
AbSciCon
2017
MESA, ARIZONA

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

2
00:00:16,690 --> 00:00:14,040
[Music]

3
00:00:18,790 --> 00:00:16,700
hi everyone I hope you can hear me I'm

4
00:00:22,060 --> 00:00:18,800
I've been a little sick so sorry if my

5
00:00:23,710 --> 00:00:22,070
voice is creaky i'm abby karen i'm a

6
00:00:25,600 --> 00:00:23,720
research assistant with greg for nia at

7
00:00:27,880 --> 00:00:25,610
MIT and i'm going to be speaking with

8
00:00:33,220 --> 00:00:27,890
you about phylogenetic proxies for the

9
00:00:35,680 --> 00:00:33,230
rise of atmospheric oxygen i can click

10
00:00:37,299 --> 00:00:35,690
it okay so the rise of oxygen as we

11
00:00:39,759 --> 00:00:37,309
currently understand it is mostly

12
00:00:41,820 --> 00:00:39,769
informed by these geochemical proxies

13
00:00:44,320 --> 00:00:41,830

like you just heard in the previous talk

14
00:00:44,950 --> 00:00:44,330
here you know two figures briefly going

15
00:00:46,510 --> 00:00:44,960
over it

16
00:00:48,880 --> 00:00:46,520
oxygen levels in the early Earth were

17
00:00:52,420 --> 00:00:48,890
really low they rose at the great

18
00:00:53,979 --> 00:00:52,430
oxidation event around 2.3 2.4 Anna did

19
00:00:55,840 --> 00:00:53,989
something in here they might have risen

20
00:00:57,970 --> 00:00:55,850
and then crashed and stayed really low

21
00:01:00,430 --> 00:00:57,980
or they might have risen to a small

22
00:01:02,999 --> 00:01:00,440
percent of modern and then rose again at

23
00:01:05,410 --> 00:01:03,009
the at the neoproterozoic oxygenation to

24
00:01:07,270 --> 00:01:05,420
approximately modern levels but what

25
00:01:10,480 --> 00:01:07,280
happened in there is still a little bit

26
00:01:13,450 --> 00:01:10,490
controversial and so I thought well

27
00:01:14,740 --> 00:01:13,460
there are two records of the history of

28
00:01:16,990 --> 00:01:14,750
life on Earth right there's this

29
00:01:18,700 --> 00:01:17,000
geochemical record that we just heard

30
00:01:19,780 --> 00:01:18,710
about and they're isotopes preserved and

31
00:01:21,850 --> 00:01:19,790
you can look at how they change over

32
00:01:23,980 --> 00:01:21,860
time through the stratigraphy I'm not

33
00:01:26,080 --> 00:01:23,990
super cool but there's also this totally

34
00:01:28,450 --> 00:01:26,090
separate record preserved in the genetic

35
00:01:29,950 --> 00:01:28,460
diversity in modern organisms so you can

36
00:01:31,899 --> 00:01:29,960
look at what genes they have and when

37
00:01:33,399 --> 00:01:31,909
they evolved different pathways to use

38
00:01:35,679 --> 00:01:33,409

different things in the in the

39
00:01:38,320 --> 00:01:35,689
atmosphere and so that's what I'm

40
00:01:39,820 --> 00:01:38,330
working on and our idea was can we take

41
00:01:41,920 --> 00:01:39,830
this genomic data and use it as a

42
00:01:45,039 --> 00:01:41,930
totally independent test for these

43
00:01:47,140 --> 00:01:45,049
oxygen hypotheses and you might be

44
00:01:48,730 --> 00:01:47,150
wondering how to do that and we decided

45
00:01:50,230 --> 00:01:48,740
to sort of attack this question by

46
00:01:52,569 --> 00:01:50,240
looking at really informative gene

47
00:01:54,280 --> 00:01:52,579
histories so for the question of oxygen

48
00:01:56,260 --> 00:01:54,290
we can look at when genes that are

49
00:01:59,020 --> 00:01:56,270
involved in dealing with oxygen toxicity

50
00:02:00,999 --> 00:01:59,030
evolved so oxygen is really dangerous

51
00:02:03,940 --> 00:02:01,009
right it damages your cells that dam it

52
00:02:05,740 --> 00:02:03,950
causes mutations and you need to be able

53
00:02:09,190 --> 00:02:05,750
to deal with that if you're living in an

54
00:02:11,020 --> 00:02:09,200
