Anne Farrell

Growth Temperature of the Last Common Ancestor of a Deeply-branching Bacterial Lineage

1
00:00:00,240 --> 00:00:10,900
[Music]

2
00:00:15,619 --> 00:00:13,369
so the optimum temperature of the last

3
00:00:18,560 --> 00:00:15,629
Universal common ancestor has been hotly

4
00:00:20,660 --> 00:00:18,570
debated over time there's been lots of

5
00:00:24,440 --> 00:00:20,670
lots of papers sort of that say it's hot

6
00:00:26,540 --> 00:00:24,450
say it's cold but we you know have very

7
00:00:28,010 --> 00:00:26,550
little access to actual and it any real

8
00:00:30,350 --> 00:00:28,020
data about what the physical

9
00:00:34,639 --> 00:00:30,360
characteristics of the environment we're

10
00:00:39,229 --> 00:00:34,649
for early life if we look at a cartoon

11
00:00:41,869 --> 00:00:39,239
diagram here of Tree of Life and look at

12
00:00:44,000 --> 00:00:41,879
extant temperatures a data that we can

13
00:00:46,759 --> 00:00:44,010

actually measure we noticed that

14
00:00:49,069 --> 00:00:46,769
temperature Optima across the tree is

15
00:00:50,630 --> 00:00:49,079
really diverse so we see sicker files

16
00:00:52,399 --> 00:00:50,640
all the way through hyperthermophiles

17
00:00:54,799 --> 00:00:52,409
search spread throughout the tree on

18
00:00:56,959 --> 00:00:54,809
multiple different branches which means

19
00:00:59,630 --> 00:00:56,969
that temperature Optima have evolved

20
00:01:02,299 --> 00:00:59,640
multiple times independently and this

21
00:01:03,529 --> 00:01:02,309
makes us ask the question how do how did

22
00:01:07,420 --> 00:01:03,539
that temperate how do those temperature

23
00:01:09,560 --> 00:01:07,430
Optima evolve and where did we come from

24
00:01:11,780 --> 00:01:09,570
one thing to notice about the tree of

25
00:01:14,210 --> 00:01:11,790
life is that if we look where the sort

26
00:01:16,160 --> 00:01:14,220
of the inferred root would be we have a

27
00:01:17,840 --> 00:01:16,170
lot of deeply branching

28
00:01:22,430 --> 00:01:17,850
hyperthermophilic and thermophilic

29
00:01:27,020 --> 00:01:22,440
groups here we often infer that we have

30
00:01:29,720 --> 00:01:27,030
to use steeply branching taxa on trees

31
00:01:31,610 --> 00:01:29,730
to infer ancestral States we assume that

32
00:01:35,780 --> 00:01:31,620
they've retained some of the ancestral

33
00:01:37,960 --> 00:01:35,790
characteristics over early life and that

34
00:01:40,580 --> 00:01:37,970
so we can use them as a proxy for that

35
00:01:44,740 --> 00:01:40,590
so given that so given that clustering

36
00:01:47,300 --> 00:01:44,750
is Luca was Luca a hot organism or not

37
00:01:49,000 --> 00:01:47,310
well there's three computational metrics

38
00:01:51,650 --> 00:01:49,010

and I only do computational value

39
00:01:53,720 --> 00:01:51,660
there's three computational ways that we

40
00:01:55,280 --> 00:01:53,730
can sort of infer temperature of

41
00:01:56,660 --> 00:01:55,290
organisms and I'm just going to go

42
00:02:00,590 --> 00:01:56,670
through those three matches pretty

43
00:02:02,450 --> 00:02:00,600
quickly the first is just doing looking

44
00:02:04,250 --> 00:02:02,460
at quantitative values doing

45
00:02:05,690 --> 00:02:04,260
quantitative trait prediction so one

46
00:02:08,059 --> 00:02:05,700
example of that would just be looking at

47
00:02:10,059 --> 00:02:08,069
parsimony parsimony means looking at the

48
00:02:11,619 --> 00:02:10,069
existing data that we have

49
00:02:13,360 --> 00:02:11,629
trying to come up with a model or an

50
00:02:16,080 --> 00:02:13,370
explanation that requires the fewest

51
00:02:18,369 --> 00:02:16,090
number of changes to arrive at that data

52
00:02:19,959 --> 00:02:18,379
except for ever loose evolution doesn't

53
00:02:22,690 --> 00:02:19,969
really work like that it rarely takes

54
00:02:24,819 --> 00:02:22,700
the simplest path so another way we can

55
00:02:27,490 --> 00:02:24,829
do this is by doing Bayesian inferences

56
00:02:29,860 --> 00:02:27,500
and that requires taking phylogenetic

57
00:02:32,699 --> 00:02:29,870
trees which we can do pretty you know

58
00:02:36,399 --> 00:02:32,709
