

```

In [9]: import pandas as pd
import numpy as np
from IPython.display import display
import matplotlib.pyplot as plt
from sklearn.metrics import precision_recall_curve, auc, roc_curve, roc_auc_score

root_path = "data/";
bacteria = ["MYCOBACTER", "CLOSTRIDIUM", "RHODOBACTER_1", "RHODOBACTER_2", "BACILLUS" ];

aveauprc = list()
ave_roc = list()

for i_b, b in enumerate(bacteria):
    print("_"*10, i_b, " ", b, "_"*10)
    bacteria_results_path = root_path + b + "/RESULTS.csv"
    print(bacteria_results_path)
    results_df = pd.read_csv(bacteria_results_path)
    display(results_df.head())

    precision, recall, pr_thresholds = precision_recall_curve(results_df["Y"], results_df["PROB_1"])
    pr_auc = auc(recall, precision)

    fpr, tpr, roc_thresholds = roc_curve(results_df["Y"], results_df["PROB_1"])
    roc_auc = roc_auc_score(results_df["Y"], results_df["PROB_1"])

    aveauprc.append(pr_auc)
    ave_roc.append(roc_auc)

    fig, (ax1, ax2) = plt.subplots(nrows=1, ncols=2)
    fig.suptitle(b + " - AUPRC: {:.3f} ROC: {:.3f}".format(pr_auc, roc_auc), fontsize=16)
    ax1.set_xlabel('REC', fontsize=12)
    ax1.set_ylabel('PRE', fontsize=12)
    ax1.plot(recall, precision)

    ax2.set_xlabel('FPR', fontsize=12)
    ax2.set_ylabel('TPR', fontsize=12)
    ax2.plot(fpr, tpr)

print("_"*20, "SUMMARY", "_"*20)
print("\n\nAVERAGE AUPRC : {:.3f} \nAVERAGE ROC : {:.3f}\n\n".format(np.mean(aveauprc), np.mean(ave_roc) ))
print("_"*40, "\n\n\n")

```

```
plt.show()
```

_____ 0 MYCOBACTER _____
data/MYCOBACTER/RESULTS.csv

	CHROM	SEQ	Y	Y_PRED	PROB_1	PROB_2
0	NC_008596.1:5734081-5734121(+)	GGCGTTGATCTGGTTGATCACCTCACGCGCGGCGTTCTGC	0	0	0.037882	0.962118
1	NC_008596.1:6504642-6504682(-)	GATCCGCGCCAGCGCGGTGGGATGTGGCCGGCAGTCGTT	0	0	0.016838	0.983162
2	NC_008596.1:87244-87284(+)	CACCTCGACCGACGAGGACCAGGCGGGCACGCTTGCGTCG	1	0	0.016476	0.983524
3	NC_008596.1:3900038-3900078(-)	AGGGCCACGCCTGCCGCGTTGACCAGGTAGTCCAGTCCGC	0	0	0.251376	0.748624
4	NC_008596.1:3545979-3546019(-)	TGCCC GCGGTCGGCTCTTCTGCCCCCGGCTTCACGCTGAC	0	0	0.037882	0.962118

_____ 1 CLOSTRIDIUM _____
data/CLOSTRIDIUM/RESULTS.csv

	CHROM	SEQ	Y	Y_PRED	PROB_1	PROB_2
0	NC_010001.1:3515375-3515415(-)	CGATTAATTTTTCTATCGTAGGTAGGGCAAATCGCATCGA	0	0	0.357151	0.642849
1	NC_010001.1:3101731-3101771(+)	TTGACGTTACAAAATTTCTATTCATGAAATCACTCTCCT	1	1	0.816822	0.183178
2	NC_010001.1:3299696-3299736(-)	AATTAAGATATTAGCGTAAAAATCTAAGGTATTTTATACG	1	0	0.316996	0.683004
3	NC_010001.1:1900079-1900119(+)	TAATTTTATTGTACTTACATTACTTATTTTATTAACATC	0	0	0.130747	0.869253
4	NC_010001.1:1233167-1233207(+)	GAAGCAGATTCTGAATATCCGTTTGACACAGTTCCTTGA	0	0	0.361083	0.638917

_____ 2 RHODOBACTER_1 _____
data/RHODOBACTER_1/RESULTS.csv

	CHROM	SEQ	Y	Y_PRED	PROB_1	PROB_2
0	NC_014034.1:762652-762692(+)	CGGGTCAAGTTCTCCACGGGTGGGATGAGCCCCTCGTGGT	1	0	0.107955	0.892045
1	NC_014034.1:1370676-1370716(-)	ACGGGGCACAAGGTAGCTGTGCGCAAGAGCCCCGTGGTC	0	0	0.223969	0.776031
2	NC_014034.1:1173027-1173067(-)	CCCGTCGCCACCACCACCTTGCCGCCCGCAAGATCCCGT	0	0	0.088009	0.911991
3	NC_014034.1:3512430-3512470(+)	TTCGCCCTAGCGCAACAAAAAATGTGCTATTATCTTTAA	1	1	0.744681	0.255319