oxygen vironment so these genes that can

55
00:02:13,060 --> 00:02:11,030
take superoxides and make them not

56
00:02:14,410 --> 00:02:13,070
destroy your organism are really

57
00:02:17,050 --> 00:02:14,420
important and there will be a really

58
00:02:19,000 --> 00:02:17,060
high selection pressure for any lineage

59
00:02:20,319 --> 00:02:19,010
that can deal with oxygen is going to

60
00:02:21,880 --> 00:02:20,329
survive much better than those that

61
00:02:24,820 --> 00:02:21,890
can't so we should be able to see that

62
00:02:26,350 --> 00:02:24,830
in the history of these genes

63
00:02:27,670 --> 00:02:26,360

similarly off

64
00:02:29,710 --> 00:02:27,680
in metabolism genes are really

65
00:02:32,020 --> 00:02:29,720
interesting because they're super useful

66
00:02:33,580 --> 00:02:32,030
if you have them you can use oxygen they

67
00:02:35,260 --> 00:02:33,590
might have even been required for

68
00:02:37,150 --> 00:02:35,270
complex life they're just really energy

69
00:02:39,730 --> 00:02:37,160
efficient but they do require more

70
00:02:41,740 --> 00:02:39,740
oxygen than you would expect to need to

71
00:02:43,900 --> 00:02:41,750
have to have these oxygen toxicity genes

72
00:02:45,010 --> 00:02:43,910
so the idea was maybe we'll see these

73
00:02:49,750 --> 00:02:45,020
genes rise and then the oxygen

74
00:02:52,060 --> 00:02:49,760
metabolism genes rise later and so we

75
00:02:53,890 --> 00:02:52,070
can look at which lineages acquire these

76
00:02:55,780 --> 00:02:53,900
genes and when they acquire them so you

77
00:02:57,760 --> 00:02:55,790
can look both in time and in like

78
00:03:01,660 --> 00:02:57,770
ecological space who is getting these

79
00:03:03,100 --> 00:03:01,670
genes when and so I'm going to do a

80
00:03:05,020 --> 00:03:03,110
really quick run-through of what

81
00:03:06,130 --> 00:03:05,030
molecular phylogenetics is just for

82
00:03:08,860 --> 00:03:06,140
those of you in the audience who might

83
00:03:10,449 --> 00:03:08,870
not have that as their specialty so

84
00:03:12,670 --> 00:03:10,459
basically what I'm doing is i acquire

85
00:03:15,370 --> 00:03:12,680
the amino acid sequences for whatever

86
00:03:17,680 --> 00:03:15,380
gene we're caring about from modern

87
00:03:19,360 --> 00:03:17,690
organisms and each letter is like an

88
00:03:21,130 --> 00:03:19,370

amino acid sequence and that is our data

89
00:03:22,660 --> 00:03:21,140
sets so eventually we're going to see

90
00:03:24,970 --> 00:03:22,670
how different these sequences are from

91
00:03:26,380 --> 00:03:24,980
each other you align them so that

92
00:03:28,510 --> 00:03:26,390
similar parts of the protein are being

93
00:03:30,640 --> 00:03:28,520
compared across species so I'm not like

94
00:03:32,500 --> 00:03:30,650
comparing you know this blue section to

95
00:03:34,479 --> 00:03:32,510
a purple section just because there's an

96
00:03:35,650 --> 00:03:34,489
insertion or something in the gene so we

97
00:03:37,300 --> 00:03:35,660
deal with that and then we create a

98
00:03:38,860 --> 00:03:37,310
phylogenetic tree and there's a bunch of

99
00:03:43,660 --> 00:03:38,870
ways to do that that I'm not going to go

100
00:03:46,420 --> 00:03:43,670
into now but in general our tips in the

101
00:03:50,680 --> 00:03:46,430
tree are going to be species the edges

102
00:03:52,090 --> 00:03:50,690
are longer if the nodes are more

103
00:03:53,920 --> 00:03:52,100
different from each other and the nodes

104
00:03:55,830 --> 00:03:53,930
represent speciation events

105
00:03:58,120 --> 00:03:55,840
so when lineages diverge from each other

106
00:04:00,009 --> 00:03:58,130
we also have support values in the

107
00:04:01,660 --> 00:04:00,019
topology of any tree that we put up and

108
