**Simulation study of the phylogenetic placement problem:**

**Sequence Divergence:**

Let us consider two sequences S1 and S2.

P-distance (S1, S2) =

Previously, what I was doing =

P-distance (S1, S2) =

|  |  |  |  |
| --- | --- | --- | --- |
| Mutation  Rate | Number of reticulations | Average P-distance | Std Error |
| 0.02 | 0 | 0.0848 | 0.002 |
| 5 | 0.0844 | 0.002 |
| 10 | 0.0846 | 0.003 |
| 0.06 | 0 | 0.2181 | 0.006 |
| 5 | 0.2175 | 0.006 |
| 10 | 0.2168 | 0.005 |
| 0.2 | 0 | 0.4707 | 0.008 |
| 5 | 0.4693 | 0.008 |
| 10 | 0.4695 | 0.008 |

**Gene Tree discordances:**

|  |  |  |  |
| --- | --- | --- | --- |
| Mutation  Rate | Number of reticulations | Average Normalized RF distance | Std Error |
| 0.02 | 0 | 0.47 | 0.0003 |
| 5 | 0.53 | 0.0004 |
| 10 | 0.59 | 0.0004 |
| 0.06 | 0 | 0.47 | 0.0003 |
| 5 | 0.53 | 0.0004 |
| 10 | 0.59 | 0.0004 |
| 0.2 | 0 | 0.47 | 0.0003 |
| 5 | 0.53 | 0.0004 |
| 10 | 0.59 | 0.0004 |

**Theoretical Baseline:**

For each of the query sequence, I generated a random sequence of the same length. Both the backbone tree and the backbone alignment were unchanged. Then the random sequence was placed in the backbone tree using APPLES to get the placement tree.

**Delta Errors:**

Delta Error = Nakhleh distance (N\*, P) – Nakhleh distance (N\*L, T)

Here,

N\* = Model network having {L,q} leaves

P= Placement tree of a query q having {L,q} leaves

N\*L= Model network having the L leaves

T= Backbone tree having L leaves

Let’s use the following 2 symbols to save spaces in the table:

A = Nakhleh distance (N\*, P) [ Placement tree error]

B= Nakhleh distance (N\*L, T) [Reference tree error]

**Genomic Data Results with Delta Error:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | AAPPLES | | APPLACER | | B | |
| Avg | Std Error | Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 14.1 | 0.308 | 14.522 | 0.314 | 13.85 | 0.308 |
| 5 | 36.174 | 0.152 | 36.228 | 0.155 | 35.01 | 0.169 |
| 10 | 52.088 | 0.111 | 52.214 | 0.106 | 50.828 | 0.117 |
| 0.06 | 0 | 13.962 | 0.179 | 14.208 | 0.189 | 13.762 | 0.189 |
| 5 | 35.704 | 0.179 | 35.792 | 0.168 | 34.616 | 0.181 |
| 10 | 51.044 | 0.133 | 51.018 | 0.131 | 49.682 | 0.140 |
| 0.2 | 0 | 13.206 | 0.295 | 13.42 | 0.302 | 12.982 | 0.295 |
| 5 | 36.062 | 0.190 | 36.192 | 0.195 | 34.986 | 0.204 |
| 10 | 50.83 | 0.118 | 50.716 | 0.119 | 49.38 | 0.129 |

Table 1: Delta error constituents (A and B) for 50 taxa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Delta Error (APPLES) | | Delta Error (PPLACER) | |
| Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 0.25 | 0.076 | 0.672 | 0.077 |
| 5 | 1.164 | 0.086 | 1.218 | 0.083 |
| 10 | 1.26 | 0.064 | 1.386 | 0.062 |
| 0.06 | 0 | 0.2 | 0.070 | 0.446 | 0.062 |
| 5 | 1.088 | 0.089 | 1.176 | 0.083 |
| 10 | 1.362 | 0.067 | 1.336 | 0.064 |
| 0.2 | 0 | 0.224 | 0.069 | 0.438 | 0.060 |
| 5 | 1.076 | 0.087 | 1.206 | 0.085 |
| 10 | 1.45 | 0.067 | 1.336 | 0.063 |

Table 2: Delta errors for 50 taxa.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | AAPPLES | | APPLACER | | B | |
| Avg | Std Error | Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 14.1 | 0.308 | 14.522 | 0.314 | 13.85 | 0.308 |
| 5 | 36.174 | 0.152 | 36.228 | 0.155 | 35.01 | 0.169 |
| 10 | 52.088 | 0.111 | 52.214 | 0.106 | 50.828 | 0.117 |
| 0.06 | 0 | 13.962 | 0.179 | 14.208 | 0.189 | 13.762 | 0.189 |
| 5 | 35.704 | 0.179 | 35.792 | 0.168 | 34.616 | 0.181 |
| 10 | 51.044 | 0.133 | 51.018 | 0.131 | 49.682 | 0.140 |
| 0.2 | 0 | 13.206 | 0.295 | 13.42 | 0.302 | 12.982 | 0.295 |
| 5 | 36.062 | 0.190 | 36.192 | 0.195 | 34.986 | 0.204 |
| 10 | 50.83 | 0.118 | 50.716 | 0.119 | 49.38 | 0.129 |

