
AbSciCon
2017
MESA, ARIZONA

1
00:00:12,250 --> 00:00:06,150
you

2
00:00:16,660 --> 00:00:14,100
[Music]

3
00:00:19,510 --> 00:00:16,670
thank you very much Beth and Alexis for

4
00:00:21,280 --> 00:00:19,520
the opportunity to share my latest

5
00:00:24,970 --> 00:00:21,290
research in progress with you guys here

6
00:00:27,429 --> 00:00:24,980
today so today I'm going to talk about

7
00:00:35,430 --> 00:00:27,439
some meta genomic data from a tropical

8
00:00:39,670 --> 00:00:37,360
thank you

9
00:00:41,439 --> 00:00:39,680
so we've already heard about the process

10
00:00:43,780 --> 00:00:41,449
of serpentinization multiple times today

11
00:00:46,930 --> 00:00:43,790
so I won't take too much time here to

12
00:00:49,000 --> 00:00:46,940
belabor the point however as far as

13
00:00:51,130 --> 00:00:49,010

astrobiology is concerned it's

14
00:00:53,290 --> 00:00:51,140
essentially a process where we've got

15
00:00:56,530 --> 00:00:53,300
minerals and water coming together and

16
00:00:59,440 --> 00:00:56,540
producing life molecules hydrogen

17
00:01:01,600 --> 00:00:59,450
methane and small organic carbons and so

18
00:01:03,760 --> 00:01:01,610
from an astro biological point of view

19
00:01:05,079 --> 00:01:03,770
that has implications for both the

20
00:01:08,800 --> 00:01:05,089
origin of life here on earth and

21
00:01:10,810 --> 00:01:08,810
potentially life on other planets one of

22
00:01:12,820 --> 00:01:10,820
the best characterized sites of

23
00:01:14,980 --> 00:01:12,830
serpentinization is the lost city

24
00:01:17,920 --> 00:01:14,990
hydrothermal field which is located off

25
00:01:21,690 --> 00:01:17,930
of the mid-atlantic ridge and the

26
00:01:25,960 --> 00:01:21,700
communities there are dominated by

27
00:01:28,749 --> 00:01:25,970
methane cycling archaea on the interior

28
00:01:31,300 --> 00:01:28,759
of the chimneys and they're called

29
00:01:33,670 --> 00:01:31,310
velocity Misano sarsen alleys and then

30
00:01:35,520 --> 00:01:33,680
the exterior of the vent chimneys are

31
00:01:38,020 --> 00:01:35,530
slightly more diverse and contain

32
00:01:40,030 --> 00:01:38,030
aerobic methane atrophic bacteria and

33
00:01:41,980 --> 00:01:40,040
what this suggests is that methane

34
00:01:46,020 --> 00:01:41,990
cycling by microbes is incredibly

35
00:01:49,090 --> 00:01:46,030
important to communities in velocity

36
00:01:51,010 --> 00:01:49,100
this is a map of some sites that our

37
00:01:53,609 --> 00:01:51,020
research group either has projects at or

38
00:01:56,920 --> 00:01:53,619

collaborations with other groups at and

39
00:02:00,100 --> 00:01:56,930
we've been studying serpentinization in

40
00:02:02,260 --> 00:02:00,110
a continental setting for years and so

41
00:02:04,630 --> 00:02:02,270
these results from the lost city beg the

42
00:02:06,520 --> 00:02:04,640
question is methane cycling by microbes

43
00:02:12,370 --> 00:02:06,530
important in these continental sites of

44
00:02:15,060 --> 00:02:12,380
certain ization as well so methanogens

45
00:02:16,840 --> 00:02:15,070
have been detected both at Oman and

46
00:02:20,050 --> 00:02:16,850
delivery and ophiolite

47
00:02:24,760 --> 00:02:20,060
in Italy using either 16s or meta

48
00:02:26,230 --> 00:02:24,770
genomics data sets in Portugal the table

49
00:02:28,150 --> 00:02:26,240
Enzo filet in Canada

50
00:02:30,100 --> 00:02:28,160
and the coaster angel feel light in

51
00:02:33,550 --> 00:02:30,110
California they have not been detected

52
00:02:35,350 --> 00:02:33,560
and that's either from 16s datasets

53
00:02:37,870 --> 00:02:35,360