00:04:03,729 --> 00:04:01,670
those support values in my analysis are

109
00:04:06,430 --> 00:04:03,739
done through boots droppings we take a

110
00:04:08,860 --> 00:04:06,440
sub sampling of that amino acid sequence

111
00:04:10,690 --> 00:04:08,870
create a tree with that that small

112
00:04:12,400 --> 00:04:10,700
section do it a hundred times and see

113
00:04:16,240 --> 00:04:12,410

how many of those times support this

114
00:04:18,820 --> 00:04:16,250
given topology so you can do this with

115
00:04:21,099 --> 00:04:18,830
some really slow evolving genes you can

116
00:04:22,480 --> 00:04:21,109
take like 30 ribosomal proteins stick

117
00:04:24,430 --> 00:04:22,490
them all together put them through this

118
00:04:27,010 --> 00:04:24,440
process and make a tree and that tree

119
00:04:29,050 --> 00:04:27,020
will generally really closely reflect

120
00:04:31,180 --> 00:04:29,060
actual evolutionary events so like if

121
00:04:32,680 --> 00:04:31,190
you look up here the cat next to the dog

122
00:04:34,390 --> 00:04:32,690
it's a bit further away from the mouse

123
00:04:37,290 --> 00:04:34,400
but further away from the reptiles and

124
00:04:38,580 --> 00:04:37,300
so on but if you look at just one gene

125
00:04:41,129 --> 00:04:38,590
it might

126
00:04:42,750 --> 00:04:41,139
different so in this example maybe the

127
00:04:44,790 --> 00:04:42,760
gene was lost on the lineage going to

128
00:04:46,860 --> 00:04:44,800
snakes maybe it evolved in the tetrapods

129
00:04:49,770 --> 00:04:46,870
the fish never had a chance to get it

130
00:04:50,940 --> 00:04:49,780
and you know weirder things can happen

131
00:04:52,890 --> 00:04:50,950
right you can have horizontal gene

132
00:04:54,300 --> 00:04:52,900
transfer events where you get the cat in

133
00:04:56,129 --> 00:04:54,310
the dog next to the bird and everybody

134
00:04:58,260 --> 00:04:56,139
knows that's not right

135
00:05:00,000 --> 00:04:58,270
and so you could say okay well if this

136
00:05:01,500 --> 00:05:00,010
is really highly supported they got the

137
00:05:03,330 --> 00:05:01,510
ancestor of cat and dog got the genes

138
00:05:04,470 --> 00:05:03,340

from the bird lineage and that might

139
00:05:06,000 --> 00:05:04,480
make sense to you but if you imagine

140
00:05:08,190 --> 00:05:06,010
these are all bacteria that you've never

141
00:05:10,110 --> 00:05:08,200
studied before you need that species

142
00:05:11,460 --> 00:05:10,120
tree to be able to compare the two same

143
00:05:13,080 --> 00:05:11,470
if you're a computer you need that

144
00:05:16,350 --> 00:05:13,090
species tree to be able to compare the

145
00:05:19,230 --> 00:05:16,360
two and you can also time transfers so

146
00:05:20,550 --> 00:05:19,240
in this kind of silly example you could

147
00:05:22,379 --> 00:05:20,560
say okay well we know birds and

148
00:05:24,090 --> 00:05:22,389
crocodiles split at this time and cats

149
00:05:26,250 --> 00:05:24,100
and dogs split at that time so this

150
00:05:28,440 --> 00:05:26,260
transfer event happened in that like 200

151
00:05:30,540 --> 00:05:28,450
million year window but if we apply this

152
00:05:32,879 --> 00:05:30,550
to all of history we might find

153
00:05:34,500 --> 00:05:32,889
something interesting so the first game

154
00:05:36,900 --> 00:05:34,510
we put through our pipeline is called

155
00:05:38,640 --> 00:05:36,910
superoxide disney at Ace this is one of

156
00:05:42,000 --> 00:05:38,650
those oxygen toxicity genes I was

157
00:05:45,140 --> 00:05:42,010
talking about and so we should see that

158
00:05:47,580 --> 00:05:45,150
as soon as a group of microbes

159
00:05:49,380 --> 00:05:47,590
experiences oxygen suddenly it really

160
00:05:50,760 --> 00:05:49,390
needs this gene and so any lineage that

