

1

0:00:00,000 --> 0:00:02.000

Hi everyone.

2

0:00:02,000 --> 0:00:04.000

I hope you're doing well. My name is Christopher Good

3

0:00:04,000 --> 0:00:06.000

and I am really excited to talk to you today

4

0:00:06,000 --> 0:00:08.000

about my Honors Capstone Project,

5

0:00:08,000 --> 0:00:10.000

the title of which is "Using

6

0:00:10,000 --> 0:00:12.000

BLAST to detect Horizontal Gene Transfer

7

0:00:12,000 --> 0:00:14.000

in Pathogenic Fungi".

8

0:00:14,000 --> 0:00:16.000

And I completed this project under the direction of Dr. Theresa Grana

9

0:00:16,827 --> 0:00:18.827

in the Biology Department.

10

0:00:18,946 --> 0:00:20.946

In terms of the rationale, my interest was

11

0:00:20,946 --> 0:00:22.946

definitely peaked in my

12

0:00:22,946 --> 0:00:24.946

Bioinformatics course with Dr. Grana,

13

0:00:25,064 --> 0:00:27.064

when we participated in a community science project

14

0:00:27,064 --> 0:00:29.064

developed by a team called Genome Solver

15

0:00:29,173 --> 0:00:31.173

and they have used

16

0:00:31,173 --> 0:00:33.173

a python code to

17

0:00:33,173 --> 0:00:35.173

develop this

18

0:00:35,173 --> 0:00:37.173

BLAST, or Basic

19

0:00:37,173 --> 0:00:39.173

Local Alignment Search Tool Pipeline

20

0:00:39,173 --> 0:00:41.173

in order to be able to detect potential

21

0:00:41,173 --> 0:00:43.173

instances of Horizontal Gene Transfer.

22

0:00:43,173 --> 0:00:45.173

And in this project,

23

0:00:45,173 --> 0:00:47.173

we were looking at horizontal gene transfer

24

0:00:47,173 --> 0:00:49.173

between bacteria

25

0:00:49,173 --> 0:00:51.173

and bacteriophages.

26

0:00:51,173 --> 0:00:53.173

Then later in the course,

27

0:00:53,173 --> 0:00:55.173

we read a research article

28

0:00:55,173 --> 0:00:57.173  
about "Microbial Dark Matter"

29

0:00:57,173 --> 0:00:59.173  
and then in the

30

0:00:59,173 --> 0:01:01.173  
context of this article,

31

0:01:01,548 --> 0:01:03.548  
it was about

32

0:01:03,548 --> 0:01:05.548  
the failure to identify causative

33

0:01:05,548 --> 0:01:07.548  
pathogens in cases of neonatal

34

0:01:07,548 --> 0:01:09.548  
sepsis,

35

0:01:09,570 --> 0:01:11.570  
this inflammatory and immunosuppressive

36

0:01:11,570 --> 0:01:13.570  
condition resulting from

37

0:01:13,570 --> 0:01:15.570  
untreated infection.

38

0:01:15,570 --> 0:01:17.570  
And the article really described

39

0:01:17,570 --> 0:01:19.570  
how scientists

40

0:01:19,570 --> 0:01:21.570  
are currently using metagenomic

41

0:01:21,570 --> 0:01:23.570

analytical tools to further our

42

0:01:23,570 --> 0:01:25.570  
understanding of

43

0:01:25,570 --> 0:01:27.570  
microorganisms

44

0:01:27,570 --> 0:01:29.570  
that are pathogenic to humans.

45

0:01:29,628 --> 0:01:31.628  
And that definitely peaked my interest

46

0:01:31,628 --> 0:01:33.628  
because

47

0:01:33,628 --> 0:01:35.628  
throughout my studies in biology,

48

0:01:35,628 --> 0:01:37.628  
as a graduating senior now,

49

0:01:37,628 --> 0:01:39.628  
we've talked about microbiomes a few times

50

0:01:39,628 --> 0:01:41.628  
and I've always found it

51

0:01:41,628 --> 0:01:43.628  
really interesting because

52

0:01:43,628 --> 0:01:45.628  
the composition of microorganisms that

53

0:01:45,628 --> 0:01:47.628  
inhabit different parts of the body,

54

0:01:47,686 --> 0:01:49.686  
they play important

55

0:01:49,686 --> 0:01:51.686  
physiological roles and

56

0:01:51,686 --> 0:01:53.686  
important roles in a number of

57

0:01:53,686 --> 0:01:55.686  
processes like immunity.