metagenomic data sets are a combination

54
00:02:40,180 --> 00:02:37,880
of the two and I just want to use this

55
00:02:41,770 --> 00:02:40,190
moment to plug a poster that's at this

56
00:02:43,510 --> 00:02:41,780
evening if you'd like to hear more about

57
00:02:47,020 --> 00:02:43,520
the table m sophia light please see

58
00:02:49,540 --> 00:02:47,030
poster 33 11 by a technician in the

59
00:02:51,210 --> 00:02:49,550
Brazelton lab and so for the remainder

60
00:02:55,180 --> 00:02:51,220
of the talk I would like to talk about

61
00:02:57,160 --> 00:02:55,190
this one site in Costa Rica and whether

62
00:03:01,870 --> 00:02:57,170
or not we see methane cycling going on

63
00:03:04,990 --> 00:03:01,880

there so the Santa Elena alveoli is

64
00:03:08,680 --> 00:03:05,000
located in the North Pacific corner of

65
00:03:11,350 --> 00:03:08,690
Costa Rica a paper by Sanchez Rio at all

66
00:03:12,730 --> 00:03:11,360
looked at various geological and

67
00:03:15,160 --> 00:03:12,740
geochemical and microbiological

68
00:03:17,350 --> 00:03:15,170
parameters at a variety of sites this

69
00:03:23,230 --> 00:03:17,360
location today I'll be talking about

70
00:03:26,650 --> 00:03:23,240
sites located right here so two sites in

71
00:03:30,430 --> 00:03:26,660
particular we've got spring 9 and a very

72
00:03:33,450 --> 00:03:30,440
close proximity

73
00:03:36,640 --> 00:03:33,460
upstream site for a background sample

74
00:03:38,680 --> 00:03:36,650
spring 9 has the typical travertine

75
00:03:42,670 --> 00:03:38,690
calcium carbonate deposits that you'd

76
00:03:44,500 --> 00:03:42,680
expect when the high calcium fluids from

77
00:03:47,100 --> 00:03:44,510
serpentinization come into contact with

78
00:03:49,390 --> 00:03:47,110
the inorganic carbon in the air and

79
00:03:51,010 --> 00:03:49,400
despite they're very close proximity of

80
00:03:52,450 --> 00:03:51,020
a couple of meters from each other they

81
00:03:55,960 --> 00:03:52,460
have different geochemistry

82
00:03:57,270 --> 00:03:55,970
a spring 9 has a pH of 11.5 and an

83
00:04:00,340 --> 00:03:57,280
abundance of hydrogen and methane

84
00:04:02,680 --> 00:04:00,350
whereas the background sample still has

85
00:04:06,460 --> 00:04:02,690
an elevated pH but it's much lower and

86
00:04:08,470 --> 00:04:06,470
has less of these volatiles and bio

87
00:04:10,540 --> 00:04:08,480
energetic calculations have suggested

88
00:04:12,220 --> 00:04:10,550

that Mehsana trophy and Mehsana genesis

89
00:04:15,880 --> 00:04:12,230
are both favorable at both of these

90
00:04:19,599 --> 00:04:15,890
sites so now to get into a little bit of

91
00:04:22,030 --> 00:04:19,609
the microbiological data a lot of

92
00:04:25,120 --> 00:04:22,040
continental serpentinization sites have

93
00:04:28,420 --> 00:04:25,130
either no or very low abundance of

94
00:04:31,630 --> 00:04:28,430
archaeal toxa and so here we looked at

95
00:04:33,280 --> 00:04:31,640
some quantitative PCR data and i've

96
00:04:34,870 --> 00:04:33,290
highlighted here the two samples that

97
00:04:37,500 --> 00:04:34,880
we'll be talking about in this talk and

98
00:04:39,670 --> 00:04:37,510
well the archaea

99
00:04:41,770 --> 00:04:39,680
well archaea are less of

100
00:04:43,689 --> 00:04:41,780
and the bacteria they still make up a

101
00:04:45,520 --> 00:04:43,699
substantial portion of the microbial

102
00:04:49,870 --> 00:04:45,530
communities which is contrary to what

103
00:04:51,400 --> 00:04:49,880
we've seen in other sites to explore the

104
00:04:55,930 --> 00:04:51,410
methane cycling a little bit further I

105
00:04:59,080 --> 00:04:55,940
mind some 16s rRNA amplicon datasets to

106
00:05:01,210 --> 00:04:59,090
look for potential methane cyclers here

107
00:05:04,150 --> 00:05:01,220
we're looking at the bacterial

108
00:05:06,640 --> 00:05:04,160
communities and I did not want to do

109
00:05:08,620 --> 00:05:06,650
that I just want to orient you here that

110
00:05:10,960 --> 00:05:08,630
the units are the percent of the total

111
00:05:13,060 --> 00:05:10,970
bacterial community but the axis goes

112
00:05:15,100 --> 00:05:13,070
from zero to ten percent and that's

113
00:05:18,580 --> 00:05:15,110

because these were not in very high

114
00:05:21,640 --> 00:05:18,590
abundance at all we see some aerobic

115
00:05:24,100 --> 00:05:21,650
Masana Tropes in spring nine and a

116
00:05:25,930 --> 00:05:24,110
slightly higher diversity and higher

117
00:05:27,879 --> 00:05:25,940
abundance in the background sample but

118
00:05:30,969 --> 00:05:27,889
it still makes up less than 1% of the

119
00:05:33,760 --> 00:05:30,979
total community the archaea tell a

120
00:05:35,379 --> 00:05:33,770
slightly different story again we're

121
00:05:38,140 --> 00:05:35,389
looking at the percent of the total of

122
00:05:41,770 --> 00:05:38,150
peer community but here the the access

123
00:05:44,230 --> 00:05:41,780
ranges from zero to 100% in the warm

124
00:05:46,149 --> 00:05:44,240
colors were looking at methanogens and

125
00:05:48,670 --> 00:05:46,159
in the cooler colors their putative

126
00:05:51,520 --> 00:05:48,680
mechanic tropes and so here we can see

127
00:05:54,790 --> 00:05:51,530
that in our background our in our high

128
00:05:56,950 --> 00:05:54,800
pH spring 9 the methane cycling archaea

129
00:05:58,990 --> 00:05:56,960
make up between forty to eighty percent

130
00:06:00,430 --> 00:05:59,000
of the total archaeal community and in

131
00:06:02,230 --> 00:06:00,440
the background sample they're still

132
00:06:06,279 --> 00:06:02,240
significant making up about thirty to

133
00:06:08,790 --> 00:06:06,289
forty percent of that community so since

134
00:06:11,170 --> 00:06:08,800
not everybody in the room right now is a

135
00:06:12,939 --> 00:06:11,180
molecular biologist I thought I'd take

136
00:06:14,980 --> 00:06:12,949
just a few moments to explain some of

137
00:06:17,200 --> 00:06:14,990
the methodology for the data I'm about

138
00:06:22,800 --> 00:06:17,210

to present so we use meta genomic

139
00:06:25,180 --> 00:06:22,810
sequencing and we filter the natural

140
00:06:27,100 --> 00:06:25,190
microbial communities we then extract

141
00:06:28,480 --> 00:06:27,110
DNA and chop that up into little pieces

142
00:06:31,659 --> 00:06:28,490
that can be sequenced on a

143
00:06:33,700 --> 00:06:31,669
high-throughput sequencer the ultimate

144
00:06:36,010 --> 00:06:33,710
goal for metagenomics is to try to

145
00:06:39,189 --> 00:06:36,020
reassemble complete genomes of

146
00:06:41,589 --> 00:06:39,199
individuals from the environment and so

147
00:06:43,719 --> 00:06:41,599
we take these small little pieces of DNA

148
00:06:46,149 --> 00:06:43,729
and try to put them back together in

149
00:06:48,610 --> 00:06:46,159
larger chunks through assembly methods

150
00:06:50,300 --> 00:06:48,620
those larger trucks chunks of DNA are

151
00:06:52,790 --> 00:06:50,310
called con tags

152
00:06:55,490 --> 00:06:52,800
binning is the step where you try to put

153
00:06:58,420 --> 00:06:55,500
back together these genomes either

154
00:07:02,659 --> 00:06:58,430
complete or more likely near complete

155
00:07:04,879 --> 00:07:02,669
and so here this represents potential

156
00:07:07,310 --> 00:07:04,889
individual organisms where you can then

157
00:07:11,060 --> 00:07:07,320
annotate and predict their different

158
00:07:14,810 --> 00:07:11,070
metabolic pathways and so this is an

159
00:07:18,200 --> 00:07:14,820
example of the assembly from spring 9

160
00:07:20,840 --> 00:07:18,210
the high ph spring from this assembly we

161
00:07:23,030 --> 00:07:20,850
have eleven metagenomic bins ten of them

162
00:07:25,190 --> 00:07:23,040
were classified as bacteria one was

163
00:07:28,220 --> 00:07:25,200

classified as unknown which means that

164
00:07:31,580 --> 00:07:28,230
it wasn't even able to be determined at

165
00:07:35,120 --> 00:07:31,590
the domain level and the coverage

166
00:07:37,280 --> 00:07:35,130
represents how complete I'm sorry the

167
00:07:39,170 --> 00:07:37,290
completeness represents how complete

168
00:07:42,170 --> 00:07:39,180
these genomes are since we're trying to

169
00:07:43,850 --> 00:07:42,180
obtain ideally full genomes and so you

170
00:07:48,680 --> 00:07:43,860
can see here we have a wide range from

171
00:07:54,350 --> 00:07:48,690
17 to 83% completeness which of the

172
00:07:56,390 --> 00:07:54,360
lower end is is not great and so would

173
00:07:59,719 --> 00:07:56,400
these bacteria I should point out that

174
00:08:02,540 --> 00:07:59,729
number one none of these were identified

175
00:08:05,690 --> 00:08:02,550
as explicit Masana tropes and also we

176
00:08:07,070 --> 00:08:05,700
don't see any are keel bone here however

177
00:08:08,990 --> 00:08:07,080
as I mentioned before from this

178
00:08:10,880 --> 00:08:09,000
metagenomic data we can also not only

179
00:08:13,610 --> 00:08:10,890
look at who's there but what their

180
00:08:17,779 --> 00:08:13,620
metabolic potential is and so this is

181
00:08:20,120 --> 00:08:17,789
the Masana genesis pathway these numbers

182
00:08:23,420 --> 00:08:20,130
here all represent potential genes in

183
00:08:25,520 --> 00:08:23,430
the pathway the genes highlighted in

184
00:08:29,090 --> 00:08:25,530
blue were found in both the background

185
00:08:30,469 --> 00:08:29,100
and the high ph spring whereas the genes

186
00:08:32,899 --> 00:08:30,479
highlighted in orange were found

187
00:08:35,540 --> 00:08:32,909
exclusively in the high ph sample and i

188
00:08:37,130 --> 00:08:35,550

should note that there were no no genes

189
00:08:39,020 --> 00:08:37,140
involved in methanogenesis that were

190
00:08:41,240 --> 00:08:39,030
found exclusively in the background

191
00:08:44,390 --> 00:08:41,250
sample and so what we can see here is

192
00:08:46,780 --> 00:08:44,400
that we have complete meth an agenda

193
00:08:50,590 --> 00:08:46,790
this fast way in the high pH Springs

194
00:08:53,450 --> 00:08:50,600
also there were no I searched for

195
00:08:57,800 --> 00:08:53,460
bacterial Nathaniel genes in these meta

196
00:08:59,240 --> 00:08:57,810
genomes and none were found so as I

197
00:09:01,310 --> 00:08:59,250
mentioned there's been a lot of research

198
00:09:03,600 --> 00:09:01,320
going on in the past few years into the

199
00:09:05,490 --> 00:09:03,610
microbial diversity

200
00:09:07,860 --> 00:09:05,500
metabolism in continental

201
00:09:10,460 --> 00:09:07,870
serpentinization sites and from that

202
00:09:12,740 --> 00:09:10,470
there some models have been developed

203
00:09:16,740 --> 00:09:12,750
some things that are found almost

204
00:09:20,610 --> 00:09:16,750
universally is an abundance of these

205
00:09:23,310 --> 00:09:20,620
beta Proteobacteria and also anaerobic

206
00:09:24,690 --> 00:09:23,320
clostridia and this data set was no

207
