Additional Materials

OCTAL: Optimal Completion of Gene Trees in Polynomial Time

Sarah Christensen, Erin K. Molloy, Pranjal Vachaspati, and Tandy Warnow

List of Figures

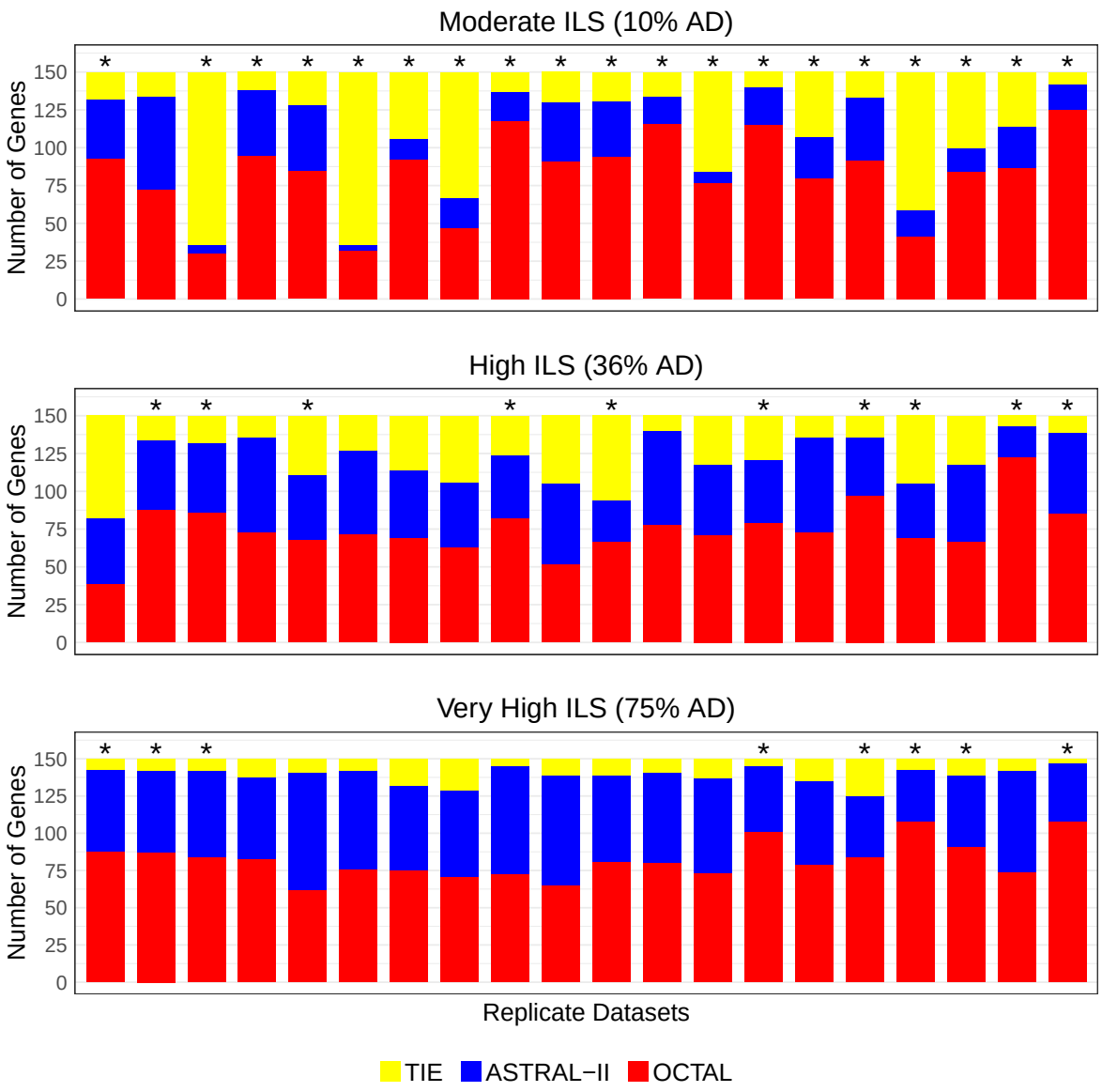

Figure S1: **The performance of OCTAL and ASTRAL-II across replicate datasets with the matching distance evaluation criterion.** Each subfigure shows the relative performance of OCTAL (using ASTRID as the reference tree) and ASTRAL-II where the matching distance was used to compare the estimated, completed gene trees to the true gene trees. The number of gene trees for which OCTAL is better than ASTRAL-II is shown in red, the number of gene trees for which ASTRAL-II is better is shown in blue, and ties are indicated by yellow. OCTAL has a statistically significant improvement over ASTRAL-II on replicates indicated with an asterisk (*).