161
00:05:52,440 --> 00:05:50,770
happens to get the gene is going to be

162
00:05:56,100 --> 00:05:52,450
really highly selected for they're going

163
00:05:57,900 --> 00:05:56,110

to live the rest of them maybe not but

164
00:05:59,550 --> 00:05:57,910
if we also if we ever see oxygen levels

165
00:06:01,260 --> 00:05:59,560
go back down we might expect to see

166
00:06:02,820 --> 00:06:01,270
losses because suddenly the gene isn't

167
00:06:06,750 --> 00:06:02,830
important anymore and we're optimizing

168
00:06:10,529 --> 00:06:06,760
for short genomes okay and this is like

169
00:06:11,760 --> 00:06:10,539
a quick idea so maybe when testing

170
00:06:15,180 --> 00:06:11,770
hypotheses about the great oxidation

171
00:06:18,990 --> 00:06:15,190
event we might see like two bump oops

172
00:06:20,779 --> 00:06:19,000
wrong button two bumps here of transfer

173
00:06:23,909 --> 00:06:20,789
events and that might be support for a

174
00:06:25,800 --> 00:06:23,919
rise at the great oxidation event steady

175
00:06:28,140 --> 00:06:25,810
state and then another rise at the

176
00:06:30,180 --> 00:06:28,150
neoproterozoic oxygenation event whereas

177
00:06:31,440 --> 00:06:30,190
if we see a whole bunch of losses right

178
00:06:33,360 --> 00:06:31,450
after the goe

179
00:06:35,370 --> 00:06:33,370
then maybe that would be a support for

180
00:06:38,100 --> 00:06:35,380
like oxygen levels go up and then they

181
00:06:40,650 --> 00:06:38,110
crash and then they go up again so that

182
00:06:42,719 --> 00:06:40,660
was the idea turns out that this is

183
00:06:44,909 --> 00:06:42,729
really really messy so my previous work

184
00:06:46,950 --> 00:06:44,919
was all on like eukaryotes that have

185
00:06:48,180 --> 00:06:46,960
just a few transfers are like okay we

186
00:06:48,980 --> 00:06:48,190
can just count them this will be a

187
00:06:51,379 --> 00:06:48,990
breeze

188
00:06:53,689 --> 00:06:51,389

it's not bacteria are sharing their

189
00:06:56,629 --> 00:06:53,699
genes all over the place and so we end

190
00:06:58,969 --> 00:06:56,639
up with this really complicated tree

191
00:07:02,450 --> 00:06:58,979
that infers like thousands of transfers

192
00:07:04,700 --> 00:07:02,460
events at really low support and so we

193
00:07:06,320 --> 00:07:04,710
needed some way to deal with this and

194
00:07:07,700 --> 00:07:06,330
the main problem is that we just have a

195
00:07:10,689 --> 00:07:07,710
short gene it's only a few hundred

196
00:07:13,279 --> 00:07:10,699
characters and we have so many sequences

197
00:07:15,469 --> 00:07:13,289
so we came up with this approach in

198
00:07:17,570 --> 00:07:15,479
which we identify so all of this is

199
00:07:19,909 --> 00:07:17,580
automated so once I do it once I should

200
00:07:22,219 --> 00:07:19,919
be able to do it a bunch of time it

201
00:07:23,960 --> 00:07:22,229
identifies someplace by taxonomic

202
00:07:25,850 --> 00:07:23,970
information so you can pull out like

203
00:07:26,990 --> 00:07:25,860
just this one creative cyanobacteria and

204
00:07:28,939 --> 00:07:27,000
there's another clade down there and

205
00:07:31,490 --> 00:07:28,949
we're not ignoring it but it's going to

206
00:07:33,920 --> 00:07:31,500
be processed separately then we create

207
00:07:36,260 --> 00:07:33,930
that ribosomal like species tree and

208
00:07:38,420 --> 00:07:36,270
also a gene tree for every single one of

209
00:07:40,999 --> 00:07:38,430
these subclades we boot them

210
00:07:44,240 --> 00:07:41,009
appropriately we include candidates to

211
00:07:46,219 --> 00:07:44,250
measure losses and then we can identify

212
00:07:48,080 --> 00:07:46,229
the transfers and losses by running it

213
00:07:50,420 --> 00:07:48,090

