Acknowledgements

- Sara Walker (ASU)
- Lee Cronin (Glasgow)
- ELIFE@ASU, Cronin Lab
- AbGradCon community!

1
00:00:07,510 --> 00:00:04,630
all right hello everyone

2
00:00:09,509 --> 00:00:07,520
my name is john levoy i'm a phd student

3
00:00:11,669 --> 00:00:09,519
at arizona state university

4
00:00:13,110 --> 00:00:11,679
um i'm also studying for the semester at

5
00:00:15,829 --> 00:00:13,120
the university of glasgow

6
00:00:16,470 --> 00:00:15,839
so i'm kind of a dual phd student right

7
00:00:18,550 --> 00:00:16,480
now

8
00:00:19,670 --> 00:00:18,560
and i'm looking at universal life

9
00:00:22,150 --> 00:00:19,680
detection

10
00:00:26,390 --> 00:00:22,160
as is revealed by small molecule

11
00:00:30,230 --> 00:00:28,230
mine's going to be a little high level

12
00:00:32,069 --> 00:00:30,240
and i think if we want to talk about

13
00:00:34,069 --> 00:00:32,079

universal life detection i think we have

14
00:00:38,150 --> 00:00:34,079
to start off with what is life

15
00:00:40,709 --> 00:00:38,160
there is a really obvious definition

16
00:00:42,150 --> 00:00:40,719
here on earth we see a human if we see

17
00:00:44,869 --> 00:00:42,160
something like this

18
00:00:45,190 --> 00:00:44,879
if we even see a virus it's very easy to

19
00:00:47,830 --> 00:00:45,200
say

20
00:00:49,590 --> 00:00:47,840
those things are alive like we know when

21
00:00:52,549 --> 00:00:49,600
we see it

22
00:00:53,750 --> 00:00:52,559
even when it comes to biosignatures we

23
00:00:54,950 --> 00:00:53,760
start seeing

24
00:00:56,709 --> 00:00:54,960
if you see something like highway

25
00:00:59,110 --> 00:00:56,719
interchange or see something like an

26
00:01:01,430 --> 00:00:59,120
oxygenated planet with lots of water and

27
00:01:03,029 --> 00:01:01,440
green things everywhere we know that

28
00:01:06,070 --> 00:01:03,039
that thing is alive as well

29
00:01:06,950 --> 00:01:06,080
it's very easy to say this thing's not

30
00:01:09,830 --> 00:01:06,960
random

31
00:01:11,670 --> 00:01:09,840
this is somehow alive where it starts

32
00:01:14,070 --> 00:01:11,680
getting a little weird

33
00:01:15,670 --> 00:01:14,080
is that we don't know the chemistry we

34
00:01:17,030 --> 00:01:15,680
don't know what a bio signature would

35
00:01:20,070 --> 00:01:17,040
look like

36
00:01:20,950 --> 00:01:20,080
um for ref we're on venus what does a

37
00:01:22,630 --> 00:01:20,960
life

38
00:01:24,870 --> 00:01:22,640

what does a living molecule even look

39
00:01:25,990 --> 00:01:24,880
like if we're on an exoplanet we have no

40
00:01:28,630 --> 00:01:26,000
idea what kind of chemistry

41
00:01:30,390 --> 00:01:28,640
is there you have no idea what kind of

42
00:01:31,510 --> 00:01:30,400
um thermodynamics kind of energetic

43
00:01:33,670 --> 00:01:31,520
constraints

44
00:01:35,910 --> 00:01:33,680
kind of energy systems there are we

45
00:01:39,190 --> 00:01:35,920
really have no idea about any of that

46
00:01:42,469 --> 00:01:39,200
and the life that i can introduce

47
00:01:44,230 --> 00:01:42,479
beginning this is very much earth biased

48
00:01:45,590 --> 00:01:44,240
we don't have a universal definition of

49
00:01:46,469 --> 00:01:45,600
life yet where we can look at something

50