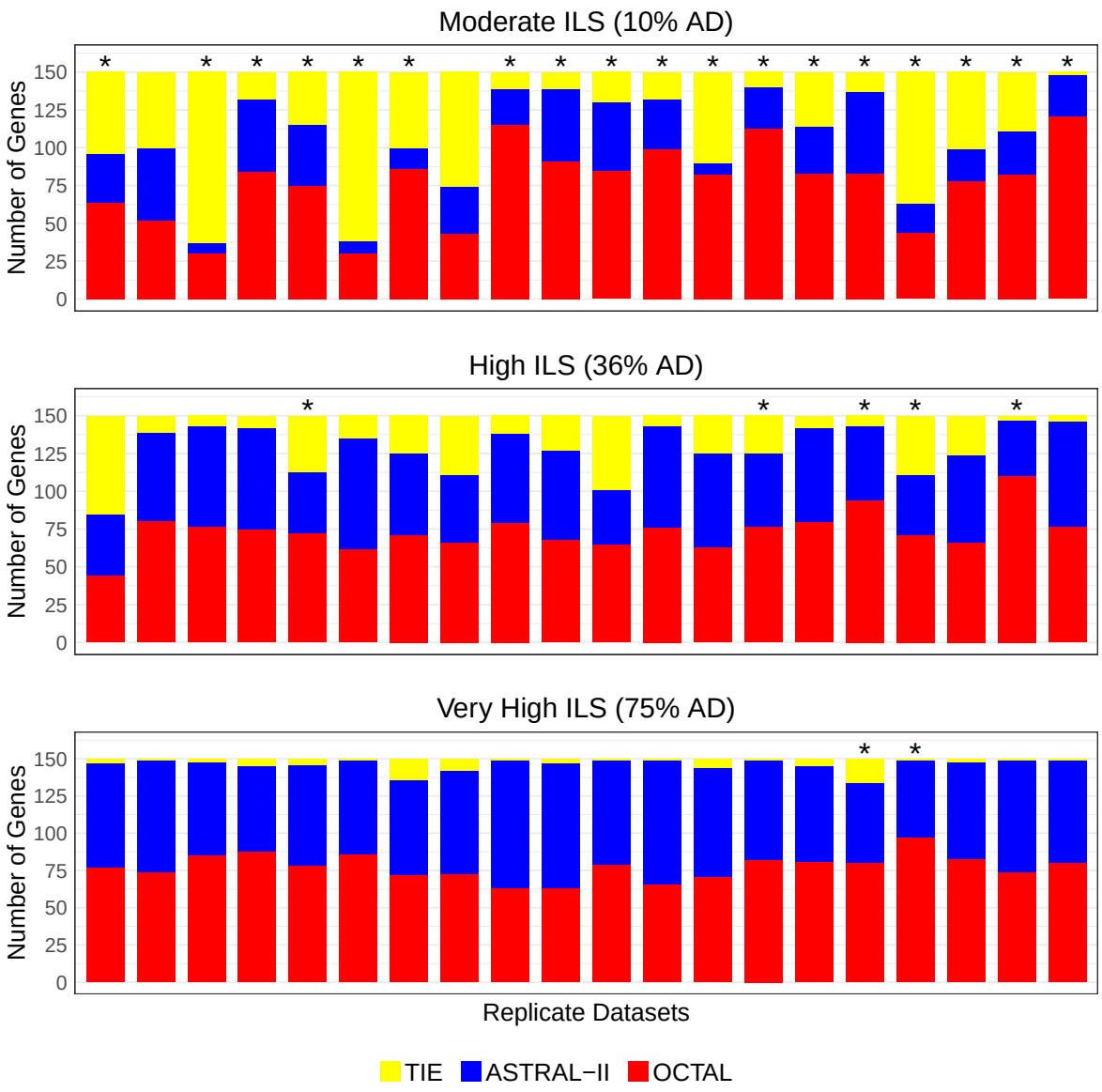

Figure S2: **The performance of OCTAL and ASTRAL-II across replicate datasets with the quartet distance evaluation criterion.** Each subfigure shows the relative performance of OCTAL (using ASTRID as the reference tree) and ASTRAL-II where the quartet distance was used to compare the estimated, completed gene trees to the true gene trees. The number of gene trees for which OCTAL is better than ASTRAL-II is shown in red, the number of gene trees for which ASTRAL-II is better is shown in blue, and ties are indicated by yellow. OCTAL has a statistically significant improvement over ASTRAL-II on replicates indicated with an asterisk (*).