through another program like Ranger DPL

214
00:07:53,120 --> 00:07:50,430
and we do this across 100 different

215
00:07:55,430 --> 00:07:53,130
bootstrap tree topology is to sort of

216
00:07:57,409 --> 00:07:55,440
get the range of variation there to

217
00:07:59,089 --> 00:07:57,419
really make sure that if we're

218
00:08:03,020 --> 00:07:59,099
identifying a transfer it's definitely a

219
00:08:04,760 --> 00:08:03,030
transfer and then from each of these we

220
00:08:05,930 --> 00:08:04,770
can each of these like little subsample

221
00:08:09,560 --> 00:08:05,940
triangles we can take somewhere between

222
00:08:11,749 --> 00:08:09,570
my two and ten sequences and turn them

223
00:08:14,089 --> 00:08:11,759
into this slightly smaller like only a

224
00:08:16,430 --> 00:08:14,099
couple hundred sequences tree instead of

225
00:08:18,290 --> 00:08:16,440
that eight thousand one so we can see

226
00:08:20,540 --> 00:08:18,300
those really deep splits well with

227
00:08:22,640 --> 00:08:20,550
better support and this ends up just

228
00:08:24,200 --> 00:08:22,650
being a huge amount of data so you I can

229
00:08:25,820 --> 00:08:24,210
look at the subsample tree I can look at

230
00:08:27,379 --> 00:08:25,830
the node support on all these deeper

231
00:08:30,020 --> 00:08:27,389
slits but I can also look at every

232
00:08:32,949 --> 00:08:30,030
single one of these smaller triangles

233
00:08:34,639 --> 00:08:32,959
and I'm going to show some of those now

234
00:08:37,130 --> 00:08:34,649
because they're actually really

235
00:08:38,899 --> 00:08:37,140
interesting so for example in this is

236
00:08:41,420 --> 00:08:38,909
that first clade of cyanobacteria that I

237
00:08:42,469 --> 00:08:41,430
pulled out and you don't really need to

238
00:08:44,870 --> 00:08:42,479

understand this it's looking at the

239
00:08:46,940 --> 00:08:44,880
topology and the red branches are

240
00:08:51,350 --> 00:08:46,950
representing transfers into that lineage

241
00:08:52,819 --> 00:08:51,360
so here you can see that FOD has a

242
00:08:55,340 --> 00:08:52,829
really complex deep history in

243
00:08:57,500 --> 00:08:55,350
cyanobacteria we have a couple well

244
00:08:59,800 --> 00:08:57,510
supported transfers and something really

245
00:09:01,780 --> 00:08:59,810
interesting is that so Blio vector is

246
00:09:03,340 --> 00:09:01,790
supposedly like the earliest branching a

247
00:09:05,619 --> 00:09:03,350
sign of bacteria so we should have that

248
00:09:07,660 --> 00:09:05,629
coming out first but we don't if you

249
00:09:09,400 --> 00:09:07,670
look at it first there's a transfer from

250
00:09:10,840 --> 00:09:09,410
this group of Santa bacteria to this

251
00:09:13,720 --> 00:09:10,850
group and then there's a transfer from

252
00:09:15,819 --> 00:09:13,730
this group to a Siddal bacteria and then

253
00:09:20,079 --> 00:09:15,829
we have actor finally gets the gene in

254
00:09:20,410 --> 00:09:20,089
the transfer from a Siddal bacteria hold

255
00:09:24,220 --> 00:09:20,420
on

256
00:09:26,049 --> 00:09:24,230
and that's odd right if it all of

257
00:09:27,850 --> 00:09:26,059
cyanobacteria were evolving in an oxygen

258
00:09:30,009 --> 00:09:27,860
ik environment they would need this gene

259
00:09:31,660 --> 00:09:30,019
and so they would preserve that gene and

260
00:09:34,600 --> 00:09:31,670
you would get a topology like this but

261
00:09:36,610 --> 00:09:34,610
instead we get this topology and all of

262
00:09:38,549 --> 00:09:36,620
those deep splits are not vertically

263
00:09:41,230 --> 00:09:38,559

inherited so we can say okay well maybe

264
00:09:43,389 --> 00:09:41,240
they happened before that before

265
00:09:44,799 --> 00:09:43,399
wherever they were living was oxygenated

266
