

Hello there,

General culture about bioinformatic is pretty low. For most it is a dark arcana done by weird people able to understand cryptic language as python or R. Here I'll try to make it as much understandable for the common soul as possible.

My main goal will be to show you the steps to find promoters motifs and RBS (ribosome binding site) form a given genome.

First where to find the genome? Gladly you will not need to search the mega structure fighting safeguard to find a genetic terminal. No, you just need to go on NCBI and search the annotated genome of the organisme you want to study. search the genome section of NCBI

<https://www.ncbi.nlm.nih.gov/genome/>

And download the chosen genome in a genbank file (full).

Now what do you need?

First Python, you can download a python distribution like anaconda or winpython.

Secondly you need biopython, for this follow the instructions, it can be a bit hard

<https://biopython.org/wiki/Download>

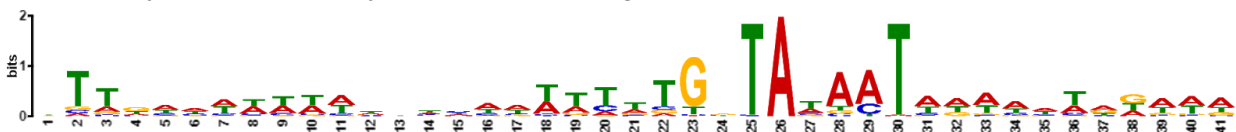
Then you can use my script named Killy.py, to gave you the sequences of promoters, just gave it your .gb or .gbk (genbank) file as argument (with only letters and numbers in the name) and it will write you a fasta file with the set of promoter sequences named as the input file +prom.fasta. It also give you the number of selected sequences.

What do you do next?

You feed it to MEME : <http://meme-suite.org/tools/meme>

With the good parameters, ie Zoop mode, and a minimum of 35 sequence (around 11 percent of the sequences given) to build a motif.

And tadam, you have motifs, you shall found things like promoters:

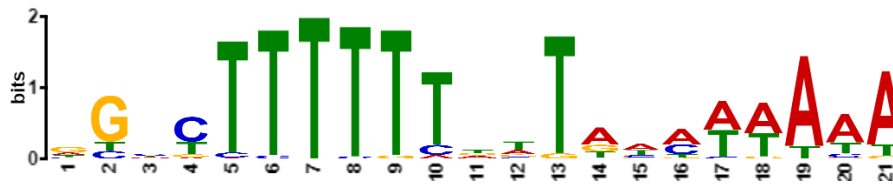


Recognizable by their -10 homology motif. But also RBS:



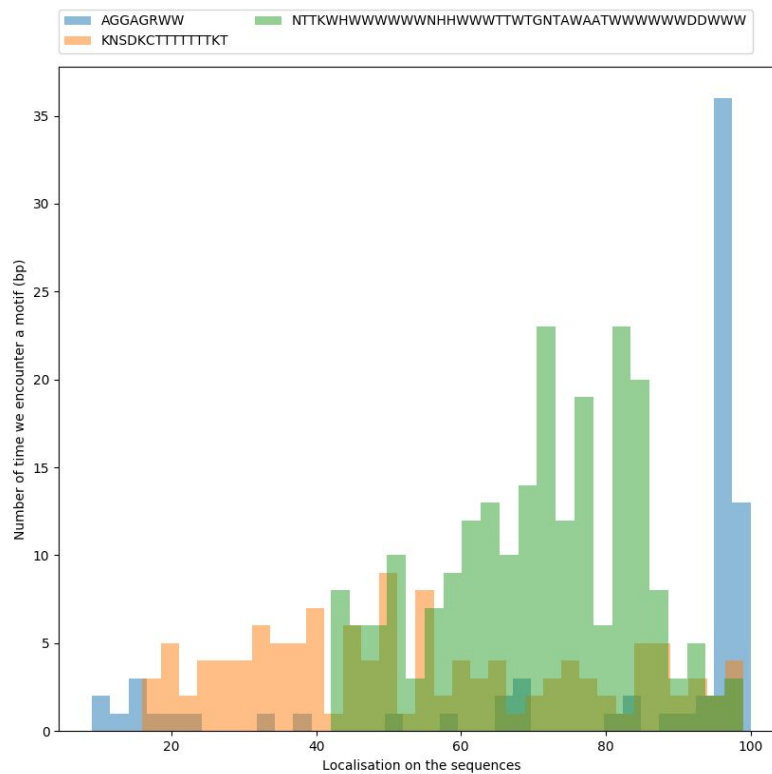
That are identifiable as the Shine Dalgarno sequence.

But also aberrations, as tails of terminators:



With the known motif of repetition of T.

I also did a script to have the histogram of the localisation of the patterns Cibo.py just download the XML file of mast result on the same folder as Cibo.py, name it mast.xml, and the script will run and gave you the histogram.



Both script are well commented for you to understand the method I use to build them.