00:01:49,510 --> 00:01:46,479
and say

51
00:01:52,550 --> 00:01:49,520
that thing is alive and that thing's not

52
00:01:55,670 --> 00:01:52,560
so part of what i'm trying to do here

53
00:01:57,990 --> 00:01:55,680
is to say here is a system for

54
00:01:59,190 --> 00:01:58,000
measuring chem here's a measurement of

55
00:02:02,310 --> 00:01:59,200
chemical systems

56
00:02:06,469 --> 00:02:02,320
that says something is alive or

57
00:02:09,350 --> 00:02:06,479
something's not um

58
00:02:10,790 --> 00:02:09,360
and crucially it's going to be chemical

59
00:02:13,990 --> 00:02:10,800
agnostic

60
00:02:17,430 --> 00:02:14,000
it's going to say cool

61
00:02:18,790 --> 00:02:17,440
i think that no matter what kind of

62
00:02:20,309 --> 00:02:18,800
chemistry we have

63
00:02:22,309 --> 00:02:20,319

whether it's on venus whether it's on

64
00:02:23,270 --> 00:02:22,319
exoplanet whatever

65
00:02:25,670 --> 00:02:23,280
where we don't know what kind of

66
00:02:26,390 --> 00:02:25,680
chemistry there is we can still look at

67
00:02:33,270 --> 00:02:26,400
the

68
00:02:35,910 --> 00:02:33,280
regardless of energy systems

69
00:02:38,150 --> 00:02:35,920
we're going to define and detect life

70
00:02:42,790 --> 00:02:38,160
using

71
00:02:46,550 --> 00:02:44,710
and the way i'm going to do that this is

72
00:02:48,470 --> 00:02:46,560
one proposal

73
00:02:49,670 --> 00:02:48,480
one particular way that i believe has

74
00:02:51,910 --> 00:02:49,680
some

75
00:02:53,110 --> 00:02:51,920
potential interest i'll explain a few

76
00:02:54,710 --> 00:02:53,120
others at the end

77
00:02:58,149 --> 00:02:54,720
um it's something called the maximal

78
00:03:01,270 --> 00:02:58,159
common substructure algorithm

79
00:03:04,149 --> 00:03:01,280
all it does is look at

80
00:03:06,229 --> 00:03:04,159
two chemical species nothing to do with

81
00:03:08,869 --> 00:03:06,239
the environments nothing with anything

82
00:03:10,550 --> 00:03:08,879
besides just the two chemical structures

83
00:03:13,350 --> 00:03:10,560
it's a graph based algorithm

84
00:03:15,350 --> 00:03:13,360
and what it does is takes it converts

85
00:03:16,790 --> 00:03:15,360
molecules to graphs

86
00:03:18,949 --> 00:03:16,800
it's done computationally behind the

87
00:03:21,270 --> 00:03:18,959
scenes and it takes the largest

88
00:03:22,630 --> 00:03:21,280

common substructure that's shared

89
00:03:26,869 --> 00:03:22,640
between those two

90
00:03:28,949 --> 00:03:26,879
so here we have two drugs and this

91
00:03:31,270 --> 00:03:28,959
carbon chain with nitrogen benzene ring

92
00:03:34,149 --> 00:03:31,280
with oxygen attached

93
00:03:35,589 --> 00:03:34,159
those that that's just that substructure

94
00:03:37,190 --> 00:03:35,599
is found between

95
00:03:38,710 --> 00:03:37,200
both of these two and it's the largest

96
00:03:40,789 --> 00:03:38,720
substructure that is shared between

97
00:03:43,430 --> 00:03:40,799
these two compounds

98
00:03:44,630 --> 00:03:43,440
and nothing to do with the chemistry we

99
00:03:45,509 --> 00:03:44,640
know we didn't even know where these

100
00:03:47,750 --> 00:03:45,519
came from

101
00:03:50,789 --> 00:03:47,760
and that's a good thing we're able to

102
00:03:52,630 --> 00:03:50,799