Table 3: Baseline delta error constituents (A and B) for 50 taxa

|  |  |  |  |
| --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Baseline Delta Error(avg)  (using APPLES) | **Normalized**  **Delta error (APPLES) with Baseline Error** |
| 0.02 | 0 | 5.163 | 0.04842146 |
| 5 | 3.65 | 0.31890411 |
| 10 | 2.416 | 0.52152318 |
| 0.06 | 0 | 4.989 | 0.04008819 |
| 5 | 3.462 | 0.31426921 |
| 10 | 2.532 | 0.53791469 |
| 0.2 | 0 | 4.487 | 0.049922 |
| 5 | 2.87 | 0.37491289 |
| 10 | 2.31 | 0.62770563 |

Table 4: Baseline delta errors and Normalized Delta error with baseline errors for genomic for 50 taxa

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | AAPPLES | | APPLACER | | B | |
| Avg | Std Error | Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 28.231 | .255 | 27.817 | .27 | 27.336 | .27 |
| 5 | 54.898 | .175 | 54.843 | .17 | 54.025 | .18 |
| 10 | 75.887 | .177 | 76.303 | .18 | 75.272 | .18 |
| 0.06 | 0 | 26.195 | .23 | 26.004 | .22 | 23.037 | .29 |
| 5 | 53.775 | .22 | 53.705 | .22 | 53.117 | .16 |
| 10 | 74.693 | .15 | 74.512 | .16 | 72.4 | .18 |
| 0.2 | 0 | 23.525 | .29 | 23.305 | .29 | 23.037 | .29 |
| 5 | 53.658 | .15 | 53.896 | .16 | 53.117 | .162 |
| 10 | 73.628 | .16 | 73.413 | .17 | 72.4 | .18 |

Table 5: Delta error constituents (A and B) for 100 taxa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Baseline Delta Error(avg)  (using APPLES) | Delta Error (APPLES) | | **Normalized**  **Delta error with Baseline Error** |
| Avg | Std Error |
| 0.02 | 0 | 6.85 | 0.895 | 0.05 | **0.13065693** |
| 5 | 4.36 | 0.873 | 0.05 | **0.20022936** |
| 10 | 3.00 | 0.71 | 0.05 | **0.24** |
| 0.06 | 0 | 6.29 | 0.49 | 0.04 | **0.07790143** |
| 5 | 3.99 | 0.849 | 0.06 | **0.21278195** |
| 10 | 3.30 | 1.167 | 0.05 | **0.35363636** |
| 0.2 | 0 | 5.18 | 0.488 | 0.04 | **0.09420849** |
| 5 | 2.77 | 0.541 | 0.05 | **0.19530686** |
| 10 | 2.6 | 1.229 | 0.05 | **0.47269231** |

Table 6: Delta error, Baseline delta error and normalized delta error for 100 taxa

**Metagenomic Data Results with Delta Error:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mutation Rate | Number of reticulations | AAPPLES | | APPLACER | | B | |
| Avg | Std Error | Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 16.388 | 0.148 | 18.63 | 0.13 | 13.34 | 0.14 |
| 5 | 37.524 | 0.123 | 39.39 | 0.12 | 35.6 | 0.13 |
| 10 | 52.31 | 0.069 | 53.36 | 0.07 | 50.26 | 0.08 |
| 0.06 | 0 | 15.652 | 0.167 | 18.16 | 0.14 | 12.84 | 0.16 |
| 5 | 37.037 | 0.106 | 38.69 | .103 | 34.82 | 0.12 |
| 10 | 52.358 | 0.063 | 53.53 | 0.07 | 50.35 | 0.07 |
| 0.2 | 0 | 15.118 | 0.153 | 17.32 | 0.14 | 11.45 | 0.14 |
| 5 | 37.38 | 0.117 | 39.09 | 0.10 | 35.11 | 0.12 |
| 10 | 51.17 | 0.067 | 52.43 | 0.07 | 49.08 | 0.07 |

Table 7: Delta error constituents (A and B) for 50 taxa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Delta Error (APPLES) | | Delta Error (PPLACER) | |
| Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 3.032 | 0.079 | 5.275 | 0.064 |
| 5 | 1.924 | 0.052 | 3.797 | 0.046 |
| 10 | 2.05 | 0.035 | 3.1 | 0.030 |
| 0.06 | 0 | 2.81 | 0.081 | 5.31 | 0.064 |
| 5 | 2.21 | 0.054 | 3.865 | 0.047 |
| 10 | 2.01 | 0.036 | 3.183 | 0.032 |
| 0.2 | 0 | 3.67 | 0.083 | 5.875 | 0.065 |
| 5 | 2.27 | 0.058 | 3.98 | 0.051 |
| 10 | 2.09 | 0.043 | 3.35 | 0.037 |

Table 8: Delta errors for 50 taxa.

