**Supplementary Table 4.** Comparison of the performance (RMSE) of regression tree and random forest models in predicting Topt of the bacterial community based on the relative abundance of commonly observed ASVs, at various levels of shared soils for common ASVs (i.e. n = 2 means that only ASVs present in at least 2 soils are included in the modeling). RMSE values are presented for the performance across the training set (RMSE\_train\_soils) and validation set (RMSE\_test\_soils).

|  |  |  |  |
| --- | --- | --- | --- |
| ASV shared in n= soils | model | RMSE\_train\_soils | RMSE\_test\_soils |
| 2 | full tree | 1.09 | 6.10 |
| 2 | random forest | 4.03 | 3.68 |
| 3 | full tree | 0.97 | 3.68 |
| 3 | random forest | 3.41 | 3.28 |
| 4 | full tree | 1.25 | 1.98 |
| 4 | random forest | 3.12 | 4.89 |
|  | linear model | 1.62 | 2.71 |