say we don't know what is going on here

103
00:03:54,470 --> 00:03:52,640
but there are shared compounds that

104
00:03:55,750 --> 00:03:54,480
these have potentially some kind of

105
00:04:00,869 --> 00:03:55,760
shared

106
00:04:03,830 --> 00:04:00,879
something like that

107
00:04:05,429 --> 00:04:03,840
and so what i'm doing to kind of say hey

108
00:04:08,229 --> 00:04:05,439
this is something that we could do

109
00:04:09,509 --> 00:04:08,239
is i'm starting with earth starting with

110
00:04:12,710 --> 00:04:09,519
earth kind of getting

111
00:04:13,750 --> 00:04:12,720
a picture of what life looks like here

112
00:04:13,990 --> 00:04:13,760
on earth and start comparing it to

113
00:04:17,110 --> 00:04:14,000

different

114
00:04:19,030 --> 00:04:17,120
sources um

115
00:04:21,590 --> 00:04:19,040
because life as we know is my favorite

116
00:04:23,430 --> 00:04:21,600
earth is only i have n equals one

117
00:04:24,629 --> 00:04:23,440
way of saying this is what we have for a

118
00:04:26,310 --> 00:04:24,639
living system

119
00:04:28,070 --> 00:04:26,320
um so what we do here is take all the

120
00:04:30,550 --> 00:04:28,080
biochemical compounds from keg

121
00:04:32,230 --> 00:04:30,560
it's a kyoto encyclopedia of genes and

122
00:04:35,510 --> 00:04:32,240
genomes

123
00:04:38,550 --> 00:04:35,520
it's one of the standards for

124
00:04:39,830 --> 00:04:38,560
small molecule biochemistry

125
00:04:41,909 --> 00:04:39,840
and we're going to take all these

126
00:04:43,110 --> 00:04:41,919
molecules whether they're acetyl coa

127
00:04:46,629 --> 00:04:43,120
some lipids

128
00:04:48,469 --> 00:04:46,639
some amino acids all this stuff and

129
00:04:50,870 --> 00:04:48,479
going to take all about 17 000 of those

130
00:04:52,390 --> 00:04:50,880
compounds and put them all through this

131
00:04:55,189 --> 00:04:52,400
mcs algorithm so you're going to

132
00:04:56,870 --> 00:04:55,199
pairwise match every single one

133
00:04:58,390 --> 00:04:56,880
ends up being a couple billion pairs so

134
00:05:01,749 --> 00:04:58,400
it takes a while to run

135
00:05:04,950 --> 00:05:01,759
and you end up with a lot of

136
00:05:05,430 --> 00:05:04,960
common substructures which return all of

137
00:05:06,790 --> 00:05:05,440
those

138
00:05:08,870 --> 00:05:06,800

remove duplicates because there's a ton

139
00:05:10,790 --> 00:05:08,880
of them and

140
00:05:13,029 --> 00:05:10,800
then count how many times each fragment

141
00:05:15,189 --> 00:05:13,039
appears in this data set

142
00:05:18,150 --> 00:05:15,199
um you're going to end up with a

143
00:05:21,749 --> 00:05:18,160
distribution pattern of how often

144
00:05:23,990 --> 00:05:21,759
certain chemical substructures appear

145
00:05:25,590 --> 00:05:24,000
and we're going to statistically

146
00:05:27,430 --> 00:05:25,600
distinguish those current patterns

147
00:05:30,150 --> 00:05:27,440
see if we can say biochemistry is

148
00:05:33,110 --> 00:05:30,160
different than say an abiotic data set

149
00:05:33,430 --> 00:05:33,120
um what this could potentially look like

150
00:05:37,749 --> 00:05:33,440
is

151
00:05:38,950 --> 00:05:37,759
has some kind of like shared of

152
00:05:42,870 --> 00:05:38,960
evolutionary