pretty nicely today even at large scales

59
00:02:39,729 --> 00:02:36,409
and mapping the extant data that we have

60
00:02:42,909 --> 00:02:39,739
on that tree so for example temperature

61
00:02:45,699 --> 00:02:42,919
I give it a tree enteric a program tree

62
00:02:47,110 --> 00:02:45,709
in temperature and it considers the

63
00:02:49,599 --> 00:02:47,120

rates of evolution are you the branch

64
00:02:53,369 --> 00:02:49,609
lengths and we can infer what the

65
00:02:55,479 --> 00:02:53,379
ancestral nodes are on that tree a

66
00:02:57,789 --> 00:02:55,489
second method the computational method

67
00:03:01,229 --> 00:02:57,799
would be to look at the ribosomal RNA GC

68
00:03:07,000 --> 00:03:01,239
content so this is a little diagram of

69
00:03:10,420 --> 00:03:07,010
episomal stem and loop structure RNA

70
00:03:14,349 --> 00:03:10,430
stem and loop structure and since these

71
00:03:16,449 --> 00:03:14,359
are single-stranded pieces the GC bonds

72
00:03:19,179 --> 00:03:16,459
which are three hydrogen bonds provide a

73
00:03:21,550 --> 00:03:19,189
little extra stability compared to the

74
00:03:25,300 --> 00:03:21,560
two bonds of an au pair and so what we

75
00:03:29,110 --> 00:03:25,310
see is enrichment in the stems for GC in

76
00:03:30,999 --> 00:03:29,120
ribosomal RNA this is shown been shown

77
00:03:35,649 --> 00:03:31,009
in multiple papers to correlate really

78
00:03:37,420 --> 00:03:35,659
well with optimal growth temperature the

79
00:03:40,390 --> 00:03:37,430
third computational metric would be to

80
00:03:43,770 --> 00:03:40,400
look at the amino acid bias of a genome

81
00:03:45,939 --> 00:03:43,780
so in this case this particular example

82
00:03:48,339 --> 00:03:45,949
we might explode we might expect that

83
00:03:51,879 --> 00:03:48,349
amino acids are might differ between

84
00:03:53,619 --> 00:03:51,889
thermophiles and meso files in the same

85
00:03:54,999 --> 00:03:53,629
way that GC content might differ based

86
00:03:59,800 --> 00:03:55,009
on just the physical and chemical

87
00:04:02,499 --> 00:03:59,810
characteristics of the environment and

88
00:04:04,119 --> 00:04:02,509

so in 2007 zelda biologists did a

89
00:04:05,649 --> 00:04:04,129
statistical correlation to try to find

90
00:04:07,929 --> 00:04:05,659
amino acid sets that seemed to be

91
00:04:10,149 --> 00:04:07,939
correlated with temperature in this set

92
00:04:12,969 --> 00:04:10,159
of seven amino acids the IBEW round set

93
00:04:16,509 --> 00:04:12,979
was correlated well across bacteria in

94
00:04:18,909 --> 00:04:16,519
archaea so these three metrics have been

95
00:04:22,330 --> 00:04:18,919
used in again multiple papers to make

96
00:04:25,720 --> 00:04:22,340
arguments for and against a hot

97
00:04:28,600 --> 00:04:25,730
universe hot Lucca or a cold Lucca so

98
00:04:31,210 --> 00:04:28,610
for example in 1996 that it all uses 16s

99
00:04:33,730 --> 00:04:31,220
ribosomal RNA tree map these

100
00:04:36,250 --> 00:04:33,740
quantitative traits that you put you

101
00:04:37,600 --> 00:04:36,260
both may be these branches pulled where

102
00:04:39,730 --> 00:04:37,610
there were hyperthermophiles and

103
00:04:41,550 --> 00:04:39,740
therefore insertive had this person Moni

104
00:04:43,810 --> 00:04:41,560
as explanation and that since all the

105
00:04:48,120 --> 00:04:43,820
deeply branching lineages were

106
00:04:56,110 --> 00:04:52,590
another example is in 1999 Gaultier used

107
00:04:58,930 --> 00:04:56,120
20 this 23 s ribosomal RNA tree and

108
00:05:00,520 --> 00:04:58,940
looked at the GC content of stems and

109
00:05:04,720 --> 00:05:00,530
inferred that the last Universal common

110
00:05:05,980 --> 00:05:04,730
ancestor was a meso file but even in

111
00:05:07,300 --> 00:05:05,990
this model even in this with this

112
00:05:11,230 --> 00:05:07,310
conclusion we're looking at deeply

113
00:05:13,330 --> 00:05:11,240

branching lineages that are thermophilic

114
00:05:14,920 --> 00:05:13,340
so one one to notice is the thermal toga

115
00:05:18,159 --> 00:05:14,930
which is a the deepest trench in the

116
00:05:22,330 --> 00:05:18,169
