iRecSpot-EF: Effective Sequence Based Features for Recombination Hotspot Prediction

Md Rafsan Jani, Md Toha Khan Mozlish,
Sajid Ahmed, Niger Sultana Tahniat, Dewan Md Farid, and
Swakkhar Shatadba

Department of Computer Science and Engineering, United International University, Dhaka, Bangladesh

1 Supplementary Information (Sp3): How does our web service work?

1.1 Step #1:

Access the service by clicking the link at `http://irecspot.pythonanywhere.com/`. We will see the home-page.

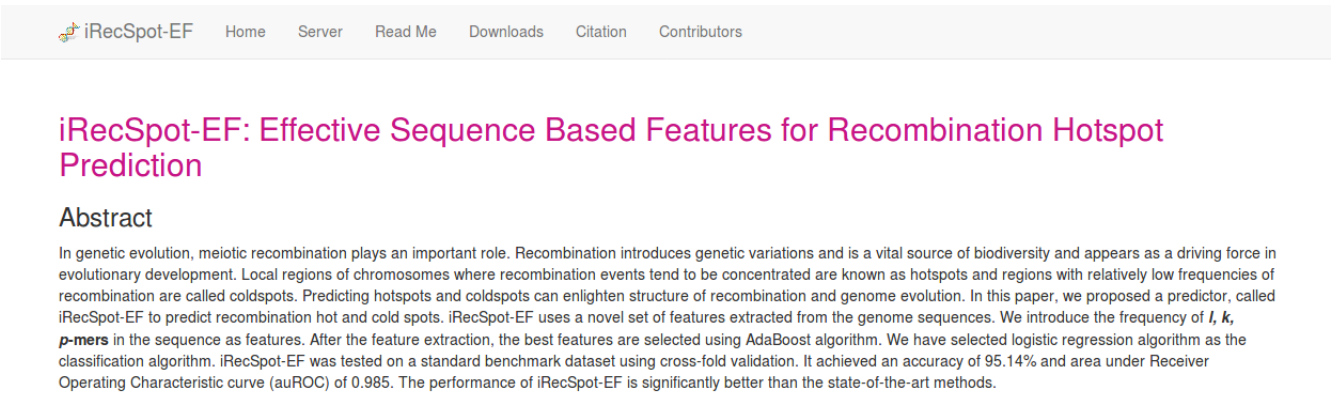

Figure 1: Home Page

1.2 Step #2:

Click on the 'Server' button on the top menu or go to this link `http://irecspot.pythonanywhere.com/server`. We will see the server page.