153
00:05:46,150 --> 00:05:42,880
got a shared chemistry

154
00:05:49,350 --> 00:05:46,160
throughout and some abiotic

155
00:05:50,070 --> 00:05:49,360
chemical fragments would be kind of the

156
00:05:51,749 --> 00:05:50,080
same

157
00:05:53,909 --> 00:05:51,759
throughout an entire data set so we can

158
00:05:56,870 --> 00:05:53,919
say life is different than

159
00:05:58,629 --> 00:05:56,880
whatever this abiotic set is if you see

160
00:06:01,670 --> 00:05:58,639
something like this we could say

161
00:06:03,270 --> 00:06:01,680
cool there is some evidence that

162
00:06:05,749 --> 00:06:03,280
this method could distinguish life from

163
00:06:08,150 --> 00:06:05,759

non-life living chemical systems from

164
00:06:11,670 --> 00:06:08,160
non-living chemical systems

165
00:06:13,670 --> 00:06:11,680
and you kind of see that which is well

166
00:06:15,110 --> 00:06:13,680
biochemistry looks like kind of what we

167
00:06:18,710 --> 00:06:15,120
expect um

168
00:06:21,270 --> 00:06:18,720
there are very few shared compounds

169
00:06:22,629 --> 00:06:21,280
so a carbon-carbon bond carbon double

170
00:06:26,710 --> 00:06:22,639
bond benzene ring

171
00:06:28,469 --> 00:06:26,720
those things appear a lot so they end up

172
00:06:30,150 --> 00:06:28,479
and they are shared almost throughout

173
00:06:31,749 --> 00:06:30,160
all of chemistry i think

174
00:06:34,150 --> 00:06:31,759
something like nearly 100 chemistry has

175
00:06:36,790 --> 00:06:34,160
a carbon-carbon bond unsurprising

176
00:06:38,469 --> 00:06:36,800
but there are very few of those shared

177
00:06:41,749 --> 00:06:38,479
chemical substructures

178
00:06:43,830 --> 00:06:41,759
and then there are very many of these

179
00:06:45,909 --> 00:06:43,840
kind of complicated substructures

180
00:06:47,110 --> 00:06:45,919
so that that drug-based substructure we

181
00:06:48,950 --> 00:06:47,120
found

182
00:06:50,790 --> 00:06:48,960
somewhere down here there's very very

183
00:06:54,629 --> 00:06:50,800
very few biochemical compounds

184
00:06:56,070 --> 00:06:54,639
that have this substructure so this is

185
00:06:57,670 --> 00:06:56,080
very similar to a power law pattern we

186
00:06:58,870 --> 00:06:57,680
see this in other forms

187
00:07:00,710 --> 00:06:58,880
of life so i'm not really going to get

188
00:07:02,230 --> 00:07:00,720

into that too much but

189
00:07:04,309 --> 00:07:02,240
this is potentially a distinguishing

190
00:07:07,990 --> 00:07:04,319
significant future of life

191
00:07:10,070 --> 00:07:08,000
um and i think thankfully

192
00:07:11,510 --> 00:07:10,080
um the three domains of life when you

193
00:07:16,629 --> 00:07:11,520
start looking at

194
00:07:18,710 --> 00:07:16,639
genomes accessed through the joint

195
00:07:19,670 --> 00:07:18,720
genome institute

196
00:07:21,670 --> 00:07:19,680
and start linking them up with the

197
00:07:22,790 --> 00:07:21,680
compounds that those genomes have be

198
00:07:23,990 --> 00:07:22,800
able to say that the three domains of

199
00:07:27,909 --> 00:07:24,000
life are pretty similar

200
00:07:29,029 --> 00:07:27,919
so this is able to one say life is

201
00:07:30,309 --> 00:07:29,039
similar which is good

202
00:07:31,110 --> 00:07:30,319