bacterial group he used and is labeled

117
00:05:26,440 --> 00:05:22,340
as a vial and the final example is from

118
00:05:27,880 --> 00:05:26,450
2008 boo so at all looked at again last

119
00:05:30,760 --> 00:05:27,890
Universal common ancestors temperature

120
00:05:34,390 --> 00:05:30,770
using both 16s ribosomal RNA GC content

121
00:05:35,080 --> 00:05:34,400
and an amino acid bias metric and found

122
00:05:37,090 --> 00:05:35,090
in mesophilic

123
00:05:39,490 --> 00:05:37,100
concluded that it was an as a file for

124
00:05:42,390 --> 00:05:39,500
low temperature organism but again even

125
00:05:44,230 --> 00:05:42,400
in this tree we have deeply branching

126
00:05:47,140 --> 00:05:44,240
hyperthermophilic lineages like the

127
00:05:51,310 --> 00:05:47,150
thermic alga showing up so this is a

128
00:05:53,050 --> 00:05:51,320
consistent so again we keep seeing this

129
00:05:54,760 --> 00:05:53,060
group the thermo chugga Therma toga

130
00:05:57,990 --> 00:05:54,770
which we the phylum is now called a

131
00:06:02,290 --> 00:05:58,000
thermo tokido represents it as an early

132
00:06:07,390 --> 00:06:02,300
branching a deeply branching lineage on

133
00:06:09,040 --> 00:06:07,400
the Tree of Life these are deep

134
00:06:11,730 --> 00:06:09,050
biosphere bacteria as well so that's

135
00:06:21,100 --> 00:06:11,740
another sort of tangential interest to

136
00:06:23,350 --> 00:06:21,110
astrobiologists so the family to go to

137
00:06:26,200 --> 00:06:23,360
phylum is no longer just

138
00:06:28,960 --> 00:06:26,210

hyperthermophilic early isolates

139
00:06:30,969 --> 00:06:28,970
just were simply from the genus thermo

140
00:06:34,480 --> 00:06:30,979
tota which a lot of those papers were

141
00:06:35,860 --> 00:06:34,490
just using thermo toget taxa or

142
00:06:39,490 --> 00:06:35,870
taxer from the server to galleys which

143
00:06:40,900 --> 00:06:39,500
is this broader group here which are all

144
00:06:42,430 --> 00:06:40,910
hyperthermophiles so it makes sense to

145
00:06:45,400 --> 00:06:42,440
label that should that branch as

146
00:06:47,680 --> 00:06:45,410
hyperthermophilic but we now have

147
00:06:50,559 --> 00:06:47,690
temperatures ranging from 80 degrees all

148
00:06:55,629 --> 00:06:50,569
the way to some Meza files at 37 degrees

149
00:06:57,339 --> 00:06:55,639
so that's pretty diverse so if we want

150
00:06:59,140 --> 00:06:57,349
to keep using this it is a deeply

151
00:07:02,320 --> 00:06:59,150
branching lineage that's true that's

152
00:07:04,390 --> 00:07:02,330
been established but if you want to keep

153
00:07:06,879 --> 00:07:04,400
sort of inferring it using the branch

154
00:07:08,710 --> 00:07:06,889
and categorizing it is one thing it

155
00:07:11,350 --> 00:07:08,720
might be best if we got a better

156
00:07:13,990 --> 00:07:11,360
estimate for the temperature of the

157
00:07:15,339 --> 00:07:14,000
ancestor of this phylum so that's what

158
00:07:16,870 --> 00:07:15,349
I'm going to talk today not about Luka

159
00:07:21,159 --> 00:07:16,880
necessarily but about the temperature of

160
00:07:23,980 --> 00:07:21,169
this phylum so in 2009 my adviser Olga's

161
00:07:27,210 --> 00:07:23,990
active a Ava took the data that she had

162
00:07:32,379 --> 00:07:27,220
available which again was mostly

163
00:07:34,719 --> 00:07:32,389

organisms from this clade and used

164
00:07:36,460 --> 00:07:34,729
genomic composition metrics to estimate

165
00:07:39,040 --> 00:07:36,470
an optimum growth temperature of 80

166
00:07:41,020 --> 00:07:39,050
degrees Celsius for the node with a red

167
00:07:43,659 --> 00:07:41,030
star here

168
00:07:45,339 --> 00:07:43,669
it makes again parsimonious sense that

169
00:07:47,499 --> 00:07:45,349
these are all hyperthermophiles so that

170
00:07:51,219 --> 00:07:47,509
this new hyperthermophiles that seemed

171
00:07:55,600 --> 00:07:51,229
great at the time in 2013 and a green at

172
00:07:58,570 --> 00:07:55,610
all had an updated set of data and used

173
00:08:00,760 --> 00:07:58,580
16s ribosomal RNA to estimate the

174
00:08:02,350 --> 00:08:00,770
temperature for this node here so it's

175
00:08:05,920 --> 00:08:02,360
sort of a new ancestral node of the

176
00:08:06,219 --> 00:08:05,930
phylum to be 76 degrees Celsius a little

177
00:08:08,110 --> 00:08:06,229
lower

178
00:08:12,010 --> 00:08:08,120
still firma file but it's a little bit

179
00:08:14,170 --> 00:08:12,020
more of a moderate thermal file but

180
00:08:17,830 --> 00:08:14,180
since Ana even did her paper in 2013

181
00:08:19,059 --> 00:08:17,840
which is not going on six years ago we

182
00:08:24,070 --> 00:08:19,069
have had it a whole bunch of new

183
00:08:24,969 --> 00:08:24,080
saratoga this list is not exhaustive it

184
00:08:26,800 --> 00:08:24,979
doesn't really matter what their names

185
00:08:28,629 --> 00:08:26,810
are it just matters that we keep adding

186
00:08:30,189 --> 00:08:28,639
temperature diversity we're not just

187
00:08:31,809 --> 00:08:30,199
adding a single temperature to this tree

188
00:08:35,190 --> 00:08:31,819

we're adding temperatures ranging from

189
00:08:38,260 --> 00:08:35,200
79 degrees Celsius to 45 degrees Celsius

190
00:08:41,860 --> 00:08:38,270
and one particular organism of interest

191
00:08:43,779 --> 00:08:41,870
that has been added since this last this

192
00:08:46,639 --> 00:08:43,789
laughs analysis with them is mezu

193
00:08:50,760 --> 00:08:46,649
acidity total Owens's

194
00:08:54,060 --> 00:08:50,770
so mezzo acidity oka is up here at the

195
00:08:55,079 --> 00:08:54,070
top of the tree and when I rotate the

196
00:08:56,639 --> 00:08:55,089
tree in a minute it's gonna be on the

197
00:08:57,720 --> 00:08:56,649
left so I'm just gonna rotate this tree

198
00:09:00,720 --> 00:08:57,730
I'm gonna keep using the same tree but

199
00:09:05,940 --> 00:09:00,730
rotate it kind of clockwise so mezzo

200
00:09:06,810 --> 00:09:05,950
saratoga is awesome it's it was found in

201
00:09:09,240 --> 00:09:06,820
a hydrothermal vent

202
00:09:11,940 --> 00:09:09,250
Wow spreading better you guys probably

203
00:09:15,090 --> 00:09:11,950
know more about that thank you so the

204
00:09:16,769 --> 00:09:15,100
deep-sea organism and it is deeply

205
00:09:19,230 --> 00:09:16,779
branching and that's true when we do

206
00:09:21,420 --> 00:09:19,240
ribosomal protein phylogeny x' and 16s

207
00:09:23,519 --> 00:09:21,430
ribosomal RNA phylogenies as well as a

208
00:09:25,199 --> 00:09:23,529
bunch of other genes so it seems to be

209
00:09:27,420 --> 00:09:25,209
pretty well established as deeply

210
00:09:28,530 --> 00:09:27,430
branching I can talk more about that

211
00:09:31,710 --> 00:09:28,540
later

212
00:09:33,210 --> 00:09:31,720
and since as people you branching we

213
00:09:34,710 --> 00:09:33,220

might assume again that it would retain

214
00:09:37,949 --> 00:09:34,720
some characteristics of early life in

215
00:09:39,449 --> 00:09:37,959
this violin um but it's actually a

216
00:09:41,850 --> 00:09:39,459
moderate thermophiles it's still

217
00:09:44,550 --> 00:09:41,860
thermofoil but just barely at 58 degrees

218
00:09:46,050 --> 00:09:44,560
Celsius so much lower than we would have

219
00:09:48,600 --> 00:09:46,060
anticipated given its deep blue

220
00:09:50,640 --> 00:09:48,610
branching placement so the question is

221
00:09:52,949 --> 00:09:50,650
does the addition of this mezzo city

222
00:09:55,560 --> 00:09:52,959
galleys change the predicted optimum

223
00:09:57,930 --> 00:09:55,570
temperature of the phylum of the violins

224
00:09:59,280 --> 00:09:57,940
last common ancestor and again I use the

225
00:10:03,000 --> 00:09:59,290
three methods that I've already gone

226
00:10:05,490 --> 00:10:03,010
over so what I did quantitative trait

227
00:10:06,900 --> 00:10:05,500
estimation I use a program called bass

228
00:10:10,170 --> 00:10:06,910
traits which I'm happy to talk to people

229
00:10:13,230 --> 00:10:10,180
about it interested later and again I

230
00:10:15,630 --> 00:10:13,240
provided it with a reliable phylogeny

231
00:10:19,110 --> 00:10:15,640
this is a ribosomal protein phylogeny hi

232
00:10:20,880 --> 00:10:19,120
bootstrap support and all the extant

233
00:10:23,970 --> 00:10:20,890
temperature data that we have

234
00:10:28,050 --> 00:10:23,980
experimental verified in the lab based

235
00:10:29,250 --> 00:10:28,060
rates infers the ancestral knows the

236
00:10:32,670 --> 00:10:29,260
temperatures of the necessarily knows

237
00:10:34,760 --> 00:10:32,680
and inferred that the temp the optimum

238
00:10:37,710 --> 00:10:34,770

temperature of the last common ancestor

239
00:10:41,069 --> 00:10:37,720
was around 65 degrees Celsius so this is

240
00:10:42,750 --> 00:10:41,079
lower than the past two estimates the

241
00:10:46,740 --> 00:10:42,760
previous estimates were for this node

242
00:10:49,949 --> 00:10:46,750
and this node so that was kind of

243
00:10:51,870 --> 00:10:49,959
interesting we also see that

244
00:10:54,210 --> 00:10:51,880
hyperthermophilic had to arise and

245
00:10:56,490 --> 00:10:54,220
cooling mezzo philly or just a general

246
00:10:58,569 --> 00:10:56,500
cooling down of certain branches also

247
00:10:59,979 --> 00:10:58,579
rose separately so we see sort of we

248
00:11:04,090 --> 00:10:59,989
one general temperature trend in this

249
00:11:07,210 --> 00:11:04,100
violin we definitely see variation when

250
00:11:09,009 --> 00:11:07,220
I use the ribosomal RNA GC content I

251
00:11:10,840 --> 00:11:09,019
first checked that the thermit ago today

252
00:11:13,269 --> 00:11:10,850
data still correlates with temperature

253
00:11:16,689 --> 00:11:13,279
so this is just thermo to go to 16s

254
00:11:20,079 --> 00:11:16,699
ribosomal RNA stems has a nice

255
00:11:22,960 --> 00:11:20,089
correlation temperature with temperature

256
00:11:27,280 --> 00:11:22,970
on the x-axis and the stem GC content on

257
00:11:29,019 --> 00:11:27,290
the y-axis and then I so then I so that

258
00:11:30,789 --> 00:11:29,029
was you get the sequences and I want to

259
00:11:33,189 --> 00:11:30,799
first infer the ancestral sequences for

260
00:11:35,139 --> 00:11:33,199
each node calculate the GC content and

261
00:11:38,859 --> 00:11:35,149
then use the GC content to estimate

262
00:11:42,519 --> 00:11:38,869
temperature when I did that I recessed

263
00:11:45,460 --> 00:11:42,529

a'mma to Greece Celsius for the last

264
00:11:47,530 --> 00:11:45,470
common ancestor of this violin and again

265
00:11:49,900 --> 00:11:47,540
hyperthermophilic had to arise at one

266
00:11:53,199 --> 00:11:49,910
point and cooling down of other branches

267
00:11:54,639 --> 00:11:53,209
had to occur separately but again this

268
00:11:56,460 --> 00:11:54,649
is pretty close to that base trait

269
00:11:59,410 --> 00:11:56,470
estimate 65 degrees

270
00:12:00,879 --> 00:11:59,420
the final metric that I the final

271
00:12:03,220 --> 00:12:00,889
computational method that I used was

272
00:12:06,309 --> 00:12:03,230
looking at the amino acid bias and I

273
00:12:09,249 --> 00:12:06,319
used that I've URL amino acid set and

274
00:12:10,780 --> 00:12:09,259
again I first checked that in the genes

275
00:12:12,759 --> 00:12:10,790
that I used in the from the thermo

276
00:12:16,419 --> 00:12:12,769
Takota specifically that there is a

277
00:12:17,799 --> 00:12:16,429
correlation with temperature and so

278
00:12:21,100 --> 00:12:17,809
temperatures on the x-axis and the

279
00:12:25,720 --> 00:12:21,110
fraction of IBL in medium fraction of

280
00:12:27,189 --> 00:12:25,730
IBL in the genes is on the y-axis I had

281
00:12:28,989 --> 00:12:27,199
to use a pretty small data set for this

282
00:12:31,679 --> 00:12:28,999
analysis because I'm still working on

283
00:12:34,629 --> 00:12:31,689
cleaning up gene treaties and so I used

284
00:12:38,079 --> 00:12:34,639
175 homologous genes that are shared

285
00:12:39,429 --> 00:12:38,089
across the across all taxa these

286
00:12:42,759 --> 00:12:39,439
organisms have genomes that are about

287
00:12:46,359 --> 00:12:42,769
2,000 genes long so 175 out of 2000s a

288
00:12:48,280 --> 00:12:46,369

pretty small subset but because they

289
00:12:49,539 --> 00:12:48,290
have a huge pan genome there's a there's

290
00:12:51,669 --> 00:12:49,549
a lot of genes that some of them have

291
00:12:54,879 --> 00:12:51,679
