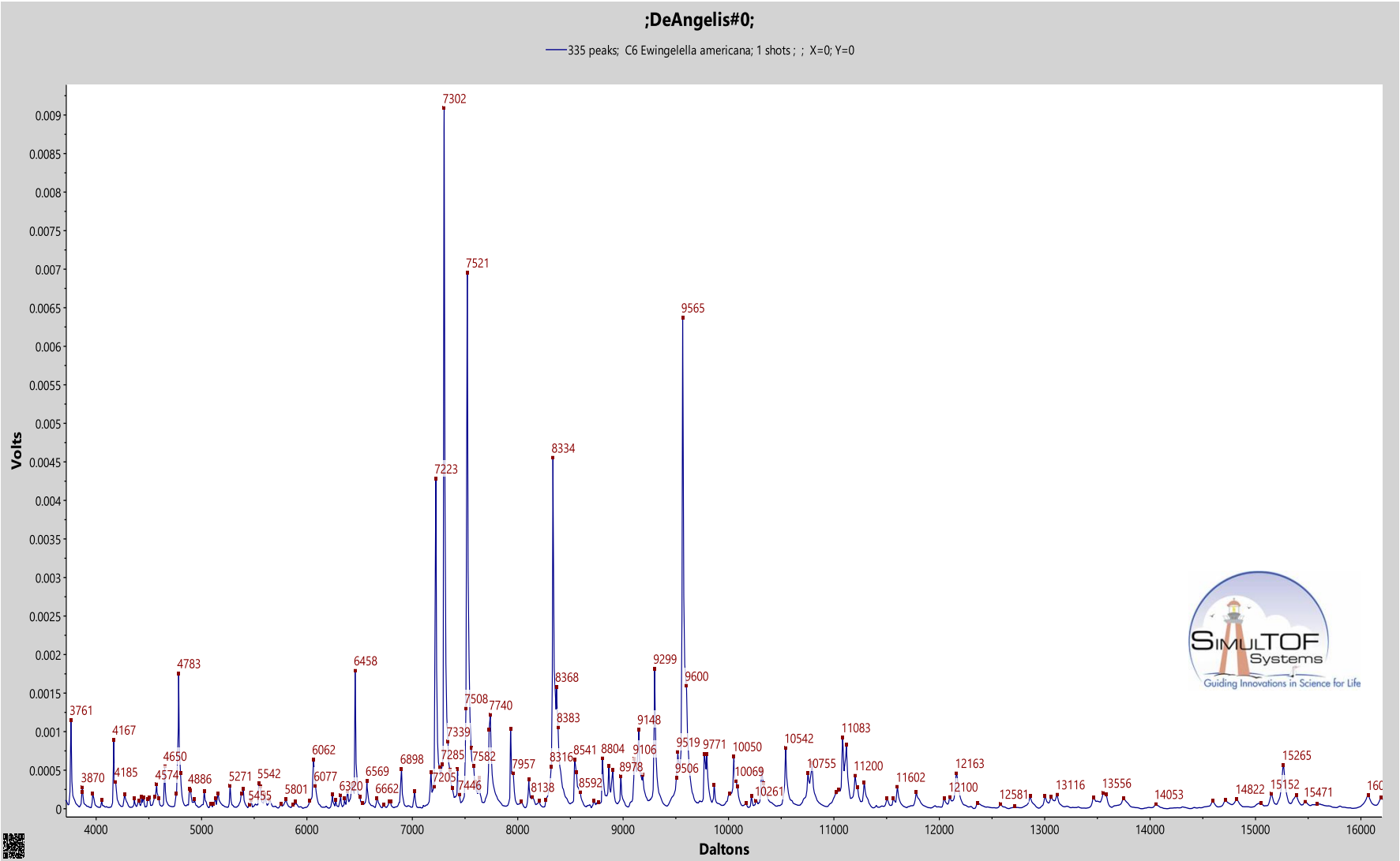
SimulTOF Systems
Guiding Innovations in Science for Life

Searching DNA databases

- Leads to understanding of species relationships
  - Returns scores of all organisms in the database (if desired)
  - Useful for exploring relationships of closely related organisms
  - Returns sequences and names for all matches
- Leads to understanding biochemistry
- Robust regarding search parameters
- Non-robust identifications indicate no closely related organism currently available at NCBI.
- Can restrict search to restricted taxa (saves time)

SimulTOF Systems
Guiding Innovations in Science for Life

MALDI spectrum of Ewingella americana



mz 3800

mz 16,100

Matches to 157 peaks from spectrum of Ewingella americana

Proteins	Matches	Score	% Intensity	% Proteins	ppm	Organism
121	74	10754	50	74	101	Ewingella americana BRK18a
116	65	7686	44	69	108	Ewingella americana ATCC
123	44	2605	28	48	95	Rahnella aquatilis HX2
122	41	1926	25	45	104	Rahnella sp. AA
117	46	1851	20	52	119	Rahnella victoriana BRK18a
121	44	1494	20	48	125	Rahnella sp. Y9602
123	34	1323	35	39	161	Serratia plymuthica HRO-C48
123	34	1323	35	39	161	Serratia plymuthica V4
125	34	1316	35	39	158	Serratia plymuthica 3Rp8
123	33	1313	35	39	156	Serratia plymuthica 3Re4-18
124	34	1301	35	39	162	Serratia plymuthica A153
125	32	1288	35	38	151	Serratia plymuthica AS9
125	32	1288	35	38	151	Serratia sp. AS12
125	32	1288	35	38	151	Serratia sp. AS13
128	34	1233	35	38	158	Serratia plymuthica tumat
123	31	1215	29	36	128	Serratia sp. 14-2641
121	32	1133	34	38	169	Serratia plymuthica S13
117	39	1121	18	45	126	Rahnella aquatilis CIP 78.65 =
121	32	1102	30	38	157	Serratia fonticola DSM 4576
114	24	1034	26	32	81	Serratia symbiotica SCt-VLC
123	22	85	8	23	188	Escherichia coli 5.0959

Top 20 matches all belong to
Proteobacteria; Gammaproteobacteria; Enterobacterales; Yersiniaceae

SimulTOF Systems
Guiding Innovations in Science for Life

Organisms most similar to Ewingella americana Brk18A

I	Proteins	Shared	Score	Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	67	67	10754	Ewingella americana	67	62	25	24	23	23	21	20	16	15	14	13	11	7	7	7	7	7	7	7
2	66	62	7686	Ewingella americana ATCC	62	66	22	21	21	20	18	19	17	13	13	12	10	5	6	5	5	5	5	5
3	68	25	1494	Rahnella sp. Y9602	25	22	68	64	41	46	48	31	42	15	14	14	12	8	7	8	9	9	8	8
4	69	24	2604	Rahnella aquatilis HX2	24	21	64	69	40	45	47	29	41	15	14	14	12	8	7	8	9	9	8	8
5	68	23	1926	Rahnella sp. AA	23	21	41	40	68	40	39	28	34	14	12	11	10	7	6	7	7	7	7	7
6	65	23	1851	Rahnella victoriana	23	20	46	45	40	65	45	32	40	14	11	12	10	8	6	8	8	8	8	8
7	66	21	1120	Rahnella aquatilis CIP	21	18	48	47	39	45	66	29	55	15	14	14	12	8	7	8	9	9	8	8
8	67	20	1011	Serratia sp. Leaf51	20	19	31	29	28	32	29	67	26	12	12	12	10	7	5	7	7	7	7	7
9	66	16	620	Rahnella aquatilis CIP	16	17	42	41	34	40	55	26	66	11	11	12	9	5	5	5	6	6	5	5
10	70	15	595	Rouxiella badensis DSM	15	13	15	15	14	14	15	12	11	70	27	28	26	7	5	7	7	7	7	7
11	69	14	542	Rouxiella silvae 213	14	13	14	14	12	11	14	12	11	27	69	65	32	7	6	7	8	8	7	7
12	67	13	542	Serratia sp. Leaf50	13	12	14	14	11	12	14	12	12	28	65	67	32	7	6	7	8	8	7	7
13	70	11	665	Serratia sp. M24T3	11	10	12	12	10	10	12	10	9	26	32	32	70	6	6	6	7	7	6	6
14	67	7	316	Serratia fonticola GS2	7	5	8	8	7	8	8	7	5	7	7	7	6	67	15	55	56	64	53	59
15	69	7	193	Serratia sp. SCBI	7	6	7	7	6	6	7	5	5	5	6	6	6	15	69	14	15	16	14	15
16	67	7	1214	Serratia sp. 14-2641	7	5	8	8	7	8	8	7	5	7	7	7	6	55	14	67	63	56	60	56
17	71	7	726	Serratia fonticola 5l	7	5	9	9	7	8	9	7	6	7	8	8	7	56	15	63	71	60	58	57
18	70	7	303	Serratia fonticola BWK15	7	5	9	9	7	8	9	7	6	7	8	8	7	64	16	56	60	70	54	61
19	66	7	1102	Serratia fonticola DSM	7	5	8	8	7	8	8	7	5	7	7	7	6	53	14	60	58	54	66	56
20	66	7	310	Serratia fonticola AeS1	7	5	8	8	7	8	8	7	5	7	7	7	6	59	15	56	57	61	56	66

SimulTOF Systems
Guiding Innovations in Science for Life

Matches to Sodalis 159R
No other organisms even close

Peaks	Protein	Matches	Score	%I	%Ch	ppm	Org
112	115	66	11157	63	68	105	Sodalis sp. 159R
112	114	23	684	28	27	121	Sodalis praecaptivus HS1
112	117	20	603	28	24	108	Candidatus Sodalis pierantonius str. SOPE
112	110	19	402	19	24	111	Lonsdalea populi CFCC 12252
112	120	23	400	36	26	240	Pectobacterium punjabense SS95
112	120	25	366	31	25	219	Pectobacterium atrosepticum 36A
112	120	25	366	31	25	219	Pectobacterium atrosepticum SS26
112	111	17	365	24	20	51	Sodalis sp. TME1
112	127	21	364	30	23	170	Pectobacterium carotovorum s0421
112	110	18	345	24	22	134	Dickeya dianthicola ME23
112	122	21	342	34	23	210	Pectobacterium carotovorum subsp. carotovorum 67
112	120	20	330	34	22	209	Pectobacterium carotovorum S1.16.01.3K
112	120	20	330	34	22	209	Pectobacterium carotovorum S1-A16
112	120	20	330	34	22	209	Pectobacterium carotovorum S4.16.03.1C
112	120	20	330	34	22	209	Pectobacterium carotovorum subsp. carotovorum ATCC 39048
112	116	19	325	34	22	209	Pectobacterium carotovorum subsp. carotovorum PCC21
112	124	20	311	34	22	209	Pectobacterium carotovorum subsp. brasiliense BZA12
112	119	19	310	34	22	209	Pectobacterium carotovorum subsp. brasiliensis ICMP 19477
112	120	19	306	33	22	206	Pectobacterium carotovorum S1.15.11.2D
112	115	16	305	23	20	119	Dickeya fangzhongdai PA1

SimulTOF Systems
Guiding Innovations in Science for Life

Only 8 ‘hot’ proteins shared between Sodalis sp. 159R and any of the 235 genomes of the Pectobacteriaceae available on 9-24-2019

ClusterViewer; Cladify: 236 organisms related to Sodalis sp. 159R; peaks 300; tol 500; autocal True; 0.6876703 seconds

Cluster centered on organism selected

	I	OID	D	C	M	MF	Sc	Organism	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P2
1	1	447	0	0	64	64	11157	Sodalis sp. 159R	64	8	8	6	6	6	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
2	2	445	0	0	66	8	683	Sodalis praecaptivus HS1	8	66	26	35	20	22	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
3	3	431	0	0	66	8	603	Candidatus Sodalis pierantonius str. SOPE	8	26	66	14	11	12	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4	4	449	0	0	60	6	365	Sodalis sp. TME1	6	35	14	60	16	16	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5	5	443	0	0	64	6	93	Sodalis glossinidius str. 'morsitans' morsit	6	20	11	16	64	59	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
6	6	441	0	0	63	6	97	Sodalis glossinidius str. 'morsitans' morsit	6	22	12	16	59	63	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
7	7	341	0	0	67	5	210	Lonsdalea britannica 477	5	4	4	3	3	4	67	41	44	43	44	41	43	42	44	43	44	42	44	44	44	43	42
8	8	401	0	0	64	5	206	Lonsdalea quercina ATCC 29281	5	3	3	3	3	3	41	64	41	41	41	39	41	40	42	41	41	41	41	41	42	41	41
9	9	377	0	0	61	5	242	Lonsdalea populi CFCC 11265	5	4	4	3	3	4	44	41	61	58	57	57	58	57	58	60	61	53	59	59	58	60	59
10	10	385	0	0	61	5	242	Lonsdalea populi CFCC 11200	5	4	4	3	3	4	43	41	58	61	58	56	61	59	60	59	58	56	61	61	60	59	58
11	11	367	0	0	63	5	213	Lonsdalea populi CFCC 13126	5	4	4	3	3	4	44	41	57	58	63	54	58	56	60	57	57	56	59	59	60	57	56
12	12	351	0	0	62	5	237	Lonsdalea populi CFCC 13800	5	4	4	3	3	4	41	39	57	56	54	62	56	55	55	58	57	52	56	56	55	58	57
13	13	371	0	0	61	5	242	Lonsdalea populi CFCC 11379	5	4	4	3	3	4	43	41	58	61	58	56	61	59	60	59	58	56	61	61	60	59	58
14	14	397	0	0	60	5	242	Lonsdalea quercina CFCC 13731	5	4	4	3	3	4	42	40	57	59	56	55	59	60	58	58	57	54	59	59	58	58	57
15	15	347	0	0	62	5	220	Lonsdalea populi CFCC 11768	5	4	4	3	3	4	44	42	58	60	60	55	60	58	62	58	58	57	61	61	62	58	57
16	16	379	0	0	61	5	242	Lonsdalea populi CFCC 12721	5	4	4	3	3	4	43	41	60	59	57	58	59	58	58	61	60	54	59	59	58	61	60
17	17	383	0	0	61	5	242	Lonsdalea populi CFCC 12786	5	4	4	3	3	4	44	41	61	58	57	57	58	57	58	60	61	53	59	59	58	60	59
18	18	373	0	0	59	5	240	Lonsdalea populi CFCC 11263	5	4	4	3	3	4	42	41	53	56	56	52	56	54	57	54	53	59	56	56	57	54	53
19	19	363	0	0	62	5	237	Lonsdalea populi CFCC 12509	5	4	4	3	3	4	44	41	59	61	59	56	61	59	61	59	59	56	62	62	61	59	58
20	20	387	0	0	62	5	237	Lonsdalea populi CFCC 13097	5	4	4	3	3	4	44	41	59	61	59	56	61	59	61	59	59	56	62	62	61	59	58
21	21	369	0	0	62	5	220	Lonsdalea populi CFCC 11196	5	4	4	3	3	4	44	42	58	60	60	55	60	58	62	58	58	57	61	61	62	58	57
22	22	381	0	0	61	5	242	Lonsdalea populi CFCC 13912	5	4	4	3	3	4	43	41	60	59	57	58	59	58	58	61	60	54	59	59	58	61	60