|  |  |  |  |
| --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Baseline Delta Error(avg)  (using APPLES) | **Normalized**  **Delta error(APPLES) with Baseline Error** |
| 0.02 | 0 | 6.02 | **0.50365449** |
| 5 | 3.16 | **0.60886076** |
| 10 | 2.94 | **0.69727891** |
| 0.06 | 0 | 5.345 | **0.52572498** |
| 5 | 3.205 | **0.68954758** |
| 10 | 2.56 | **0.78515625** |
| 0.2 | 0 | 5.02 | **0.7310757** |
| 5 | 2.83 | **0.80212014** |
| 10 | 2.31 | **0.9047619** |

Table 9: Baseline delta errors and Normalized Delta error with baseline errors for 50 taxa

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | AAPPLES | | APPLACER | | B | |
| Avg | Std Error | Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 32.038 | .133 | 34.45 | .12 | 28.639 | .13 |
| 5 | 55.503 | .12 | 57.34 | .12 | 52.94 | .12 |
| 10 | 74.78 | .096 | 76.58 | .10 | 73.007 | .11 |
| 0.06 | 0 | 29.15 | .13 | 31.363 | .12 | 28.15 | .14 |
| 5 | 55.31 | .11 | 57.074 | .11 | 53.56 | .14 |
| 10 | 76.38 | .08 | 77.82 | .08 | 71.805 | .10 |
| 0.2 | 0 | 31.30 | .15 | 33.77 | .15 | 28.15 | .14 |
| 5 | 56.02 | .13 | 57.98 | .13 | 53.56 | .14 |
| 10 | 74.13 | .09 | 75.65 | .09 | 71.81 | .10 |

Table 10: Delta error constituents (A and B) for 100 taxa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Baseline Delta Error(avg)  (using APPLES) | Delta Error (APPLES) | | **Normalized**  **Delta error with Baseline Error** |
| Avg | Std Error |
| 0.02 | 0 | 7.04 | 3.399 | 0.06 | **0.4828125** |
| 5 | 4.66 | 2.563 | 0.05 | **0.55** |
| 10 | 3.64 | 1.773 | 0.03 | **0.60714286** |
| 0.06 | 0 | 6.73 | 3.78 | 0.07 | **0.56166419** |
| 5 | 4.04 | 2.362 | 0.04 | **0.58465347** |
| 10 | 3.4 | 2.10 | 0.03 | **0.61764706** |
| 0.2 | 0 | 4.25 | 3.15 | 0.08 | **0.74117647** |
| 5 | 3.03 | 2.46 | 0.05 | **0.81188119** |
| 10 | 2.61 | 2.32 | 0.04 | **0.88888889** |

Table 11: Delta error, Baseline delta error and normalized delta error for 100 taxa

**Type1 1 and Type 2 Error:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Type1 Error | | Type 2 Error | |
| Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 14.1 | 0.31 | 15.07 | 0.31 |
| 5 | 28.2 | 0.15 | 44.17 | 0.15 |
| 10 | 37.09 | 0.11 | 68.07 | 0.11 |
| 0.06 | 0 | 13.962 | 0.18 | 14.94 | 0.18 |
| 5 | 27.73 | 0.18 | 43.70 | 0.18 |
| 10 | 36.04 | 0.13 | 67.02 | 0.13 |
| 0.2 | 0 | 13.21 | 0.29 | 14.18 | 0.29 |
| 5 | 28.09 | 0.19 | 44.06 | 0.19 |
| 10 | 35.83 | 0.12 | 66.81 | 0.12 |

Table 12: Type 1 and Type 2 error for genomic data for 50 taxa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mutation rate | Number of reticulations | Type1 Error | | Type2 Error | |
| Avg | Std Error | Avg | Std Error |
| 0.02 | 0 | 16.72 | 0.25 | 17.7 | 0.25 |
| 5 | 29.51 | 0.211 | 45.49 | 0.21 |
| 10 | 37.39 | 0.12 | 68.374 | 0.12 |
| 0.06 | 0 | 15.71 | 0.29 | 16.68 | 0.29 |
| 5 | 29.11 | 0.18 | 45.09 | 0.18 |
| 10 | 37.36 | 0.10 | 68.33 | 0.11 |
| 0.2 | 0 | 15.27 | 0.26 | 16.25 | 0.26 |
| 5 | 29.45 | 0.20 | 45.43 | 0.20 |
| 10 | 36.24 | 0.12 | 67.22 | 0.12 |

Table 13: Type 1 and Type 2 error for metagenomic data for 50 taxa

Extreme Case: Number of reticulation 50 with 50 taxa:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Data Type | Placement Tree Error (A) | | Baseline Placement Tree Error | | Normalized Placement Tree error with Baseline error | Reference Tree Error (B) | | Delta Error | |
| Avg | Std Error | Avg | Std Error | Avg | Std Error | Avg | Std Error |
| Genomic | 96.694 | 0.90 | 97.33 | 0.91 | 0.99 | 104.9 | 0.30 | -8.258 | 0.8 |
| Meta-genomic | 97.021 | 0.94 | 97.08 | 0.95 | 0.99 | 105.04 | 0.30 | -8.02 | 0.83 |