58

0:01:55,686 --> 0:01:57.686  
So, as I did more background research

59

0:01:57,686 --> 0:01:59.686  
on cases of sepsis, when

60

0:01:59,686 --> 0:02:01.686  
the causative pathogen

61

0:02:01,686 --> 0:02:03.686  
can be identified, it seems

62

0:02:03,686 --> 0:02:05.686  
the majority are bacterial. Of course,

63

0:02:05,686 --> 0:02:07.686  
there's also a fungal

64

0:02:07,686 --> 0:02:09.686  
minority and I think that's worth

65

0:02:09,686 --> 0:02:11.686  
pursuing.

66

0:02:11,686 --> 0:02:13.686  
So, horizontal gene transfer

67

0:02:13,686 --> 0:02:15.686  
is way more common

68

0:02:15,686 --> 0:02:17.686

to prokaryotic organisms.

69

0:02:17,686 --> 0:02:19.686  
But in terms of eukaryotes,

70

0:02:19,686 --> 0:02:21.686  
there is limited evidence

71

0:02:21,859 --> 0:02:23.859  
of horizontal gene transfer

72

0:02:23,859 --> 0:02:25.859  
in species of the Candida genus,

73

0:02:25,859 --> 0:02:27.859  
a type of pathogenic yeast.

74

0:02:27,859 --> 0:02:29.859  
And those instances

75

0:02:29,859 --> 0:02:31.859  
were between Candida yeast and bacteria

76

0:02:31,859 --> 0:02:33.859  
and between a

77

0:02:33,859 --> 0:02:35.859  
Candida yeast species and another yeast species.

78

0:02:35,859 --> 0:02:37.859  
So, I think that warrants

79

0:02:37,859 --> 0:02:39.859  
additional research on the evolutionary

80

0:02:39,859 --> 0:02:41.859  
forced behind

81

0:02:41,859 --> 0:02:43.859  
fungal pathogenesis.

82

0:02:43,859 --> 0:02:45.859

And quickly,

83

0:02:45,859 --> 0:02:47.859

Candida became

84

0:02:47,859 --> 0:02:49.859

the culprit of interest, because

85

0:02:50,671 --> 0:02:52.671

they are an opportunistic pathogen.

86

0:02:52,671 --> 0:02:54.671

So they do inhabit

87

0:02:54,671 --> 0:02:56.671

certain microbiomes, but

88

0:02:56,671 --> 0:02:58.671

it seems from when I was doing research that

89

0:02:58,671 --> 0:03:00.671

when some kind of disruption occurs

90

0:03:00,671 --> 0:03:02.671

like an invasive surgery,

91

0:03:02,714 --> 0:03:04.714

that can open the door to

92

0:03:04,714 --> 0:03:06.714

infection. So,

93

0:03:06,714 --> 0:03:08.714

the essential questions in this

94

0:03:08,714 --> 0:03:10.714

project were what

95

0:03:10,714 --> 0:03:12.714

bioinformatics tools can be used to

96

0:03:12,714 --> 0:03:14.714  
not only predict,

97

0:03:14,714 --> 0:03:16.714  
potential

98

0:03:16,714 --> 0:03:18.714  
or other potential

99

0:03:18,714 --> 0:03:20.714  
fungal pathogens, but

100

0:03:20,714 --> 0:03:22.714  
also more importantly

101

0:03:22,714 --> 0:03:24.714  
what is the

102

0:03:24,714 --> 0:03:26.714  
role of horizontal gene transfer

103

0:03:26,714 --> 0:03:28.714  
or fungal horizontal gene transfer

104

0:03:28,714 --> 0:03:30.714  
in the context

105

0:03:30,714 --> 0:03:32.714  
antifungal resistance

106

0:03:32,714 --> 0:03:34.714  
gene acquisition.

107

0:03:34,714 --> 0:03:36.714  
And in terms of the methodology,

108

0:03:36,714 --> 0:03:38.714  
I started by



109  
0:03:38,714 --> 0:03:40.714  
using BLAST

110  
0:03:40,714 --> 0:03:42.714  
and of course BLAST is a software,

111  
0:03:42,714 --> 0:03:44.714  
there are a number of platforms that

112  
0:03:44,714 --> 0:03:46.714  
support it,

113  
0:03:46,714 --> 0:03:48.714  
but I prefer just because the interface

114  
0:03:48,714 --> 0:03:50.714  
I think is easier to work with,

115  
0:03:50,714 --> 0:03:52.714  
the National Center for

116  
0:03:52,714 --> 0:03:54.714  
Biotechnology Information or NCBI,

117  
0:03:54,897 --> 0:03:56.897  
their BLAST program.

118  
0:03:56,897 --> 0:03:58.897  
So, I used

119  
0:03:58,897 --> 0:04:00.897  
previously-sequenced antifungal resistance

120  
0:04:00,897 --> 0:04:02.897  
genes and they are each,

121  
0:04:02,897 --> 0:04:04.897  
in Bioinformatics projects,

122  
0:04:04,897 --> 0:04:06.897

sequence data, it has

123

0:04:06,897 --> 0:04:08.897  
an accession number attached to it

124

0:04:08,897 --> 0:04:10.897  
and that's just like an ID

125

0:04:10,897 --> 0:04:12.897  
or tag if you will,

126

0:04:12,897 --> 0:04:14.897  
and the antifungal resistance

127

0:04:14,897 --> 0:04:16.897  
genes that I searched

128

0:04:16,897 --> 0:04:18.897  
in BLAST and starting

129

0:04:18,897 --> 0:04:20.897  
with a number of

130

0:04:20,897 --> 0:04:24.897  
Candida species were: ERG11,

131

0:04:24,897 --> 0:04:26.897  
ERG3, TAC1,

132

0:04:26,902 --> 0:04:28.902  
GSC1,

133

0:04:28,902 --> 0:04:30.902  
cgPDR1, FSK1

134

0:04:30,902 --> 0:04:32.902  
and FSK or

135

0:04:32,902 --> 0:04:34.902  
sorry, FKS1 or FKS2

136  
0:04:35,026 --> 0:04:37.026  
um and

137  
0:04:37,026 --> 0:04:39.026  
in doing the

138  
0:04:39,026 --> 0:04:41.026  
BLAST search,

139  
0:04:41,026 --> 0:04:43.026  
so I had to start by doing a Forward BLAST

140  
0:04:43,292 --> 0:04:45.292  
starting with the

141  
0:04:45,292 --> 0:04:47.292  
fungal sequence and seeing

142  
0:04:47,292 --> 0:04:49.292  
the viral hits that appear

143  
0:04:49,292 --> 0:04:51.292  
as well as a Reverse BLAST.

144  
0:04:51,292 --> 0:04:53.292  
So going from those viral hits, back

145  
0:04:53,292 --> 0:04:55.292  
to the

146  
0:04:55,292 --> 0:04:57.292  
fungal hits and

147  
0:04:57,292 --> 0:04:59.292  
potentially getting the same hit.

148  
0:04:59,292 --> 0:05:01.292  
And then

149  
0:05:01,292 --> 0:05:03.292

in a bioinformatics project,

150

0:05:03,292 --> 0:05:05.292

this is one thing I've found and I think this

151

0:05:05,292 --> 0:05:07.292

holds true, is

152

0:05:07,507 --> 0:05:09.507

there is a lot of data that comes at

153

0:05:09,507 --> 0:05:11.507

you and it's almost like a fire hose,

154

0:05:11,507 --> 0:05:13.507

so it's really important to establish

155

0:05:13,507 --> 0:05:15.507

cutoffs and know

156

0:05:15,507 --> 0:05:17.507

what data is actually

157

0:05:17,507 --> 0:05:19.507

worth exploring

158

0:05:19,507 --> 0:05:21.507

further

159

0:05:21,507 --> 0:05:23.507

or indicates something.