	CHROM	SEQ	Y	Y_PRED	PROB_1	PROB_2
4	NC_014034.1:2489880-2489920(-)	TCGTTCAAATATCACCGCCCGCGCGCCTTGTGCGCTCGG	1	1	0.838111	0.161889

_____ 3 RHODOBACTER_2 _____
data/RHODOBACTER_2/RESULTS.csv

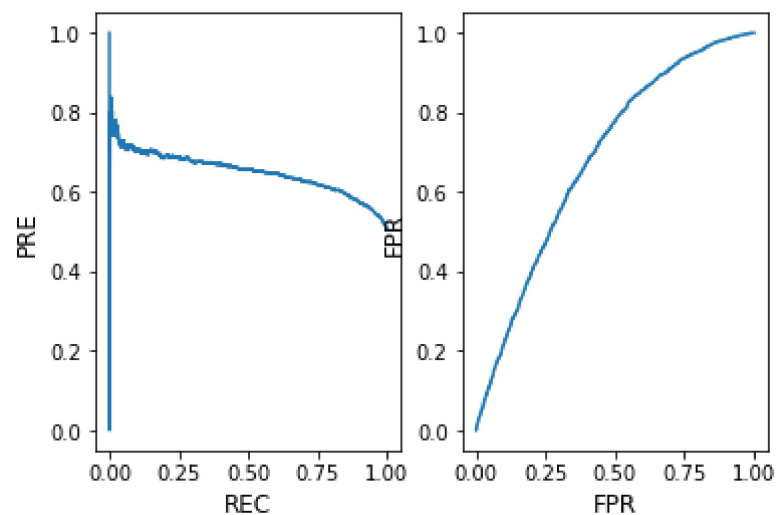
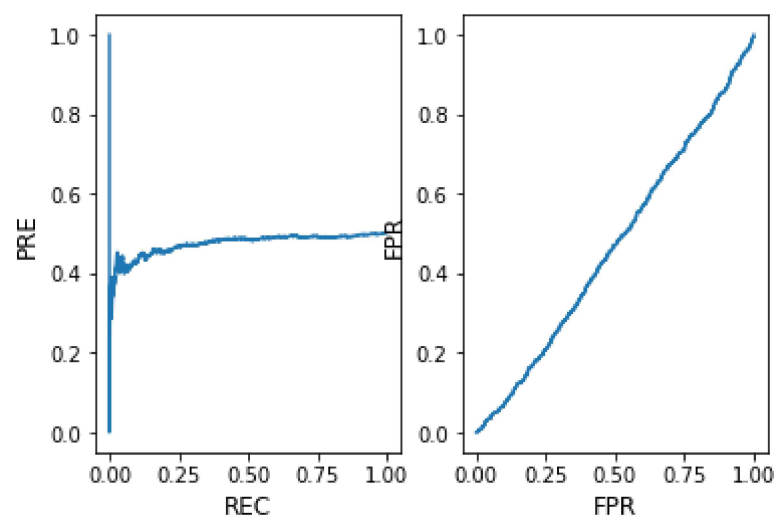
	CHROM	SEQ	Y	Y_PRED	PROB_1	PROB_2
0	NC_014034.1:2374117-2374157(+)	GACCATCGCAACGCCGACGGCGGCGCTGAAATTGTCCTGC	0	0	0.083763	0.916237
1	NC_014034.1:3657494-3657534(-)	GGCGCGTTGACGGCGGATGCAGATGGCGTGATCCGCACCA	1	0	0.237280	0.762720
2	NC_014034.1:756232-756272(+)	GTCGCAACGGCTGCATGAAATCTTGAACGCCTGCCGCTG	1	0	0.187159	0.812841
3	NC_014034.1:2964931-2964971(+)	CGGCCCCCCTTGCCCGGGTCGCGCCGCAGGGCTACGCTTG	1	0	0.040421	0.959579
4	NC_014034.1:2554357-2554397(-)	ACCTCGATGCCGTGCGCGCCGAGATCGCGCCCGAAGGGCT	1	0	0.001667	0.998333