00:09:27,450 --> 00:09:24,700
different there were an abundance of

208
00:09:30,120 --> 00:09:27,460
both of those groups one question that

209
00:09:33,240 --> 00:09:30,130
still remains is what the role and

210
00:09:35,910 --> 00:09:33,250
abundance of these archaea methanogens

211
00:09:37,350 --> 00:09:35,920
are in this system we still don't have

212
00:09:39,060 --> 00:09:37,360
an answer to that as they're found in

213
00:09:42,600 --> 00:09:39,070

some sites and not others but we're

214
00:09:45,000 --> 00:09:42,610
beginning to dive into that now so just

215
00:09:46,860 --> 00:09:45,010
to summarize the evidence for Monsanto

216
00:09:48,420 --> 00:09:46,870
Genesis at the santolina ophiolite

217
00:09:51,720 --> 00:09:48,430
include bioenergetics

218
00:09:55,410 --> 00:09:51,730
our keel 16s data both from qpcr and

219
00:09:58,430 --> 00:09:55,420
amplicons and Fulmer Masana Genesis

220
00:10:01,230 --> 00:09:58,440
pathway in the spring nine meta-genome

221
00:10:03,720 --> 00:10:01,240
however know our keel bins were found in

222
00:10:06,110 --> 00:10:03,730
that meta-genome and i just want to take

223
00:10:08,250 --> 00:10:06,120
a quick moment to explain that

224
00:10:11,640 --> 00:10:08,260
assembling metagenomes is much like

225
00:10:12,810 --> 00:10:11,650
taking hundreds of puzzles and spreading

226
00:10:14,790 --> 00:10:12,820
out all the pieces and trying to

227
00:10:17,220 --> 00:10:14,800
reassemble those puzzles again and so

228
00:10:19,170 --> 00:10:17,230
well in this go-around the puzzles

229
00:10:21,210 --> 00:10:19,180
didn't come back together very well

230
00:10:24,990 --> 00:10:21,220
there are other methods to try to try to

231
00:10:27,060 --> 00:10:25,000
improve those bins so my next steps are

232
00:10:28,920 --> 00:10:27,070
going to be to try to optimize the meta

233
00:10:30,890 --> 00:10:28,930
genomic assemblies and try to get more

234
00:10:33,510 --> 00:10:30,900
complete and representative bins

235
00:10:35,670 --> 00:10:33,520
additionally we have meta transcriptomic

236
00:10:37,170 --> 00:10:35,680
data for both of these samples and meta

237
00:10:39,150 --> 00:10:37,180
transcriptomes can tell you not only

238
00:10:41,700 --> 00:10:39,160

what genes are present in the

239
00:10:42,690 --> 00:10:41,710
environment like the metagenomes but

240
00:10:44,690 --> 00:10:42,700
they can tell you which ones are

241
00:10:48,510 --> 00:10:44,700
actively being expressed by the microbes

242
00:10:51,330 --> 00:10:48,520
so stay tuned with that I would like to

243
00:10:52,860 --> 00:10:51,340
thank my co-authors and collaborators my

244
00:10:55,470 --> 00:10:52,870
funding source of the deep carbon

245
00:10:57,750 --> 00:10:55,480
Observatory as well as everyone who

246
00:11:07,650 --> 00:10:57,760
helped out with data analysis sequencing

247
00:11:12,250 --> 00:11:09,940
Thank You Katrina we have plenty of time

248
00:11:13,720 --> 00:11:12,260
for questions if folks want to come up

249
00:11:23,290 --> 00:11:13,730
to the microphone or stand up and speak

250
00:11:26,410 --> 00:11:23,300
really loudly yeah that's very nice so

251
00:11:30,340 --> 00:11:26,420
one of the persistent issues in genome

252
00:11:31,810 --> 00:11:30,350
analysis is unassigned operating frames

253
00:11:36,310 --> 00:11:31,820
or open air nutrients with unknown

254
00:11:37,420 --> 00:11:36,320
functions do you find in your in your

255
00:11:41,769 --> 00:11:37,430
various samples from different

256
00:11:45,639 --> 00:11:41,779
environments greater or lower abundances

257
00:11:47,170 --> 00:11:45,649
of the hard to assign stuff that might

