Tarefa avaliada por colega: Bioinformatics Application Challenge

Enviado!
Seu trabalho está pronto para ser avaliado por seus colegas.
Depois, você também precisará avaliar o trabalho deles.
Enviaremos um e-mail quando sua nota estiver pronta. Ela deverá
estar pronta em **Dezembro 14, 11:59 PM PST**.

> Avaliar o trabalho de um colega

Instruções

My submission

Discussões

A comparative study of the search of motifs in datasets of different sizes

Enviado em 5 de Dezembro de 2016

Shareable Link

We will begin by running three different software tools on the dataset provided. First, upload upstream250.txt to Consensus (Hertz and Stormo, 1999). Set the desired pattern width equal to 20 (keep all other parameters the same) and click "submit".After the program has run, scroll to the bottom of the page and click "next". Under "Matrix 1", you will see 19 sequences corresponding to the substrings of the input strings having length 20 that are generated as a motif matrix. The elements in the column to the left of these sequences have the form XXX/YYY, where YYY represents the starting position of each sequence in the original string of length 250.

**Provide all of the starting positions of the strings of length 20.**

187
141
117
162
156
199
139
202
175
161
180
169
174

Central de Ajuda

120

161

198

214

157

159

In order to visualize the information contained in these sequences, we will copy them into WebLogo (http://weblogo.threeplusone.com) to generate a motif logo.

Upload the image file obtained after generating this motif logo (with default parameters).
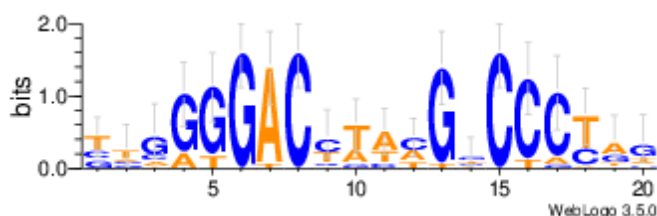
We will perform similar tasks with MEME (Bailey and Elkan, 1994). Upload upstream250.txt, and tell MEME to find 1 motif instead of 3. Then click on advanced options and change the minimum width to 20 and the maximum width to 20.After submitting the process, click on "MEME html output". Notice that the motif logo has been generated under "Discovered Motifs". Click the down arrow under "more" to see the starting positions of motifs. To download the motif logo, click the right arrow above the logo, navigate to the "Download-logo" tab, and click "Download".

If the queue on the MEME server is too long, you can use alternate instance.

Indicate the starting positions of the substrings of length 20 identified by MEME.

57

139

172

114

107

136

143

155

159

137

200

118

186

178

173

160

62

201

216

165

45

154

204
187

Upload the image file obtained after downloading this motif logo.

We will now run a third motif finder, W-AlignACE (Chen and Jiang, 2006). Choose upstream250.txt as your file to upload, enter 20 as the number of columns to align and the number of sites to expect, and change the fractional background GC-content to 0.66 (the GC-content of *Mycobacterium tuberculosis*). Then click "submit".

We will not ask you to output the results of your program. Why do you think that this is? (hint: W-AlignACE is based on Gibbs sampling). Make sure to answer in your own words.

Because Gibbs sampling uses random select of motifs, the software may produce different outputs in different runs. This occurs even when running it with the same random seed but in different computers.

Did all three programs generate similar motifs? Provide a brief (1-2 sentence) explanation.

No, they are all different. Besides all algorithms have the same settings, they should have different heuristic behavior, since they aren't brute force approach.

Although your biologist colleague told you that the motif is probably about 20 bp long, you are skeptical, so you decide to run a motif finding program that finds a motif over a wide range of different lengths.

Run MEME again on upstream250.txt, but this time, use the default parameters for minimum width (6) and maximum width (50). Note: this process may take a few minutes to run.

(a) How long is the motif produced by MEME?

(b) Is the motif logo produced by MEME similar to the one produced before for a motif of length 20?

a) As I left the default settings, it returned me three motifs. One with size 41, one with size 40 and one with size 20.
b) The size 40 motif is similar of the produced before 20 size motif.

When using motif software with fixed motif lengths, is it better to start with short motifs or long motifs? Why?

I would start with long motifs because if you compare the results in motifs of various sizes you can visualize a consensus between them.

To evaluate the statistical significance of an identified motif, we need to ensure that a motif with the same or even larger score is unlikely to occur in a collection of "typical" DNA strings (of the same length).

How would you generate these strings? Justify your answer.

We have begun to confirm our colleague's suspicion that we should consider motifs of length about 20. However, thus far, we have only analyzed the 250 bp regions upstream of each gene. This makes us wonder whether we will identify the same motif for upstream regions of different lengths. First, we will consider upstream regions of length 25 bp (upstream25.txt).

Upload the motif logo obtained by running MEME on upstream25.txt. (Remember to specify a motif of length 20 in the advanced options.)
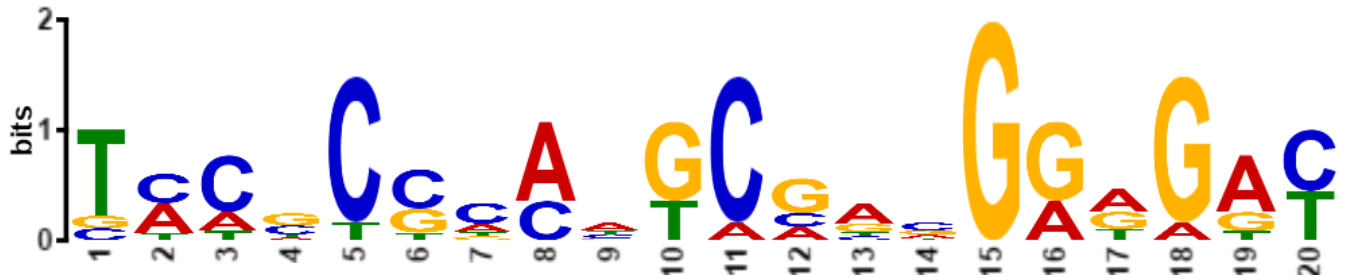
Next, we will consider upstream regions of length 100 bp (upstream100.txt).

Upload the motif logo obtained by running MEME on upstream100.txt. (Remember to specify a motif of length 20 in the advanced options.)

Next, we will consider upstream regions of length 500 bp (upstream500.txt).

Upload the motif logo obtained by running MEME on upstream500.txt. (Remember to specify a motif of length 20 in the advanced options.)

Finally, we will consider upstream regions of length 1000 bp (upstream1000.txt).

Upload the motif logo obtained by running MEME on upstream1000.txt. (Remember to specify a motif of length 20 in the advanced options.)

Central de Ajuda

We will now compare the different motif logos generated from varying the length of upstream regions.

Which of the motif logos that you created are similar to the motif logo generated from upstream250.txt?

Both logos generated with upstream500 and upstream1000 are similar to motif logo generated with upstream250

✎ Editar envio

Comentários

Visível para os colegas de turma

compartilhe suas ideias...

👍   👎   🚩

Central de Ajuda