if those domains are all over the place

203
00:07:32,790 --> 00:07:31,120
this would be a little hard to

204
00:07:36,070 --> 00:07:32,800
distinguish

205
00:07:38,309 --> 00:07:36,080
and my other data set was

206
00:07:39,990 --> 00:07:38,319
technologically produced chemistry so

207
00:07:43,110 --> 00:07:40,000
take data from

208
00:07:44,309 --> 00:07:43,120
re-access which we have access to via

209
00:07:45,670 --> 00:07:44,319
the collaboration with the university of

210
00:07:49,749 --> 00:07:45,680
glasgow

211
00:07:51,909 --> 00:07:49,759
and i'm able to say reaccess has

212
00:07:54,230 --> 00:07:51,919
very very few shared compounds just

213
00:07:57,270 --> 00:07:54,240

noted it just drops off precipitously

214
00:07:58,309 --> 00:07:57,280
and there are very many compounds or

215
00:08:01,110 --> 00:07:58,319
substructures

216
00:08:02,790 --> 00:08:01,120
that are shared throughout there are

217
00:08:03,749 --> 00:08:02,800
very many substructures that are not

218
00:08:06,710 --> 00:08:03,759
shared

219
00:08:07,830 --> 00:08:06,720
so see it's technologically produced

220
00:08:11,110 --> 00:08:07,840
chemistry is less

221
00:08:13,510 --> 00:08:11,120
shared than biology

222
00:08:14,869 --> 00:08:13,520
which which is interesting um shows that

223
00:08:17,350 --> 00:08:14,879
if something has this kind of not

224
00:08:18,469 --> 00:08:17,360
shared pattern um probably doesn't have

225
00:08:19,510 --> 00:08:18,479
this evolutionary history of what

226
00:08:22,790 --> 00:08:19,520
chemistry does

227
00:08:24,070 --> 00:08:22,800
that's my initial hypothesis um

228
00:08:25,350 --> 00:08:24,080
what's kind of saying that here is that

229
00:08:26,469 --> 00:08:25,360
biochemistry is more shared

230
00:08:28,390 --> 00:08:26,479
substructures

231
00:08:30,150 --> 00:08:28,400
than a non-living technology to produce

232
00:08:32,310 --> 00:08:30,160
system

233
00:08:33,990 --> 00:08:32,320
a few caveats that reaccess is primarily

234
00:08:35,430 --> 00:08:34,000
a pharmaceutical database

235
00:08:37,110 --> 00:08:35,440
there's a lot of materials science and

236
00:08:37,990 --> 00:08:37,120
there's a lot of weird metal chemistry

237
00:08:39,670 --> 00:08:38,000
going on there

238
00:08:41,829 --> 00:08:39,680

but a significant portion is

239
00:08:44,630 --> 00:08:41,839
pharmaceutical based so it is

240
00:08:46,389 --> 00:08:44,640
biochemical adjacent um it's not what's

241
00:08:50,470 --> 00:08:46,399
made in biochemistry

242
00:08:54,870 --> 00:08:53,590
next steps the most common substructure

243
00:08:55,350 --> 00:08:54,880
algorithm is not the only way we can do

244
00:08:57,350 --> 00:08:55,360
this

245
00:08:59,350 --> 00:08:57,360
um one of the reasons i am here in

246
00:09:00,310 --> 00:08:59,360
glasgow is to study molecular assembly

247
00:09:02,389 --> 00:09:00,320
fragments

248
00:09:03,350 --> 00:09:02,399
which um there's a paper that just came

249
00:09:08,070 --> 00:09:03,360
out

250
00:09:09,509 --> 00:09:08,080
how

251
00:09:11,350 --> 00:09:09,519
assembly theory can be used as a

252
00:09:14,310 --> 00:09:11,360
potential biosignature

253
00:09:16,550 --> 00:09:14,320
and it's also a preprint which will soon

254