160

0:05:23,507 --> 0:05:25.507

So in this project,

161

0:05:25,507 --> 0:05:27.507

the BLAST cutoffs

162

0:05:27,507 --> 0:05:29.507

we used were

163  
0:05:29,507 --> 0:05:31.507  
greater than 40% query coverage,

164  
0:05:31,507 --> 0:05:33.507  
expect value of

165  
0:05:33,551 --> 0:05:35.551  
less than 10 to the negative 5th,

166  
0:05:35,551 --> 0:05:37.551  
and originally we

167  
0:05:37,551 --> 0:05:39.551  
wanted to see greater than 80% identity

168  
0:05:39,551 --> 0:05:41.551  
of course,

169  
0:05:41,551 --> 0:05:43.551  
we were a little flexible on this because

170  
0:05:43,551 --> 0:05:45.551  
for eukaryotic

171  
0:05:45,551 --> 0:05:47.551  
horizontal gene transfer, you're comparing

172  
0:05:47,551 --> 0:05:49.551  
fungi and viruses,

173  
0:05:49,551 --> 0:05:51.551  
so because they're so different

174  
0:05:51,551 --> 0:05:53.551  
we placed less emphasis on percent

175  
0:05:53,551 --> 0:05:55.551  
identity in this case.

176  
0:05:55,551 --> 0:05:57.551

So, um

177

0:05:57,551 --> 0:05:59.551  
through the first month

178

0:05:59,551 --> 0:06:01.551  
it quickly became evident

179

0:06:01,551 --> 0:06:03.551  
the strongest BLAST

180

0:06:03,551 --> 0:06:05.551  
evidence for fungal

181

0:06:05,551 --> 0:06:07.551  
horizontal gene transfer

182

0:06:07,551 --> 0:06:09.551  
or the strongest antifungal resistance

183

0:06:09,551 --> 0:06:11.551  
gene was

184

0:06:11,551 --> 0:06:13.551  
ERG3,

185

0:06:13,551 --> 0:06:15.551  
which the gene product is

186

0:06:15,551 --> 0:06:17.551  
C-5 sterol desaturase,

187

0:06:17,551 --> 0:06:19.551  
desaturase and that's going to

188

0:06:19,551 --> 0:06:21.551  
play an important role in

189

0:06:21,551 --> 0:06:23.551  
ergosterol synthesis

190  
0:06:23,551 --> 0:06:25.551  
or cell membrane stability.

191  
0:06:25,551 --> 0:06:27.551  
And

192  
0:06:27,551 --> 0:06:29.551  
going forward with this

193  
0:06:29,551 --> 0:06:31.551  
in the next month, we

194  
0:06:31,551 --> 0:06:33.551

195  
0:06:33,551 --> 0:06:35.551  
used BLAST

196  
0:06:35,729 --> 0:06:37.729  
going from pathogenic

197  
0:06:37,729 --> 0:06:39.729  
fungi, looking at the PEP8

198  
0:06:39,729 --> 0:06:41.729  
gene, because that's

199  
0:06:41,729 --> 0:06:43.729  
a confirmed gene in fungal pathogenesis

200  
0:06:43,729 --> 0:06:45.729  
and looking at other fungi

201  
0:06:45,729 --> 0:06:47.729  
and that did not come up

202  
0:06:47,729 --> 0:06:49.729  
with a lot, at least not

203  
0:06:49,729 --> 0:06:51.729

any fungi that are

204

0:06:51,943 --> 0:06:53.943  
known to be pathogenic to humans.

205

0:06:53,943 --> 0:06:55.943  
But, just to be thorough

206

0:06:55,943 --> 0:06:57.943  
I also

207

0:06:57,943 --> 0:06:59.943  
did the same Forward and Reverse process

208

0:06:59,943 --> 0:07:01.943  
in those fungal hits,

209

0:07:01,943 --> 0:07:03.943  
um and

210

0:07:03,943 --> 0:07:05.943  
with a special interest

211

0:07:05,943 --> 0:07:07.943  
in ERG3.

212

0:07:07,943 --> 0:07:09.943  
And of course that, in terms of

213

0:07:09,943 --> 0:07:11.943  
the BLAST evidence didn't reveal

214

0:07:11,943 --> 0:07:13.943  
a whole lot.

215

0:07:13,943 --> 0:07:15.943  
So, BLAST because it's a quick

216

0:07:15,943 --> 0:07:17.943  
statistical tool,



217  
0:07:17,943 --> 0:07:19.943

218  
0:07:19,943 --> 0:07:21.943  
other research articles often

219  
0:07:21,943 --> 0:07:23.943  
say to

220  
0:07:23,943 --> 0:07:25.943  
have more support for horizontal

221  
0:07:25,943 --> 0:07:27.943  
gene transfer, you really

222  
0:07:27,943 --> 0:07:29.943  
need to analyze the data further.

223  
0:07:29,943 --> 0:07:31.943  
And in doing this, I

224  
0:07:31,943 --> 0:07:33.943  
used, first I constructed an

225  
0:07:33,943 --> 0:07:35.943  
alignment because

226  
0:07:35,943 --> 0:07:37.943  
that's just a quick visual way to

227  
0:07:37,943 --> 0:07:39.943  
look at evolutionary

228  
0:07:39,943 --> 0:07:41.943  
conservation in sequence data

229  
0:07:42,378 --> 0:07:44.378  
and I used that with,

230  
0:07:44,378 --> 0:07:46.378

using a program

231

0:07:46,378 --> 0:07:48.378  
by the European Molecular

232

0:07:48,378 --> 0:07:50.378  
Biological Laboratory (EMBL)

233

0:07:50,378 --> 0:07:52.378  
a program called MUSCLE,

234

0:07:52,378 --> 0:07:54.378  
or Multiple Sequence Comparison by

235

0:07:54,378 --> 0:07:56.378  
Log Expectation.

236

0:07:56,378 --> 0:07:58.378  
And that is featured on my

237

0:07:58,378 --> 0:08:00.378  
results next to

238

0:08:00,484 --> 0:08:02.484  
the phylogenetic tree,

239

0:08:03,037 --> 0:08:05.037  
which I will talk about right now.

240

0:08:05,037 --> 0:08:07.037  
For that, I used MEGA

241

0:08:07,037 --> 0:08:09.037  
or Molecular Evolutionary

242

0:08:09,037 --> 0:08:11.037  
Genetic Analysis,

243

0:08:11,037 --> 0:08:13.037  
a pretty-commonly used software

244  
0:08:13,037 --> 0:08:15.037  
and in doing this,

245  
0:08:15,037 --> 0:08:17.037  
I constructed a Maximum

246  
0:08:17,037 --> 0:08:19.037  
Likelihood phylogeny

247  
0:08:19,037 --> 0:08:21.037  
of the antifungal

248  
0:08:21,037 --> 0:08:23.037  
resistance gene of interest

249  
0:08:23,037 --> 0:08:25.037  
and in bold are

250  
0:08:25,037 --> 0:08:27.037  
the hits from

251  
0:08:27,037 --> 0:08:29.037  
the BLAST evidence

252  
0:08:29,037 --> 0:08:31.037  
and what this can do is tell us

253  
0:08:31,037 --> 0:08:33.037  
the likely

254  
0:08:33,037 --> 0:08:35.037  
evolutionary relationship

255  
0:08:35,037 --> 0:08:37.037  
of the sequence, based on the alignment.

256  
0:08:37,037 --> 0:08:39.037  
So

257  
0:08:39,037 --> 0:08:41.037

in terms of the BLAST evidence that was

258

0:08:41,037 --> 0:08:43.037  
strongest, ERG3

259

0:08:43,037 --> 0:08:45.037  
in *Candida albicans*

260

0:08:45,037 --> 0:08:47.037  
and *Candida californica*

261

0:08:47,037 --> 0:08:49.037  
came up as a hit

262

0:08:49,037 --> 0:08:51.037  
for the forward and reverse

263

0:08:51,037 --> 0:08:53.037  
with Orpheovirus

264

0:08:53,037 --> 0:08:55.037  
and *Candida glabrata*

265

0:08:55,037 --> 0:08:57.037  
came up with a hit,

266

0:08:57,037 --> 0:08:59.037  
came up as a hit for

267

0:08:59,037 --> 0:09:01.037  
Edafosvirus.

268

0:09:01,037 --> 0:09:03.037  
These viruses are actually

269

0:09:03,037 --> 0:09:05.037  
a type of "giant virus"

270

0:09:05,037 --> 0:09:07.037  
and known to infect algae,

271

0:09:07,037 --> 0:09:09.037  
so that was kind of surprising.

272

0:09:09,037 --> 0:09:11.037  
I kind of figured a

273

0:09:11,037 --> 0:09:13.037  
mycophage or even a bacteriophage

274

0:09:13,037 --> 0:09:15.037  
would come up as a

275

0:09:15,037 --> 0:09:17.037

276

0:09:17,037 --> 0:09:19.037  
as a viral hit in the BLAST searches,

277

0:09:19,037 --> 0:09:21.037  
but in pursuing this

278

0:09:21,037 --> 0:09:23.037  
I also thought it was worth and also

279

0:09:23,037 --> 0:09:25.037  
for the purpose of the phylogenetic tree,

280

0:09:25,037 --> 0:09:27.037  
because it's important to have an outgroup,

281

0:09:27,037 --> 0:09:29.037  
I also included

282

0:09:29,037 --> 0:09:31.037  
the algal host

283

0:09:31,037 --> 0:09:33.037  
of these viruses

284

0:09:33,037 --> 0:09:35.037

which is Acanthamoeba

285

0:09:35,037 --> 0:09:37.037  
castellanii, and

286

0:09:37,037 --> 0:09:39.037  
that is at the very bottom of the tree.

287

0:09:39,037 --> 0:09:41.037  
So in producing the tree,

288

0:09:41,037 --> 0:09:43.037  
it quickly became

289

0:09:43,037 --> 0:09:45.037  
evident that

290

0:09:45,037 --> 0:09:47.037  
the maximum likelihood

291

0:09:47,037 --> 0:09:49.037  
phylogeny results

292

0:09:49,037 --> 0:09:51.037  
contradicted the

293

0:09:51,037 --> 0:09:53.037  
BLAST results

294

0:09:53,037 --> 0:09:55.037  
in that the

295

0:09:55,120 --> 0:09:57.120  
ERG3 gene in

296

0:09:57,120 --> 0:09:59.120  
the Candida species

297

0:09:59,120 --> 0:10:01.120  
were pretty much grouped with

298  
0:10:01,120 --> 0:10:03.120  
yeast, not in

299  
0:10:03,120 --> 0:10:05.120  
a smaller

300  
0:10:05,120 --> 0:10:07.120  
ingroup

301  
0:10:07,120 --> 0:10:09.120  
or closely related

302  
0:10:09,120 --> 0:10:11.120  
with the viral

303  
0:10:11,329 --> 0:10:13.329  
ERG3 orthologs,

304  
0:10:13,329 --> 0:10:15.329  
And

305  
0:10:15,329 --> 0:10:17.329  
what's kind of weird though

306  
0:10:17,489 --> 0:10:19.489  
was another fungal species

307  
0:10:19,489 --> 0:10:21.489  
that was a PEP8 ortholog,

308  
0:10:21,489 --> 0:10:23.489  
so of that

309  
0:10:23,489 --> 0:10:25.489  
pathogenic gene that

310  
0:10:25,489 --> 0:10:27.489  
encodes a sorting protein

311  
0:10:27,489 --> 0:10:29.489

that

312

0:10:29,489 --> 0:10:31.489  
Hypopichia

313

0:10:31,489 --> 0:10:33.489  
burtonii,

314

0:10:33,540 --> 0:10:35.540  
this kind of unsuspected

315

0:10:35,540 --> 0:10:37.540  
fungal species,

316

0:10:37,540 --> 0:10:39.540  
ERG3

317

0:10:39,540 --> 0:10:41.540  
in that species was grouped

318

0:10:41,540 --> 0:10:43.540  
closer with the virus or with both

319

0:10:43,540 --> 0:10:45.540  
viruses than it was

320

0:10:45,540 --> 0:10:47.540  
with either

321

0:10:47,540 --> 0:10:49.540  
the pathogenic

322

0:10:49,540 --> 0:10:51.540  
Candida species

323

0:10:51,540 --> 0:10:53.540  
or with the algal host

324

0:10:53,540 --> 0:10:55.540  
species of both viruses.



325  
0:10:55,540 --> 0:10:57.540  
So, I tried to pursue that

326  
0:10:57,540 --> 0:10:59.540  
and in BLAST,

327  
0:10:59,595 --> 0:11:01.595  
that didn't come up with anything.

328  
0:11:01,595 --> 0:11:03.595  
But a reason for that could be

329  
0:11:03,595 --> 0:11:05.595  
due to the

330  
0:11:05,595 --> 0:11:07.595  
length of the ERG3 sequence,

331  
0:11:07,595 --> 0:11:09.595  
of course this is the protein

332  
0:11:09,595 --> 0:11:11.595  
or the gene product.

333  
0:11:11,595 --> 0:11:13.595  
The protein sequence, so that's around

334  
0:11:13,595 --> 0:11:15.595  
370 amino acids,

335  
0:11:15,947 --> 0:11:17.947  
so not terribly, not a terribly long

336  
0:11:17,947 --> 0:11:19.947  
sequence

337  
0:11:19,947 --> 0:11:21.947  
and that could explain those results.

338  
0:11:21,947 --> 0:11:23.947

I will move myself over

339

0:11:23,947 --> 0:11:25.947  
to the other corner

340

0:11:25,947 --> 0:11:27.947  
and conclude

341

0:11:27,947 --> 0:11:29.947  
this presentation.

342

0:11:29,947 --> 0:11:31.947  
So in conclusion,

343

0:11:31,947 --> 0:11:33.947  
BLAST showed limited evidence of

344

0:11:33,947 --> 0:11:35.947  
horizontal gene transfer

345

0:11:35,947 --> 0:11:37.947  
of ERG3 between

346

0:11:37,947 --> 0:11:39.947  
three *Candida* species, which were:

347

0:11:39,947 --> 0:11:41.947  
*Candida albicans*, *californica*, and

348

0:11:41,947 --> 0:11:43.947  
*glabrata* with the

349

0:11:43,947 --> 0:11:45.947  
two viruses of interest:

350

0:11:45,947 --> 0:11:47.947  
*orpheovirus* and *edafosvirus*.

351

0:11:47,947 --> 0:11:49.947

352  
0:11:49,947 --> 0:11:51.947  
The maximum likelihood phylogeny

353  
0:11:51,947 --> 0:11:53.947  
did not support these

354  
0:11:53,947 --> 0:11:55.947  
results and of course, that's based on

355  
0:11:55,947 --> 0:11:57.947  
the alignment and I did not include

356  
0:11:57,947 --> 0:11:59.947  
the full alignment on this poster.

357  
0:11:59,947 --> 0:12:01.947  
I only included the

358  
0:12:01,947 --> 0:12:03.947  
enzyme active site

359  
0:12:03,947 --> 0:12:05.947  
alignment for ERG3P,

360  
0:12:05,947 --> 0:12:07.947  
which is located

361  
0:12:07,995 --> 0:12:09.995  
inside of a histidine box according to

362  
0:12:09,995 --> 0:12:11.995  
one article and

363  
0:12:11,995 --> 0:12:13.995  
of course one species there,

364  
0:12:13,995 --> 0:12:15.995  
the one fungal species,

365  
0:12:15,995 --> 0:12:17.995

there was no observed histidine box

366

0:12:17,995 --> 0:12:19.995

but

367

0:12:19,995 --> 0:12:21.995

that is what the alignment came

368

0:12:21,995 --> 0:12:23.995

up with.

369

0:12:23,995 --> 0:12:25.995

But it didn't support the BLAST

370

0:12:25,995 --> 0:12:27.995

evidence, which is really strange.

371

0:12:27,995 --> 0:12:29.995

So considering that

372

0:12:29,995 --> 0:12:31.995

it's likely that BLAST

373

0:12:31,995 --> 0:12:33.995

for the context of

374

0:12:33,995 --> 0:12:35.995

horizontal gene transfer is probably

375

0:12:35,995 --> 0:12:37.995

better suited for prokaryotes,

376

0:12:37,995 --> 0:12:39.995

not eukaryotes like this project.

377

0:12:39,995 --> 0:12:41.995

Um

378

0:12:41,995 --> 0:12:43.995

and it's important to remember that it's a quick

379  
0:12:43,995 --> 0:12:45.995  
statistical tool, so it's not really meant

380  
0:12:46,257 --> 0:12:48.257  
to be a definitive predictor

381  
0:12:48,484 --> 0:12:50.484  
of these

382  
0:12:50,484 --> 0:12:52.484  
evolutionary events.