_____ 4 BACILLUS _____
data/BACILLUS/RESULTS.csv

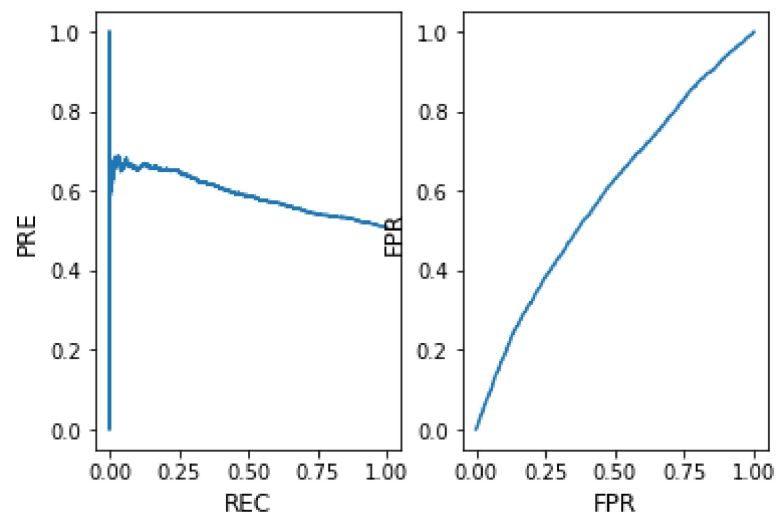
	CHROM	SEQ	Y	Y_PRED	PROB_1	PROB_2
0	CP002927.1:2362306-2362346(+)	CGCCGGTGCATAAATAATTTATATTTATACAAACCCAATA	0	1	0.863995	0.136005
1	CP002927.1:3727585-3727625(-)	GATTAGCATAATAAAGGAAAACGGGCCGAACAATCGGACA	0	0	0.455651	0.544349
2	CP002927.1:317044-317084(-)	GAATTCCAAGAACGACTGTCTCCATCAACGCTTGTCCGC	0	0	0.134536	0.865464
3	CP002927.1:3596390-3596430(+)	CAATTTGGCGCGGCGCTTTTTTGGCGATTCTCGCCAGCTG	0	0	0.018855	0.981145
4	CP002927.1:3537690-3537730(+)	GCCTTCTGATGCGAGAAATCCGACTACGAGTTCCTCTAAA	0	0	0.301800	0.698200

_____ SUMMARY _____

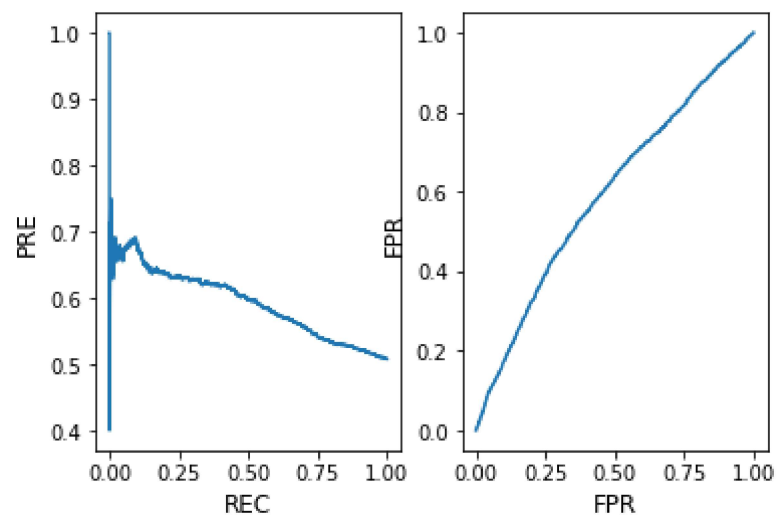
AVERAGE AUPRC : 0.592
AVERAGE ROC : 0.608

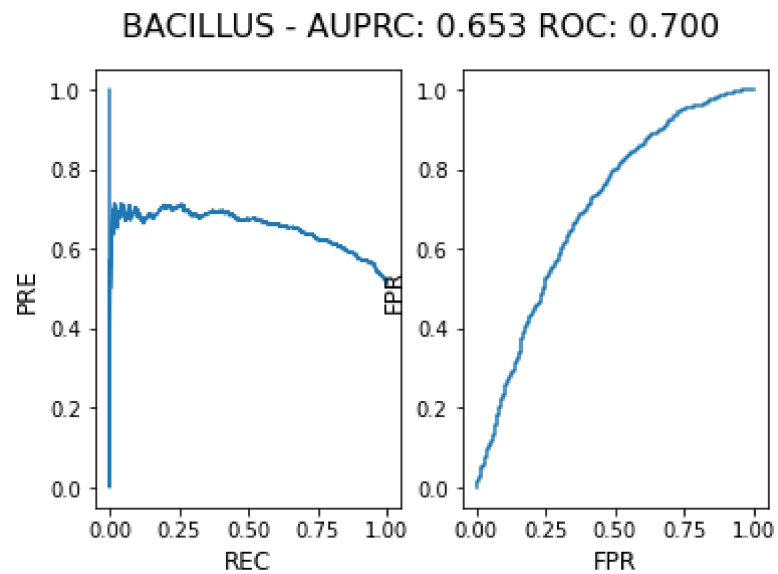
MYCOBACTER - AUPRC: 0.649 ROC: 0.684**CLOSTRIDIUM - AUPRC: 0.474 ROC: 0.470**

RHODOBACTER_1 - AUPRC: 0.591 ROC: 0.593



RHODOBACTER_2 - AUPRC: 0.593 ROC: 0.595





In []: