PNAS

Supporting Information for:

A robust, agnostic molecular biosignature based on machine learning

H. James Cleaves II, Grethe Hystad, Anirudh Prabhu, Michael L. Wong,

George D. Cody, Sophia Economon, and Robert M. Hazen

Corresponding Author: Robert M. Hazen

Email: rhazen@ciw.edu

This PDF file includes: Table S1

Table S1. Tabulation of 134 samples analyzed in this study.

Sample Number	Sample Name	Sample Description	Age (Ma)	Classification
1	ADGCONT3d.txt	Alkaline degraded β-D-glucose	0	A
2	AKG3d.txt	Commercial α-keto glutarate	0	A
3	Ala333.txt	Commercial tri-L-alanine	0	A
4	AlaMR3d.txt	Maillard reaction of DL Alanine and β-D-glucose	0	A
5	Alanine3d.txt	Commercial DL Alanine	0	A
6	Albertit3d.txt	Albertite. Asphalt like substance from Albert County, New Brunswick, Canada	359-419	B
7	Allende23d.txt	Allende Meteorite. Meteorite seen to fall near Pueblito de Allende, Mexico Feb. 1 1969. Kind gift of Professor Alan Schwartz.	~4500	A
8	Allende3d.txt	Allende Meteorite. Meteorite seen to fall near Pueblito de Allende, Mexico Feb. 1 1969. Kind gift of Professor Alan Schwartz.	~4500	A
9	AMAX.3d.txt	AMAX coal	ND	B
10	ASPHLTMM3d.txt	Asphaltum, Trinidad. Black, highly viscous liquid or semi-solid form of petroleum, commonly referred to as asphalt.	ND	B
11	ASPMR.3d.txt	Maillard reaction of DL-aspartic acid reacted with β-D-glucose	0	A
12	BARLEY.3d.txt	*Hordeum vulgare*, Barley kernal	0	B
13	BASMATI.3d.txt	*Oryza sativa*, Basmati rice grain	0	B
14	BBRsoil1.3d.txt	Soil from Broad Branch creek area 38°57'28.1"N 77°03'48.0"W	0	B
15	Bcereus3d.txt	*Bacillus cereus*. Ward's scientific. Gram-positive rod-shaped bacterium commonly found in soil, food, and marine sponges. *B. cereus* bacteria may be anaerobes or facultative anaerobes	0	B

16	Biofil.3d.txt	Biofilm from Broad Brach 38°57'27.0"N 77°03'45.9"W	0	B
17	Bonnetf3d.txt	Common Bonnet Fungus (cap, basidiomycete), *Mycena galericulata*	0	B
18	brnrice3d.txt	*Oryza sativa*, Brown rice grain,	0	B
19	Bsubtil3d.txt	*Bacillus subtilis* cells, Ward's Scientific	0	B
20	Bulgur3d.txt	*Triticum durum*, Bulgur wheat kernal	0	B
21	Cannak3d.txt	Cannak meteorite	~4500	A
22	Cannelco3d.txt	Cannel Coal. Cannel coal or candle coal is a type of bituminous, pollen-derived coal.	ND	B
23	Carnos3d.txt	Carnarum Creek Oil Shale, Australia	ND	B
24	Cedarto3d.txt	Cedar wood t0 from Cody Fogel expt.	0	B
25	Chiles3d.txt	*Capsicum chinense*, Habanero chile seed	0	B
26	Chiltin3d.txt	Chitin. Chitin is a long-chain polymer of N-acetylglucosamine. Chitin is probably the second most abundant polysaccharide in nature (after cellulose); an estimated 1 billion tons of chitin are produced each year in the biosphere It is a primary component of cell walls in fungi, the exoskeletons of arthropods, the radulae, cephalopod beaks and gladii of molluscs and in some nematodes and diatoms.	0	B
27	Citron3d.txt	Citronella candle	0	B
28	CIWSoil13d.txt	Soil from EPL campus	0	B
29	Collagen3d.txt	Commercial collagen. Collagen is the main structural protein in the extracellular matrix found in the human connective tissues. As the main component of connective tissue, it is the most abundant protein in mammals, making up from 25% to 35% of the whole-body protein content.	0	B
30	CRDPAR3d.txt	*Stereum complicatum*, “Crowded parchment” wood fungus from Rock Creek Park	0	B

31	Cyanob23d.txt	Mixed freshwater cyanobacteria, Ward's Scientific	0	B
32	Cyanob3d.txt	Mixed freshwater cyanobacteria, Ward's Scientific	0	B
33	CysMR23d.txt	Maillard reaction of DL-cysteine reacted with β-D glucose	0	A
34	CysMR3d.txt	Maillard reaction of DL-cysteine reacted with β -D glucose	0	A
35	Devons3d.txt	Devonian Shale	60-419	B
36	DGlucose3d.txt	β-D-glucose, Sigma Aldrich	0	A
37	DNA3d.txt	Commercial DNA (from *Saccharomyces cerevisiae*)	0	A
38	ECO33DB.txt	*Escherichia coli*, Ward's Scientific	0	B
39	Ecoli13d.txt	*Escherichia coli*, Ward's Scientific	0	B
40	ED513933d.txt	Miller Urey experiment	0	A
41	Elaterit3d.txt	Elaterite. Also known as Aeonite, 'elastic bitumen', 'mineral caoutchouc' or Wurtzilite) is a brown hydrocarbon varying somewhat in consistency, being sometimes soft, elastic and sticky, like India rubber, and occasionally hard and brittle.	ND	B
42	Entaer3d.txt	*Klebsiella aerogenes*, previously known as *Enterobacter aerogenes*, is a Gram-negative, rod-shaped bacterium.	0	B
43	Ferron3d.txt	Ferron Coalified Wood, Ferron Coal Bed, Utah	100.5-66	B
44	GCN3d.txt	Commercial glycolonitrile, Sigma Aldrich	0	A
45	Gelatin3d.txt	Commercial gelatin. Gelatin is collagen that has been irreversibly hydrolyzed.	0	B
46	GlnMR3d.txt	Maillard reaction of DL-glutamine reacted with β-D-glucose	0	A
47	GluMR23d.txt	Maillard reaction of DL-glutamic acid reacted with β-D-glucose	0	A
48	GluMR3d.txt	Maillard reaction of DL-glutamic acid reacted with β-D-glucose	0	A

49	Glycine3d.txt	Commercial glycine, Sigma Aldrich	0	A
50	GlyMR3d.txt	Maillard reaction of glycine reacted with β-D-glucose	0	A
51	Glyoxaci3d.txt	Glyoxylic acid, Sigma Aldrich	0	A
52	Grass3d.txt	Common bluegrass, *Poa pratensis*, from CIW lawn	0	B
53	Grnrivs3d.txt	Green River Shale. The Green River Formation is an Eocene geologic formation that records the sedimentation in a group of intermountain lakes in three basins along the present-day Green River in Colorado, Wyoming, and Utah.	34-56	B
54	Hair3d.txt	Human hair	0	B
55	HCN531723d.txt	HCN polymer Reaction started May 31, 1972, Stanley L. Miller collection	0	A
56	HMT3d.txt	Commercial Hexamethylenetetramine Sigma Aldrich	0	A
57	Humica3d.txt	Humic Acid, Mallinckrodt	0	B
58	Jetwhtby3d.txt	Jet from Whitby, UK. Jet is a product of decomposition of wood from millions of years ago, commonly the wood of trees of the family *Araucariaceae*.	180	B
59	Kraton23d.txt	Kraton polymer. Kraton polymers are styrenic block copolymers (SBC) consisting of polystyrene blocks and rubber blocks. The rubber blocks consist of polybutadiene, polyisoprene, or their hydrogenated equivalents.	0	A
60	Kraton3d.txt	See above	0	A
61	LAsp3d.txt A	Commercial L-Aspartic Acid	0	A
62	Lcys3d.txt	Commercial L-Cysteine	0	A

63	Lescel3d.txt	*Ficaria verna* (formerly *Ranunculus ficaria L.*), commonly known as lesser celandine or pilewort, is a low-growing, hairless perennial flowering plant in the buttercup family *Ranunculaceae* native to Europe and Western Asia. It is an invasive species in North America, where it is known by the common name fig buttercup.	0	B
64	Leuglulq3d.txt	Maillard reaction of DL-leucine reacted with β-D-glucose, liquid fraction	0	A
65	Leuglumr3d.txt	Maillard reaction of DL-leucine reacted with β-D-glucose, solid fraction	0	A
66	Lgln3d.txt	Commercial L-glutamine	0	A
67	Lglu3d.txt	Commercial L-glutamic acid	0	A
68	Lhis3d.txt	Commercial L-histidine	0	A
69	Lichen3d.txt	Unidentified lichen species from Broad Branch Road area 38°57'28.1"N 77°03'48.0"W	0	B
70	Lleu3d.txt	Commercial L-leucine	0	A
71	Llys3d.txt	Commercial L-lysine	0	A
72	Lmet3d.txt	Commercial L-methionine	0	A
73	Lphe3d.txt	Commercial L-Phenylalanine	0	A
74	Lpro3d.txt	Commercial L-proline	0	A
75	Lser3d.txt	Commercial L-Serine	0	A
76	Ltyr3d.txt	Commercial L-Tyrosine	0	A
77	Lysmr3d.txt	Maillard reaction of L-lysine reacted with β-D-glucose	0	A
78	Malic3d.txt	Commercial DL Malic Acid	0	A
79	MetMRLQ3d.txt	Maillard reaction of L-methionine reacted with β-D-glucose, liquid fraction	0	A
80	Metseqah3d.txt	Metasequoia, Sample B, from Axel Heiberg island.	45	B

81	Miclut3d.txt	*Micrococcus luteus*, Gram-positive to Gram-variable, obligate aerobe, nonmotile, tetrad-arranging, pigmented, saprotrophic coccus bacterium.	0	B
82	Moss3d.txt	Unidentified moss species from Rock Creek Park	0	B
83	Mythic3d.txt	*Mythicomyces* fungus species	0	B
84	Nrk1_RT3d.txt	Reaction of acetaldehyde, formaldehyde and ammonia conducted at 25° C	0	A
85	Nrk2_853d.txt	Reaction of acetaldehyde, formaldehyde and ammonia conducted at 85° C	0	A
86	Nrk2rt3d.txt	Reaction of acetaldehyde, formaldehyde and ammonia conducted at 25° C	0	A
87	Nrka1853d.txt	Reaction of acetaldehyde, formaldehyde and ammonia conducted at 85° C	0	A
88	Nylon3d.txt	Nylon 6,6	0	A
89	OakLeaf3d.txt	*Quercus alba,* Oak leaf	0	B
90	Oat3d.txt	*Avena sativa,* Oat	0	B
91	Okra3d.txt	*Abelmoschus esculentus*, Okra seed	0	B
92	p1165lp3d.txt	Hexamethylenetetramine (HMT) heated with Fe_2O_3	0	A
93	p1165ls3d.txt	HMT heated with Fe_2O_3	0	A
94	p118l2s3d.txt	Aqueous sodium ferrocyanide, sunlight irradiated	0	A
95	p82_30B3d.txt	Bradford Crude Petroleum; The Bradford oil field is located mainly in McKean County of northwestern Pennsylvania.	ND	B
96	p82_313d.txt	Bell Creek Petroleum, The Bell Creek Field is a Lower Cretaceous stratigraphic trap in the Muddy Sandstone	100-145	B
97	p82_31B3d.txt	Bell Creek Petroleum; The Bell Creek Field is a Lower Cretaceous stratigraphic trap in a muddy sandstone.	100-145	B
98	p82_323d.txt	Exxon 44647 Petroleum	ND	B

99	p82_403d.txt	Formose/NH_3 150° C	0	A
100	p82_473d.txt	Pyruvic Acid polymer	0	A
101	p82_513d.txt	Formose/NH_3 85° C	0	A
102	p82_523d.txt	Formose 25° C	0	A
103	p825085C3d.txt	Formose 85° C	0	A
104	p8253fr3d.txt	Formose NH_3 25° C	0	A
105	pacampa23d.txt	Polyacrylamide-co acrylic acid Ave Mw 15 million Aldrich	0	A
106	pacampac3d.txt	Polyacrylamide-co acrylic acid Ave Mw 15 million Aldrich	0	A
107	peat3d.txt	Peat, Somerset England	0.012	B
108	pemagma3d.txt	Polyethylene co-methacrylate co-glycidyl methacrylate	0	A
109	phemr3d.txt	Maillard reaction of L-phenylalanine reacted with β-D glucose	0	A
110	Pitt8Coa3d.txt	Pitt Coal, The Pittsburgh Coal Seam is the thickest and most extensive coal bed in the Appalachian Basin.	265-330	B
111	ProMR3d.txt	Maillard reaction of L-proline reacted with β-D glucose	0	A
112	Pumpkin3d.txt	Pumpkin, *Cucurbita pepo*	0	B
113	Quisqui3d.txt	Quisqueite, Peru, S-rich bitumen-like material.	ND	B
114	Rdlent3d.txt	*Lens esculenta*, Red lentil	0	B
115	RNA3d.txt	Commercial RNA from *Saccharomyces cerevisiae*	0	A
116	Scerev3d.txt	*Saccharomyces cerevesiae*, Common yeast	0	B
117	SerMar3d.txt	*Serratia marcescens*. Rod-shaped, Gram-negative bacteria in the family *Yersiniaceae*. Facultative anaerobe	0	B
118	SerMRSO3d.txt	Maillard reaction of L-serine reacted with β-D-glucose, solid fraction	0	A

119	Shungite3d.txt	Shungite, near Shunga village, Karelia, Russia. Shungite is a black and lustrous mineraloid. Stratified shungite-bearing deposits that retain sedimentary structures are interpreted as metamorphosed oil source rocks. Some diapiric structures have been identified, which are interpreted as possible mud volcanoes.	~2000	B
120	SIOM25023d.txt	Synthetic Insoluble Organic Matter prepared from dextrose, water and ammonia, heated to 250° C George Cody	0	A
121	SIOM250D3d.txt	Synthetic Insoluble Organic Matter prepared from dextrose, water and ammonia, heated to 250° C George Cody	0	A
122	Sodpyr3d.txt	Commercial sodium pyruvate, Sigma Aldrich	0	A
123	Staepi3d.txt	*Staphylococcus epidermidis*. Ward's scientific. Gram-positive, facultative anaerobic bacteria.	0	B
124	Strob3d.txt	High pressure polymer of diphenylbutadiene	0	A
125	TAGL5b3d.txt	Tagish Lake meteorite fraction 5b	~4500	A
126	TAGLIII3d.txt	Tagish Lake meteorite fraction III	~4500	A
127	TAGLIIV3d.txt	Tagish Lake meteorite fraction IIV	~4500	A
128	TAGLKIIH3d.txt	Tagish Lake meteorite fraction IIH	~4500	A
129	TAGLold3d.txt	Tagish Lake meteorite combined fractions	~4500	A
130	TglkIIH23d.txt	Tagish Lake meteorite fraction IIH	~4500	A
131	TglkIIV23d.txt	Tagish Lake meteorite fraction IIV	~4500	A
132	Torbanit3d.txt	Torbanite, near Edinburgh, Scotland. Also known as boghead coal or channel coal, is a variety of fine-grained black oil shale. Generally occurs as lenticular masses associated with Permian coal deposits. Torbanite is classified as lacustrine type oil shale.	252-299	B
133	Tyrglumr3d.txt	Maillard reaction of L-tyrosine reacted with β-D-glucose	0	A
134	Westflds3d.txt	Westfield Oil Shale, Scotland	300-360	B