258
00:11:49,030 --> 00:11:47,180
that might indicate that whatever

259
00:11:52,600 --> 00:11:49,040
they're doing this environmental

260
00:11:54,940 --> 00:11:52,610
condition needs to make use of them yeah

261
00:11:56,829 --> 00:11:54,950
that's always a problem so anytime that

262
00:12:00,220 --> 00:11:56,839
you're trying to assign whether it be

263
00:12:02,800 --> 00:12:00,230

taxonomy or in this case function you're

264
00:12:04,420 --> 00:12:02,810
limited by the databases that exist and

265
00:12:06,880 --> 00:12:04,430
the databases are limited by what is

266
00:12:09,400 --> 00:12:06,890
already known and so I think even in

267
00:12:11,860 --> 00:12:09,410
e.coli about 40% of the genes are still

268
00:12:13,660 --> 00:12:11,870
hypothetical meaning Sara Dean but we

269
00:12:15,850 --> 00:12:13,670
don't know what it does and that number

270
00:12:17,829 --> 00:12:15,860
just goes up as you get into unexplored

271
00:12:19,420 --> 00:12:17,839
environment off the top of my head I

272
00:12:22,540 --> 00:12:19,430
don't know what percentage of our Dean's

273
00:12:24,760 --> 00:12:22,550
are hypothetical but it's definitely an

274
00:12:27,519 --> 00:12:24,770
issue cultured representatives can help

275
00:12:33,040 --> 00:12:27,529
us address that but is it is it fairly

276
00:12:36,640 --> 00:12:33,050
consistent across the different samples

277
00:12:38,560 --> 00:12:36,650
sources or philo types that's a great

278
00:12:42,000 --> 00:12:38,570
question I don't know the answer to off

279
00:12:48,660 --> 00:12:45,990
Catrina I have a question the fact that

280
00:12:50,280 --> 00:12:48,670
you see some of these archaea I wonder

281
00:12:52,050 --> 00:12:50,290
if it's related to other parts of the

282
00:12:53,060 --> 00:12:52,060
chemistry of these waters in addition to

283
00:12:54,780 --> 00:12:53,070
just the methane and hydrogen

284
00:12:58,020 --> 00:12:54,790
particularly what is the sulfate

285
00:13:00,690 --> 00:12:58,030
concentration like in these fluids yeah

286
00:13:03,870 --> 00:13:00,700
I know that some measurements were taken

287
00:13:06,210 --> 00:13:03,880
I don't recall what they were off the

288
00:13:09,090 --> 00:13:06,220

top of my head but we are beginning to

289
00:13:11,100 --> 00:13:09,100
look into some of these other electron

290
00:13:16,740 --> 00:13:11,110
acceptors at these sites to answer some

291
00:13:22,950 --> 00:13:16,750
of the questions any other questions for

292
00:13:26,610 --> 00:13:22,960
Katrina there we go I got one hey

293
00:13:28,920 --> 00:13:26,620
Katrina Thanks so do I understand it

294
00:13:30,570 --> 00:13:28,930
correctly that the methane cycling genes

295
00:13:32,130 --> 00:13:30,580
were found just in the random

296
00:13:37,410 --> 00:13:32,140
meta-genome but they did they weren't

297
00:13:39,810 --> 00:13:37,420
recovered in any bin no actually

298
00:13:42,870 --> 00:13:39,820
so the annotations were done on the bins

299
00:13:45,120 --> 00:13:42,880
which raises a question of the validity

300
00:13:48,780 --> 00:13:45,130
of some of the bins that those genes

301
00:13:50,520 --> 00:13:48,790
were all archeo but all of the bins were

302
00:13:52,710 --> 00:13:50,530
identified as bacteria and so I think

303
00:13:55,680 --> 00:13:52,720
what we likely have is some archaeal

304
00:13:57,840 --> 00:13:55,690
contamination within those bins so it's

305
00:14:00,300 --> 00:13:57,850
promising that we see the full pathway

306
00:14:03,030 --> 00:14:00,310
but I think the bins need to be

307
00:14:10,020 --> 00:14:03,040
reassembled to try to get up at a little