00:09:48,030 --> 00:09:44,809
right and that's and that's a pretty

267
00:09:50,949 --> 00:09:48,040
cool observation similarly on this is a

268
00:09:54,160 --> 00:09:50,959
subclade of crenarchaeota and you can

269
00:09:55,749 --> 00:09:54,170
see my gene wasn't these thermo protti

270
00:09:58,329 --> 00:09:55,759
allies and then it was transferred into

271
00:10:01,900 --> 00:09:58,339
some really modern groups right like so

272
00:10:06,069 --> 00:10:01,910
follow ballet and arrow PI room and

273
00:10:07,329 --> 00:10:06,079
those are the learnt like I think in the

274
00:10:09,819 --> 00:10:07,339
last 800 million years

275
00:10:12,280 --> 00:10:09,829
I'm pretty sure but I'm not 100% sure

276
00:10:14,980 --> 00:10:12,290
but so these are a bunch of different

277
00:10:17,199 --> 00:10:14,990
independent transfer events into aerobic

278
00:10:19,840 --> 00:10:17,209
archaea that diverged relatively

279
00:10:22,689 --> 00:10:19,850
recently so for example our of hiram is

280
00:10:24,490 --> 00:10:22,699
a deep ocean one and so we can say well

281
00:10:26,559 --> 00:10:24,500
maybe this is actually evidence for a

282
00:10:29,530 --> 00:10:26,569
delayed neoproterozoic oxygenation of

283
00:10:32,259 --> 00:10:29,540
the deep ocean because these guys didn't

284
00:10:33,759 --> 00:10:32,269
see oxygen and then suddenly they they

285
00:10:36,189 --> 00:10:33,769
see it they need to protect themselves

286
00:10:40,030 --> 00:10:36,199
from it and they end up acquiring this

287
00:10:41,559 --> 00:10:40,040
gene and then being selected for and

288
00:10:44,499 --> 00:10:41,569

that's that's basically all I have to

289
00:10:46,389 --> 00:10:44,509
say just that genomic data is a really

290
00:10:47,949 --> 00:10:46,399
like out-of-the-box way to attack these

291
00:10:49,179 --> 00:10:47,959
problems and it's in a totally

292
00:10:50,759 --> 00:10:49,189
independent way from all this wonderful

293
00:10:53,470 --> 00:10:50,769
geochemical work that you all are doing

294
00:10:54,999 --> 00:10:53,480
and we can look at specific clades to

295
00:10:57,280 --> 00:10:55,009
see what was going on ecologically and

296
00:11:00,100 --> 00:10:57,290
temporally and eventually once I get the

297
00:11:02,139 --> 00:11:00,110
the timing thing done hopefully we'll be

298
00:11:06,100 --> 00:11:02,149
able to see like patterns bumps of

299
00:11:07,449 --> 00:11:06,110
transfers and losses over time and yeah

300
00:11:08,980 --> 00:11:07,459
and that should be it should be faster

301
00:11:10,900 --> 00:11:08,990
now that I have this automated pipeline

302
00:11:12,639 --> 00:11:10,910
that I spent forever doing to

303
00:11:13,000 --> 00:11:12,649
automatically infer when these transfer

304
00:11:15,910 --> 00:11:13,010
events

305
00:11:17,740 --> 00:11:15,920
happening and that's all thank you to

306
00:11:19,330 --> 00:11:17,750
you all for inviting me here to speak

307
00:11:31,150 --> 00:11:19,340
thanks to my sponsors and to the

308
00:11:33,760 --> 00:11:31,160
Fournier lab and MIT oh I have a

309
00:11:36,700 --> 00:11:33,770
question yeah oh quit of course you guys

310
00:11:38,350 --> 00:11:36,710
how very nice talk I have so you have a

311
00:11:40,720 --> 00:11:38,360
you get a lot of superoxide dismutase

312
00:11:42,880 --> 00:11:40,730
--is i'm assuming not all of those have

313
00:11:44,350 --> 00:11:42,890

been tested for activity is there a

314
00:11:45,610 --> 00:11:44,360
possibility that some of them don't do

315
00:11:47,590 --> 00:11:45,620
that function that they do something

316
00:11:50,530 --> 00:11:47,600
else are you pretty confident that these

317
00:11:53,170 --> 00:11:50,540
are all doing superoxide dismutase and

318
00:11:55,030 --> 00:11:53,180
so everything that I've heard of when