383  
0:12:52,484 --> 0:12:54.484  
And another thing to

384  
0:12:54,484 --> 0:12:56.484  
or an important thing to consider is that

385  
0:12:56,484 --> 0:12:58.484  
because less than 30% of

386  
0:12:58,484 --> 0:13:00.484  
*Candida albicans*

387  
0:13:00,484 --> 0:13:02.484  
so a really more-common

388  
0:13:02,484 --> 0:13:04.484  
fungal pathogen,

389  
0:13:04,484 --> 0:13:06.484  
considering that less than 30% of its genes

390  
0:13:06,484 --> 0:13:08.484  
are characterized,

391  
0:13:08,484 --> 0:13:10.484  
and that value is going to be

392  
0:13:10,484 --> 0:13:12.484

a lot smaller

393

0:13:12,484 --> 0:13:14.484  
for other Candida species

394

0:13:14,484 --> 0:13:16.484  
or less known

395

0:13:16,484 --> 0:13:18.484  
fungi,

396

0:13:18,484 --> 0:13:20.484  
really being able to predict

397

0:13:20,484 --> 0:13:22.484  
more gene products and

398

0:13:22,484 --> 0:13:24.484  
characterize

399

0:13:24,484 --> 0:13:26.484  
the

400

0:13:26,484 --> 0:13:28.484  
full genome

401

0:13:28,484 --> 0:13:30.484  
of other fungal species,

402

0:13:30,484 --> 0:13:32.484  
that could reveal new evidence of horizontal

403

0:13:32,484 --> 0:13:34.484  
gene transfer.

404

0:13:34,484 --> 0:13:36.484  
So,

405

0:13:36,484 --> 0:13:38.484  
while we didn't find

406  
0:13:38,484 --> 0:13:40.484  
a definitive um

407  
0:13:40,484 --> 0:13:42.484  
instance of horizontal gene transfer,

408  
0:13:42,484 --> 0:13:44.484  
we found

409  
0:13:44,484 --> 0:13:46.484  
BLAST

410  
0:13:46,484 --> 0:13:48.484  
evidence for ERG3

411  
0:13:48,484 --> 0:13:50.484  
horizontal gene transfer that seems to

412  
0:13:50,484 --> 0:13:52.484  
contradict

413  
0:13:52,484 --> 0:13:54.484  
um MEGA

414  
0:13:54,484 --> 0:13:56.484

415  
0:13:56,484 --> 0:13:58.484  
the maximum likelihood phylogeny's

416  
0:13:58,484 --> 0:14:00.484  
prediction of

417  
0:14:00,484 --> 0:14:02.484  
ERG3 horizontal gene transfer.

418  
0:14:02,484 --> 0:14:04.484  
So,

419  
0:14:04,484 --> 0:14:06.484

I would also like to thank the

420

0:14:06,484 --> 0:14:08.484  
biology department

421

0:14:08,484 --> 0:14:10.484  
and the honors program at the university

422

0:14:10,484 --> 0:14:12.484  
of mary washington for me giving me

423

0:14:12,484 --> 0:14:14.484  
the opportunity to conduct this project.

424

0:14:14,484 --> 0:14:16.484  
It was a really amazing opportunity.

425

0:14:16,484 --> 0:14:18.484  
And I also want to thank

426

0:14:18,484 --> 0:14:20.484  
Dr. Vinayak Mathur

427

0:14:20,484 --> 0:14:22.484  
at Cabrini University, so he's

428

0:14:22,484 --> 0:14:24.484  
part of the Genome Solver team,

429

0:14:24,484 --> 0:14:26.484  
and his expert guidance was super

430

0:14:26,484 --> 0:14:28.484  
helpful in

431

0:14:28,484 --> 0:14:30.484  
completing this project and

432

0:14:30,484 --> 0:14:32.484  
making sense of this data.



433

0:14:32,484 --> 0:14:34.484

And that is all

434

0:14:34,484 --> 0:14:36.484

I have, but thank you so much

435

0:14:36,484 --> 0:14:40.484

for listening and let me know if you have any questions I guess in the comments.

436

0:14:40,484 --> 0:14:42.484

Take care.