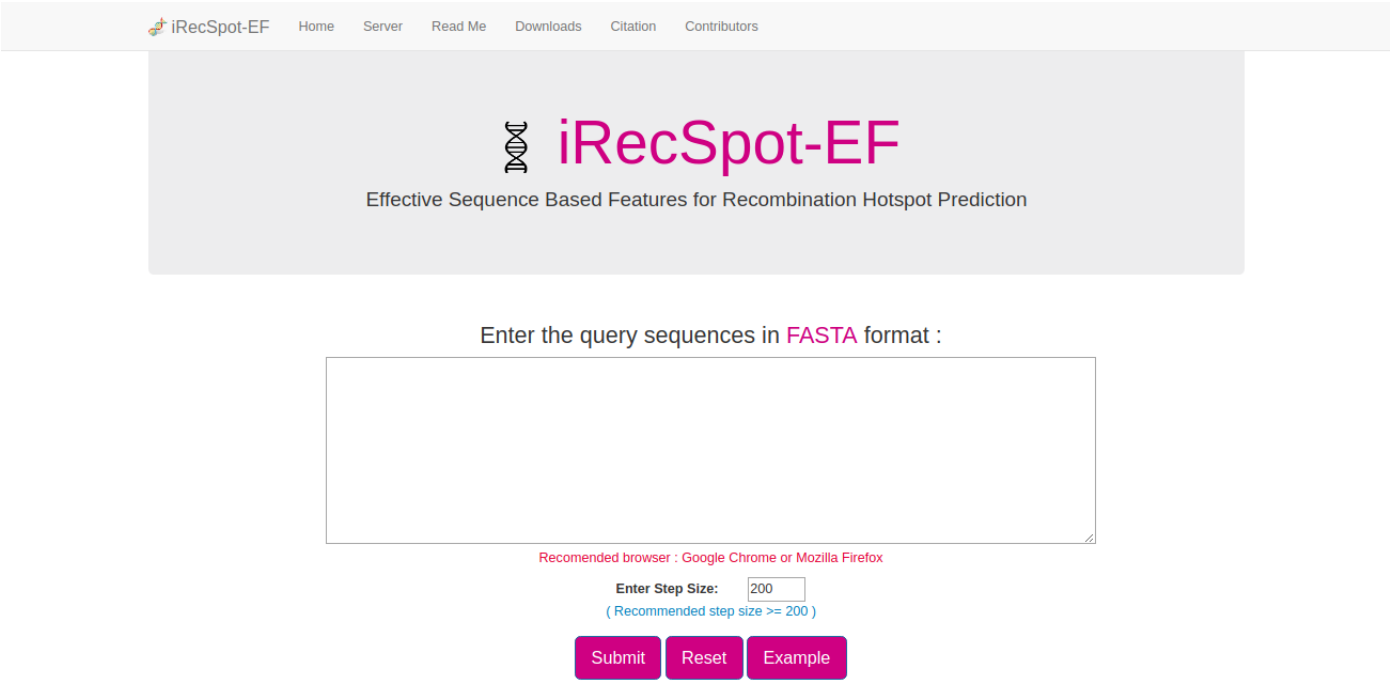

Figure 2: Server Page

1.3 Step #3:

Once in the server page, either type or copy and paste the query DNA sequence into the input box. The input sequence should be in the FASTA format. A sequence in FASTA format consists of a single initial line beginning with the symbol '>' in the first column, followed by lines of sequence data in which nucleotides are represented using single-letter codes. Except for the mandatory symbol '>', all the other characters in the single initial line are optional and only used for the purpose of identification and description. The sequence ends if another line starting with the symbol '>' appears; this indicates the start of another sequence. Example sequences in FASTA format can be seen by clicking on the Example button right above the input box.

1.4 Step #4:

Users are able to set the step size which denotes length of the sub-sequences of interest. The predictor predicts the type (hotspot, coldspot or none) for each sub-sequence according to the selected step size. Example: if sequence length is 519 and step size is 200, then the model will predict the types of three sub-sequences. First one ranging from position 1 to 200 of the input sequence, the second one from position 201 to 400 and the third one from 401 to 519.

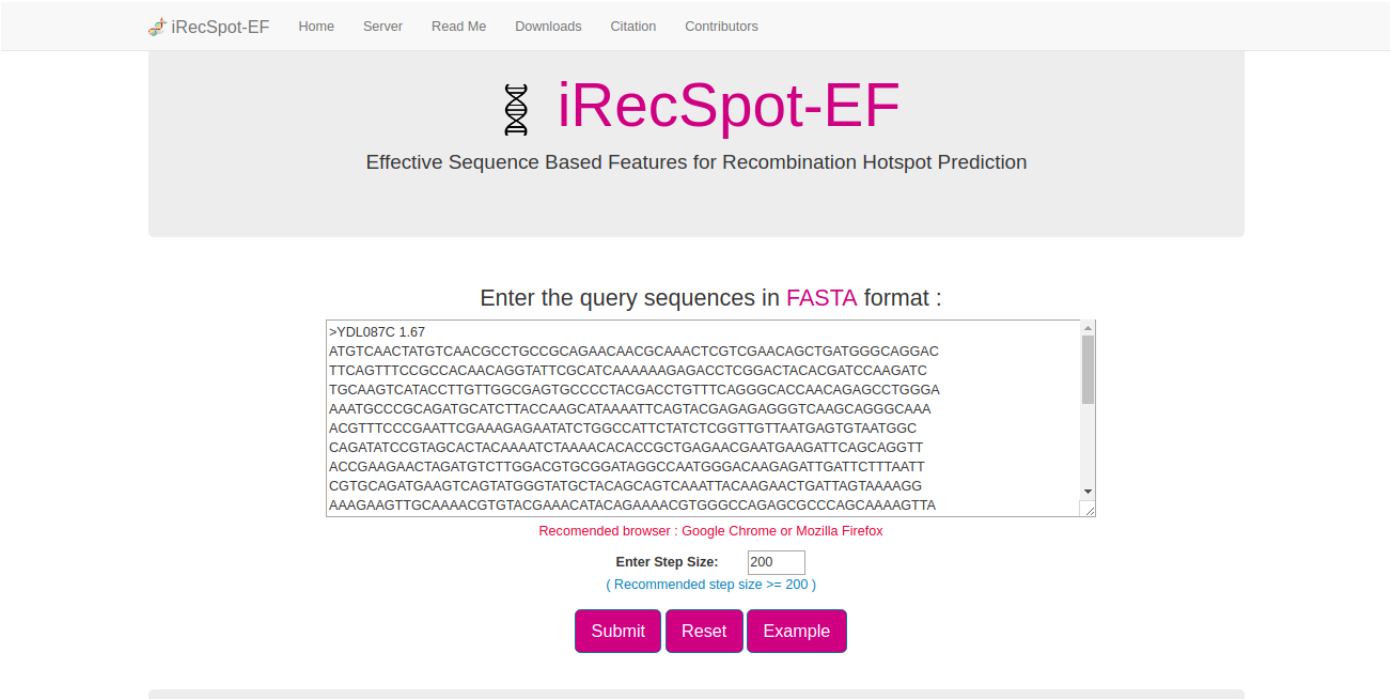

Figure 3: Enter FASTA and step size

1.5 Step #5:

Click on the 'Submit' button for predictions. Please wait, the system will take some time to generate results.