00:09:19,110 --> 00:09:16,560
be published in science advances

255
00:09:20,070 --> 00:09:19,120
of looking at modular assembly as a way

256
00:09:25,190 --> 00:09:20,080
to

257
00:09:27,590 --> 00:09:25,200
evolutionarily related compounds

258
00:09:28,630 --> 00:09:27,600
um essentially it makes fragments as

259
00:09:29,350 --> 00:09:28,640
well you're able to break down

260
00:09:31,190 --> 00:09:29,360
components

261
00:09:32,949 --> 00:09:31,200
or to break down the chemicals into

262
00:09:34,630 --> 00:09:32,959
component fragments

263
00:09:35,990 --> 00:09:34,640

assembly theory builds them back up here

264
00:09:37,590 --> 00:09:36,000
it's looking at the fragments i'm

265
00:09:39,670 --> 00:09:37,600
curious to see how they compare with the

266
00:09:41,430 --> 00:09:39,680
mcs algorithms

267
00:09:43,430 --> 00:09:41,440
any questions on this stuff specifically

268
00:09:47,350 --> 00:09:43,440
let me know and we're very happy to talk

269
00:09:52,230 --> 00:09:50,710
the what i'm trying to say but i think

270
00:09:53,910 --> 00:09:52,240
my goal is is to say

271
00:09:55,509 --> 00:09:53,920
um chemical agnostic measures can serve

272
00:09:57,509 --> 00:09:55,519
as the basis for identifying

273
00:09:59,509 --> 00:09:57,519
these limitable these these living

274
00:10:00,710 --> 00:09:59,519
chemical systems

275
00:10:02,470 --> 00:10:00,720
i think you need to have something

276
00:10:03,990 --> 00:10:02,480
that's chemically agnostic i think

277
00:10:05,030 --> 00:10:04,000
you're not able to distinguish chemical

278
00:10:07,190 --> 00:10:05,040
systems if you're looking only at

279
00:10:08,870 --> 00:10:07,200
biochemistry because you get a little or

280
00:10:10,389 --> 00:10:08,880
because you get biased with what earth

281
00:10:13,030 --> 00:10:10,399
chemistry has to offer

282
00:10:14,470 --> 00:10:13,040
um we won't know what chemistry looks

283
00:10:17,990 --> 00:10:14,480
like in other planets

284
00:10:19,829 --> 00:10:18,000
um and i think in order to have a full

285
00:10:21,190 --> 00:10:19,839
debate about what life is

286
00:10:23,269 --> 00:10:21,200
in order to distinguish life from

287
00:10:27,190 --> 00:10:23,279
chemistry you need to make sure it's a

288
00:10:32,230 --> 00:10:29,990

um work with sarah walker at arizona

289
00:10:34,470 --> 00:10:32,240
state university so thank you to her for

290
00:10:36,150 --> 00:10:34,480
helping me with this um working with lee

291
00:10:37,910 --> 00:10:36,160
cronin and glasgow

292
00:10:39,590 --> 00:10:37,920
certainly been an interesting few weeks

293
00:10:41,030 --> 00:10:39,600
so far

294
00:10:43,350 --> 00:10:41,040
and really excited to continue working

295
00:10:45,910 --> 00:10:43,360
here everyone in the

296
00:10:48,150 --> 00:10:45,920
arizona state lab everyone that i've met

297
00:10:51,110 --> 00:10:48,160
so far in the cronin group

298
00:10:51,829 --> 00:10:51,120
and also for thank you for that crack at

299
00:10:55,030 --> 00:10:51,839
grad con

300
00:10:56,550 --> 00:10:55,040
opportunity um to present here and

301
00:10:57,750 --> 00:10:56,560
meet everyone it's a wonderful

302
00:10:59,509 --> 00:10:57,760
conference i'm always having a lot of

303
00:11:02,069 --> 00:10:59,519
fun