Proving the methodology

- Can identify many colonies from the environment to species
- The method will fail if the organism is not in the database
- Genomic matching can be tested using any peak list (ideally with intensity)
  - Calibration is important
- Genomic matching can also be tested in silico
  - Generate a ribosomal mass list
  - Can now reduce mass tolerance to nearly zero.
- Challenge me with a bacterial ID problem!

SimulTOF Systems
Guiding Innovations in Science for Life

Biological confusing factors

- Some bacteria are harder to lyse so as to release ribosomal proteins
- Not all ribosomal proteins are detected
- Probably some differential degradation
- Probably at lag phase ribosomes are less prominent in the proteome

SimulTOF Systems
Guiding Innovations in Science for Life

Informatic Confounding Factors

- Some genomes have mistakes in ribosomal proteins
  - Some subunits entirely missing in some proteomes
    - In E coli and B. subtilis, only ~ ½ of ribosomal proteins are necessary for growth
  - N-terminal extensions or deletions in the databases
  - Some sequences are mis-annotated as ribosomal subunits
- Some ribosomal proteins are present in multiple forms, at least for some clades
  - L31 has a special zinc-deficient form
- Some proteins are modified
  - L33 -> methylated in many gammaproteobacteria
- Everything works better without these problems

SimulTOF Systems
Guiding Innovations in Science for Life

Bacterial species and MALDI

- Species originally defined by
  - Disease
  - microscopy
  - growth requirements / metabolism
  - ecological niche
- New consensus: genomic similarity -> proteomic similarity
  - Vertical vs Horizontal evolution
- Simplification: Substantial Ribosomal protein sequence identity
  - Highly conserved throughout prokaryotic evolution, transmitted vertically
  - 54 subunits typically
  - Complex nearly 1:1
- MALDI recognizes small, abundant, basic proteins like ribosomal subunits
- Therefore, MALDI can distinguish species
  - If organisms cannot be distinguished by MALDI, then they do not belong in different species
  - Limitations of species delineation by MALDI due to errors in species definition

SimulTOF Systems
Guiding Innovations in Science for Life

Conclusions

- MALDI can identify sequenced bacteria
- Limitations in distinguishing bacteria largely due to taxonomic confusion
  - Mixing data from strains across true taxonomic clades confounds organism identification
  - Incorrect species annotations a problem at NCBI this year
  - Would work better if ribosomal proteins were annotated more reproducibly

Collaborators

- Jean-Philippe Charrier, Biomerieux
- Joe Zhou, Each-Reach, China
- Kent Voorhees and Chris Cox, Colorado School of Mines
- Martin Polz, MIT
- Kristen DeAngelis, Umass Amherst; Hayley Green
- SimulTOF Systems

References

- Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M, Lin JS. (2003) Microorganism identification by matrix-assisted laser/desorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. Anal Chem. 75:3817-22.
- e.g. http://www.ncbi.nlm.nih.gov/genome/?term=*Shigella+boydii*

SimulTOF Systems
Guiding Innovations in Science for Life