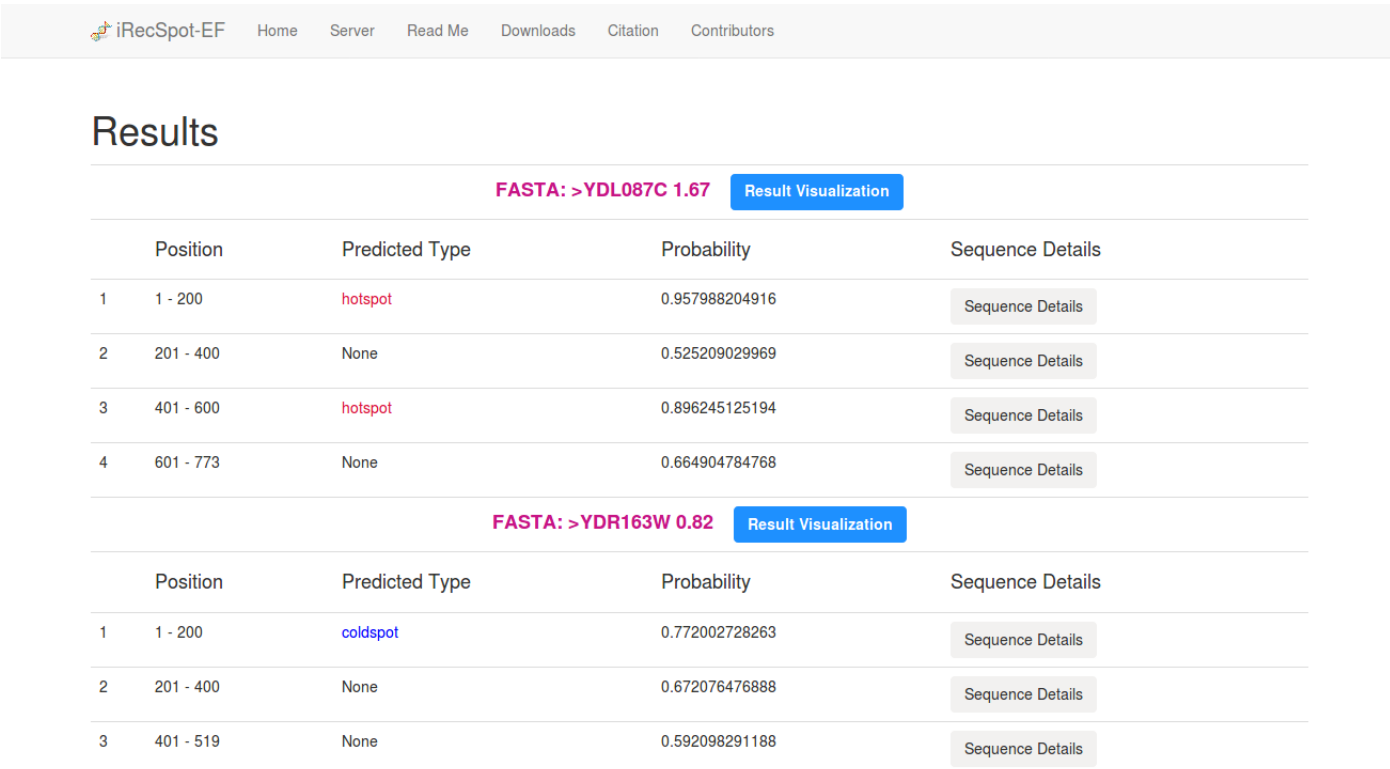

Figure 4.1: Predictions for respective sub-sequences

Note that each of the predictions is accompanied by the corresponding probability for the predicted class which can be interpreted as prediction-confidence.

1.5.1 Step #5.1:

Click 'Result visualization' to see the sequence details.

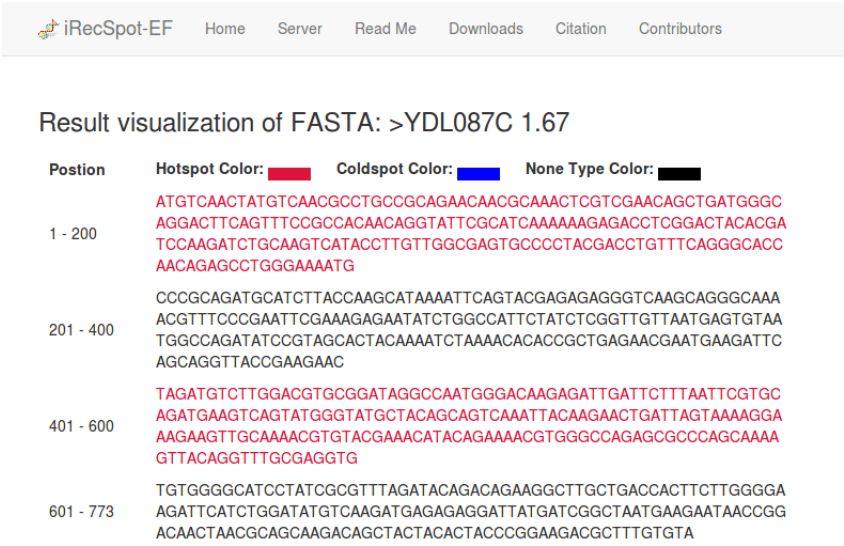

Figure 4.2: Results visualization