319
00:11:57,460 --> 00:11:55,040
there is a superoxide it is doing some

320
00:11:58,870 --> 00:11:57,470
sort of superoxide dismutase activity

321
00:12:00,610 --> 00:11:58,880
I've never heard of one that's not but

322
00:12:02,050 --> 00:12:00,620
it's certainly possible and there are a

323
00:12:05,920 --> 00:12:02,060
couple different superoxide dismutase

324
00:12:07,300 --> 00:12:05,930
genes so so when I go a bit further in

325
00:12:08,650 --> 00:12:07,310
the project we'll have analyzed like a

326
00:12:10,300 --> 00:12:08,660
whole bunch of different genes and we

327
00:12:14,320 --> 00:12:10,310
can sort of take the whole history of

328
00:12:15,580 --> 00:12:14,330
all of them to not miss anything this is

329
00:12:16,660 --> 00:12:15,590
more of a technical point that I may

330
00:12:18,580 --> 00:12:16,670
have missed but you said that you were

331
00:12:20,050 --> 00:12:18,590
subsampling or tree in order to better

332
00:12:21,610 --> 00:12:20,060
resolve those clade you're doing gene

333
00:12:23,380 --> 00:12:21,620
species trees with an individual claves

334
00:12:25,390 --> 00:12:23,390
to look for transference what if there

335
00:12:26,920 --> 00:12:25,400
were transfers between really distant

336
00:12:29,590 --> 00:12:26,930
clays how are you catching that ID

337
00:12:32,260 --> 00:12:29,600
um so I'm taking every single so so for

338
00:12:34,060 --> 00:12:32,270

example in the UM thermo archaea one

339
00:12:35,740 --> 00:12:34,070
that I just showed I take one clade

340
00:12:37,990 --> 00:12:35,750
that's all of the thermo archaea and

341
00:12:41,380 --> 00:12:38,000
everything inside it just subsample on

342
00:12:44,560 --> 00:12:41,390
the algorithm ignores the annoncer mark

343
00:12:45,940 --> 00:12:44,570
yeah and then I have another plate that

344
00:12:47,410 --> 00:12:45,950
will be identified that's like just

345
00:12:48,640 --> 00:12:47,420
those social valleys and we'll so

346
00:12:50,290 --> 00:12:48,650
example them separately

347
00:12:53,560 --> 00:12:50,300
that way when you build the entire tree

348
00:12:55,300 --> 00:12:53,570
you should get um you know like to term

349
00:12:56,800 --> 00:12:55,310
archaea to coming out and to over here

350
00:12:58,720 --> 00:12:56,810
coming out and then the sister being

351
00:13:00,430 --> 00:12:58,730
those social Ballet's also coming out I

352
00:13:04,690 --> 00:13:00,440
hope that answers your question

353
00:13:07,360 --> 00:13:04,700
thanks very quickly yes a comment mostly

354
00:13:10,900 --> 00:13:07,370
about the cyanobacteria and the super

355
00:13:13,570 --> 00:13:10,910
rock dis musik it's a little pet theory

356
00:13:17,560 --> 00:13:13,580
of mind but I want to share it with you

357
00:13:19,750 --> 00:13:17,570
yes the sign of bacterial are one of the

358
00:13:22,180 --> 00:13:19,760
only groups of bacteria that use an

359
00:13:24,080 --> 00:13:22,190
oxidative desaturation mechanism for

360
00:13:27,260 --> 00:13:24,090
making their unsaturated fatty acid

361
00:13:30,980 --> 00:13:27,270
and I've often queried in my mind

362
00:13:33,110 --> 00:13:30,990
whether this wasn't a consequence of the

363
00:13:35,660 --> 00:13:33,120

fact that they had to initially deal

364
00:13:39,710 --> 00:13:35,670
with oxygen them sadly so you might put

365
00:13:41,060 --> 00:13:39,720
that that gene into your your yeah can

366
00:13:44,570 --> 00:13:41,070
you can do this repeat what it was like

367
00:13:47,060 --> 00:13:44,580
remember it's a desaturation oxidative

368
00:13:49,190 --> 00:13:47,070
saturate okay thank you that enzyme that

369
00:13:52,010 --> 00:13:49,200
puts the on the double bond and a fatty

370
00:13:54,460 --> 00:13:52,020
acid and it requires oxygen a nice point