something to do I can talk more about

292
00:12:56,530 --> 00:12:54,889
that separately so I had to use a pretty

293
00:13:00,579 --> 00:12:56,540
small subset so I did check that the

294
00:13:02,859 --> 00:13:00,589
distribution of I've URL in the 175

295
00:13:04,869 --> 00:13:02,869
genes that I selected match the

296
00:13:08,470 --> 00:13:04,879
distribution of IBL in whole genome ax

297
00:13:09,900 --> 00:13:08,480
sets so for example in giotto go petraea

298
00:13:11,610 --> 00:13:09,910
which is one organism

299
00:13:16,080 --> 00:13:11,620
look at the IV rail fraction on the

300
00:13:18,690 --> 00:13:16,090
x-axis here and the blue line the gene

301
00:13:21,300 --> 00:13:18,700
subset matches the whole genome

302
00:13:23,400 --> 00:13:21,310
distribution pretty pretty pretty

303
00:13:25,440 --> 00:13:23,410
perfectly and that held true for all the

304
00:13:26,610 --> 00:13:25,450
data I was able to use so I felt the

305
00:13:30,800 --> 00:13:26,620
hundred and seventy-five genes were

306
00:13:33,000 --> 00:13:30,810
reliable proxy for whole genome analysis

307
00:13:35,270 --> 00:13:33,010
and the reconstruction of one hundred

308
00:13:37,920 --> 00:13:35,280
and seventy-five ancestral sequences

309
00:13:39,780 --> 00:13:37,930
provided a median IPL of about forty

310
00:13:41,460 --> 00:13:39,790
four point five percent and this value

311
00:13:48,540 --> 00:13:41,470
reflected a temperature of 73 degrees

312
00:13:50,970 --> 00:13:48,550
Celsius for the ancestor so finally in

313
00:13:53,010 --> 00:13:50,980

summary I have three different values

314
00:13:55,850 --> 00:13:53,020
that I was able to estimate 65 degrees

315
00:13:58,770 --> 00:13:55,860
63 degrees in a bit higher at 73 degrees

316
00:14:01,110 --> 00:13:58,780
these are all lower than the previous

317
00:14:04,800 --> 00:14:01,120
two estimates from Zack sadaiva and

318
00:14:06,600 --> 00:14:04,810
green at all for those earlier nodes so

319
00:14:09,090 --> 00:14:06,610
adding this data point as well as some

320
00:14:11,310 --> 00:14:09,100
other data in the tree seems to have

321
00:14:14,130 --> 00:14:11,320
lowered our estimates it's pasta it's

322
00:14:16,230 --> 00:14:14,140
still a thermophilic it still is ever

323
00:14:17,580 --> 00:14:16,240
feel like ancestor so we can still say

324
00:14:19,170 --> 00:14:17,590
it's a thermo file but it might actually

325
00:14:24,390 --> 00:14:19,180
be more of a moderate thermo file then

326
00:14:26,310 --> 00:14:24,400
we have previously assumed we'd keep

327
00:14:28,200 --> 00:14:26,320
seeing all these models the evolution of

328
00:14:30,480 --> 00:14:28,210
both hyperthermophilic and mesophilic

329
00:14:32,100 --> 00:14:30,490
were secondary so we actually see

330
00:14:34,650 --> 00:14:32,110
temperature fluctuations happening a lot

331
00:14:36,540 --> 00:14:34,660
over time and our findings are in

332
00:14:39,000 --> 00:14:36,550
concordance with several analyses that

333
00:14:40,590 --> 00:14:39,010
argue that possibly early life wasn't

334
00:14:46,020 --> 00:14:40,600
quite as hyperthermophilic as we

335
00:14:48,690 --> 00:14:46,030
initially wanted to to believe and with

336
00:14:52,050 --> 00:14:48,700
that i'd like to thank my amazing

337
00:14:57,020 --> 00:14:52,060
adviser Olga's Activia who's great the

338
00:14:58,830 --> 00:14:57,030

lab of oz all of our lab members and our

339
00:15:00,810 --> 00:14:58,840
collaborator camilla des Beaux who's

340
00:15:03,000 --> 00:15:00,820
worked a lot with Meza togas these

341
00:15:06,030 --> 00:15:03,010
moderate temperature organisms Simon's

342
00:15:07,770 --> 00:15:06,040
foundation for funding ecology evolution

343
00:15:09,870 --> 00:15:07,780
ecosystems and society program at

344
00:15:11,910 --> 00:15:09,880
Dartmouth which is a great program and

345
00:15:14,800 --> 00:15:11,920
well being there and Dartmouth for all

346
00:15:21,920 --> 00:15:14,810
our funding as well mm-hmm Thanks

347
00:15:21,930 --> 00:15:27,170
I'm happy okay questions yeah

348
00:15:33,269 --> 00:15:30,090
hi thank you uh for the talk I'm just

349
00:15:35,189 --> 00:15:33,279
curious are there any like geological

350
00:15:37,050 --> 00:15:35,199
timescales attached to these predictions

351
00:15:38,629 --> 00:15:37,060
so I have not done that I've gotten that

352
00:15:42,629 --> 00:15:38,639
question a few times and I have not done

353
00:15:44,429 --> 00:15:42,639
like um you know any kind of dating if

354
00:15:46,199 --> 00:15:44,439
someone here is an expert at bacterial

355
00:15:47,999 --> 00:15:46,209
microfossils and let's talk me through

356
00:15:50,670 --> 00:15:48,009
sort of what I could maybe do with this

357
00:15:53,100 --> 00:15:50,680
I'd be interested in that but that's not

358
00:15:54,269 --> 00:15:53,110
my area of expertise and doing something

359
00:15:55,800 --> 00:15:54,279
like applying a molecular clock

360
00:15:58,650 --> 00:15:55,810
assumption is probably gonna give a

361
00:15:59,850 --> 00:15:58,660
really crummy that's me anyways so I

362
00:16:04,939 --> 00:15:59,860
haven't that hasn't been worked my time

363
00:16:07,379 --> 00:16:04,949

or anything like that really love I

364
00:16:10,170 --> 00:16:07,389
really love this talk thank you thanks

365
00:16:13,470 --> 00:16:10,180
um I was really curious about the IV L

366
00:16:15,749 --> 00:16:13,480
amino acid bias I was wondering if you

367
00:16:20,699 --> 00:16:15,759
were aware of any driving logic behind

368
00:16:23,129 --> 00:16:20,709
that - yeah so I'm here L is definitely

369
00:16:24,629 --> 00:16:23,139
it was purely based on Cisco correlation

370
00:16:26,040 --> 00:16:24,639
like they went through all these sets of

371
00:16:31,259 --> 00:16:26,050
amino acids it were like which one's the

372
00:16:33,120 --> 00:16:31,269
best this is the best but it does

373
00:16:34,590 --> 00:16:33,130
contain a lot of a couple charged amino

374
00:16:36,120 --> 00:16:34,600
acids and we do know that actually

375
00:16:37,650 --> 00:16:36,130
charge versus polar amino acids or

376
00:16:39,900 --> 00:16:37,660
another really strong amino acid

377
00:16:43,620 --> 00:16:39,910
correlate in genome so there's some

378
00:16:45,329 --> 00:16:43,630
biological significance there um I've

379
00:16:49,290 --> 00:16:45,339
also heard that it's inexpensive amino

380
00:16:54,120 --> 00:16:49,300
acids set someone who's better and can

381
00:16:56,040 --> 00:16:54,130
probably verify that and even though

382
00:16:57,329 --> 00:16:56,050
it's an expensive amino acid set we

383
00:16:58,920 --> 00:16:57,339
think it's probably there's been some

384
00:17:01,620 --> 00:16:58,930
hypotheses that it's better to use that

385
00:17:02,759 --> 00:17:01,630
set rather than deal with like producing

386
00:17:05,159 --> 00:17:02,769
a lot of chaperones to deal with

387
00:17:17,610 --> 00:17:05,169
denatured and miss folded proteins so

388
00:17:22,620 --> 00:17:19,769

anyways I have the question so this deep

389
00:17:24,299 --> 00:17:22,630
branching organism you said that the

390
00:17:26,939 --> 00:17:24,309
phylogeny is that you have here a pretty

391
00:17:29,490 --> 00:17:26,949
high support is it consider percent

392
00:17:32,190 --> 00:17:29,500
consistent that this deep branch or ends

393
00:17:35,580 --> 00:17:32,200
up as the out group this will clean yeah

394
00:17:37,080 --> 00:17:35,590
so if I look at my ribosomal protein

395
00:17:38,850 --> 00:17:37,090
tree

396
00:17:40,560 --> 00:17:38,860
it's a wrapper so it's a concatenated

397
00:17:43,529 --> 00:17:40,570
ribosomal protein tree which often

398
00:17:45,330 --> 00:17:43,539
employed support values highly so the

399
00:17:47,880 --> 00:17:45,340
only two nodes that are pretty poor are

400
00:17:49,350 --> 00:17:47,890
below ninety five four hundred straps

401
00:17:51,510 --> 00:17:49,360
are these two nodes you can kind of see

402
00:17:52,830 --> 00:17:51,520
them in a few circles there so otherwise

403
00:17:54,840 --> 00:17:52,840
everything's above ninety five right

404
00:17:57,210 --> 00:17:54,850
it's really really well supported

405
00:17:58,769 --> 00:17:57,220
but that could possibly be an artifact

406
00:18:04,320 --> 00:17:58,779
so we definitely we're concerned about

407
00:18:06,269 --> 00:18:04,330
that if we look at the 16s tree so this

408
00:18:07,500 --> 00:18:06,279
is a similar tree I didn't draw your

409
00:18:10,260 --> 00:18:07,510
attention to it but actually instead of

410
00:18:13,560 --> 00:18:10,270
having one leaf here actually up to so

411
00:18:15,899 --> 00:18:13,570
we have at least one of the described

412
00:18:18,570 --> 00:18:15,909
organism with temperature with 16s data

413
00:18:21,440 --> 00:18:18,580

available which breaks that branch and

414
00:18:23,430 --> 00:18:21,450
it's still an out and long branch

415
00:18:25,769 --> 00:18:23,440
additionally when we break that branch

416
00:18:29,159 --> 00:18:25,779
further with more environmental clones

417
00:18:32,610 --> 00:18:29,169
it stays where it is we also have other

418
00:18:33,720 --> 00:18:32,620
genome like metagenomes data that helps

419
00:18:36,389 --> 00:18:33,730
break that branch and then I'm also

420
00:18:37,409 --> 00:18:36,399
doing an analysis of which genes I'm

421
00:18:39,389 --> 00:18:37,419
saying and where those genes are

422
00:18:41,190 --> 00:18:39,399
grouping with other organisms right now

423
00:18:43,710 --> 00:18:41,200
so I'm trying to establish that it's

424
00:18:48,360 --> 00:18:43,720
really belong that it's truly a deeply

425
00:18:50,789 --> 00:18:48,370
branching lineage all right I'm just

426
00:18:53,659 --> 00:18:50,799
went to culinary briefly on the on the

427
00:18:58,200 --> 00:18:53,669
question of dating yeah these lineages

428
00:19:01,260 --> 00:18:58,210
given that there is going to be I would

429
00:19:03,330 --> 00:19:01,270
imagine no micro fossils available from

430
00:19:06,060 --> 00:19:03,340
this particular Python this will have to

431
00:19:08,149 --> 00:19:06,070
depend almost entirely on finding

432
00:19:10,560 --> 00:19:08,159
horizontal gene transfer events from

433
00:19:12,600 --> 00:19:10,570
lineages where the fossil records such

434
00:19:14,190 --> 00:19:12,610
as cyanobacteria or some hearing that

435
00:19:15,600 --> 00:19:14,200
and I think that's something that's

436
00:19:18,899 --> 00:19:15,610
actually easy to look at have there been

437
00:19:21,180 --> 00:19:18,909
identified or Aldean transfer events

438
00:19:24,029 --> 00:19:21,190

into this phylum which would then allow

439
00:19:25,710 --> 00:19:24,039
to propagate the dating constraints from

440
00:19:26,480 --> 00:19:25,720
a lineage which does have fossil

441
00:19:29,690 --> 00:19:26,490
calibration

442
00:19:32,090 --> 00:19:29,700
okay so first of all I can say for sure

443
00:19:33,710 --> 00:19:32,100
that this phylum has undergone becomes a

444
00:19:34,850 --> 00:19:33,720
horizontal gene transfer for a while

445
00:19:36,290 --> 00:19:34,860
they thought maybe it was actually in

446
00:19:38,600 --> 00:19:36,300
archaea because that's so many a pale

447
00:19:39,620 --> 00:19:38,610
genes and then they thought maybe they

448
00:19:41,150 --> 00:19:39,630
weren't even a file open

449
00:19:42,710 --> 00:19:41,160
they were they were Firmicutes because

450
00:19:44,710 --> 00:19:42,720
there's so many from acute gain so lots

451
00:19:46,490 --> 00:19:44,720
of gene transfer terms of cyanobacteria

452
00:19:48,140 --> 00:19:46,500
specifically I'm not sure but I would

453
00:19:54,380 --> 00:19:48,150
love to talk to you more about that we

454
00:19:58,930 --> 00:19:54,390
should not we should be I'm gonna ask a

455
00:20:02,180 --> 00:19:58,940
quick question so because I'm more of

456
00:20:03,799 --> 00:20:02,190
ecologists I'm just kind of wondering so

457
00:20:05,570 --> 00:20:03,809
you know we kind of all have this idea

458
00:20:08,510 --> 00:20:05,580
of Luca being in vents and this really

459
00:20:11,270 --> 00:20:08,520
hot thing so if it's not an event could

460
00:20:13,160 --> 00:20:11,280
you speculate and you could go as wild

461
00:20:15,080 --> 00:20:13,170
as you want here but like on the type of

462
00:20:19,730 --> 00:20:15,090
environment that this Musa philic Luca

463
00:20:22,880 --> 00:20:19,740

would live in oh gosh I feel like I feel

464
00:20:26,210 --> 00:20:22,890
like it's way more rich in this cabin I