

crisper_spacer_array_files: This file contain the spacer arrays of the CRISPRs.

final_draft.txt: this file contain the complete information about the fasta that includes number of CRISPRs, forward/reverse strand, start/end of a particular CRISPR, consensus repeat and average repeat length.

Fasta files: it includes all the fasta files.

crisper_repeat_array_files: It contains the repeat arrays of the CRISPRs.

motif_crisper_spacer_array_files: same as **crisper_spacer_array_files** but generated from a different dataset.

motif_fasta_files: same as **Fasta files** but generated from a differnt dataset.

motif_crispr_files: Folder containing CRISPR files of various gnomes.

motif_crisper_repeat_array_files : same as **crisper_repeat_array_files** generated from a different set of data.

motif_encoded_file: contains encoding of the varoius CRISPR repeats.(trial_encoding.txt)

consensus_repeat.txt: contains consensus repeats of varoius CRISPRs along with their strand orientation.

MasterCrisprStrand: this folder contains all the possible subfolders(repeatfile, spacerfile, crisprfile, fastafile etc..) of a new dataset.

consensus_repeat_file: contains consensus repeat of the CRISPRs and their strand orientation

cp_consensus.txt: same as above without strand information

crispr_file: contains repeats and spacer information for all the CRISPRs.

encode_file: conatins encoding information about the CRISPR repeats.

fasta_file: contains the fasta files

mapped_file: same as **final_draft.txt** but contains data from a different dataset.

repeat_file: contains repeat arrays

spacer_file: contain spacer arrays

for_rev_crispr_seq: contains the forward and reverse orientation (**INPUT_F.SEQUENCE** **INPUT_R.SEQUENCE**)of a CRISPR repeat after applying the necesssary algorithms.

master_dataset_with_negative_positioning: this folder contains all the subfolders like in **MasterCrisprStrand** where the enconding is implemented keeping negative positions in in the encoded sequence. This folder also contains subfolders **crispr_genome** and **empty_genome** that talks about gnomes that conatin or doesnot contain CRISPRs. We also have **target_true_file** and **target_false_file** that is used by the learning algorithm during training.

master_dataset_X_pos_negative-positioning: same as above except the in the middle strand position is encoded as X.

master_dataset_with_positive_positioning: same as **master_dataset_with_negative_positioning** except while encoding negative sign is removed.

master_dataset_X_pos_positive_positioning: same as above except the in the middle strand position is encoded as X.

Test_dir_with_negative_positioning/Test_dir_with_positive_positioning/Test_dir_X_pos_positve_positioning/Test_dir_X_pos_negative-positioning: these folders are used to store the results of training process with respect to the four different kinds of encoding mentioned above.

m_r_d_specifications: contains files for the four different kinds of encoding with respect to various values of “m,r,d” and its associated measure values.