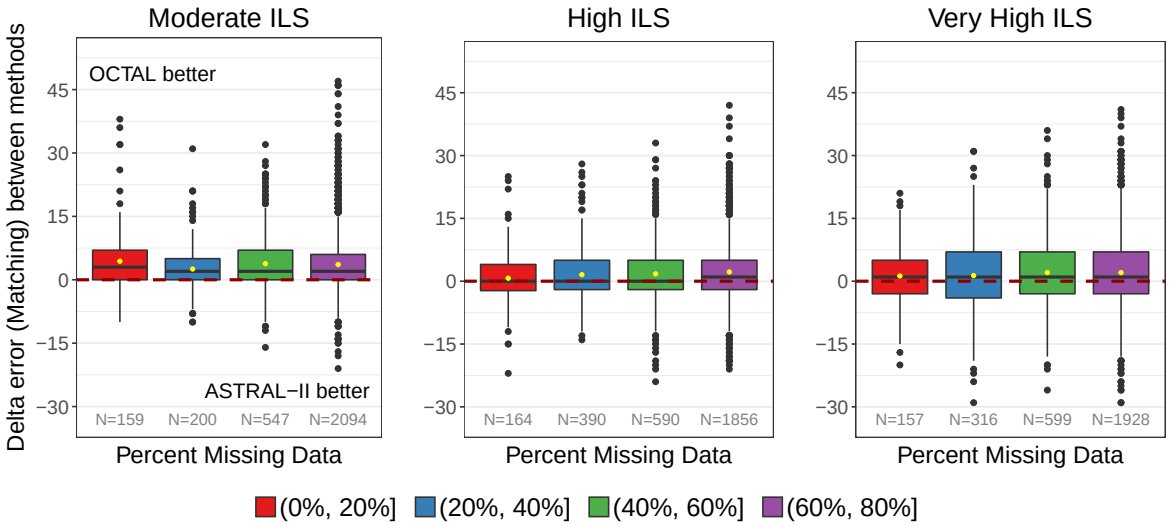

Figure S3: **The impact of degree of missing data on the relative performance of OCTAL and ASTRAL-II with the matching distance evaluation criterion.** The y-axis shows the difference in the matching distance between trees completed using OCTAL (using ASTRID as the reference tree) and ASTRAL-II. Positive values indicate that OCTAL is better than ASTRAL-II, and negative values indicate that ASTRAL-II is better. Each boxplot includes genes with the specified percent of missing data (e.g., red indicates genes are missing 0-20% of the species) for a given level of ILS. The number N of genes in each plot is also provided on the x-axis.

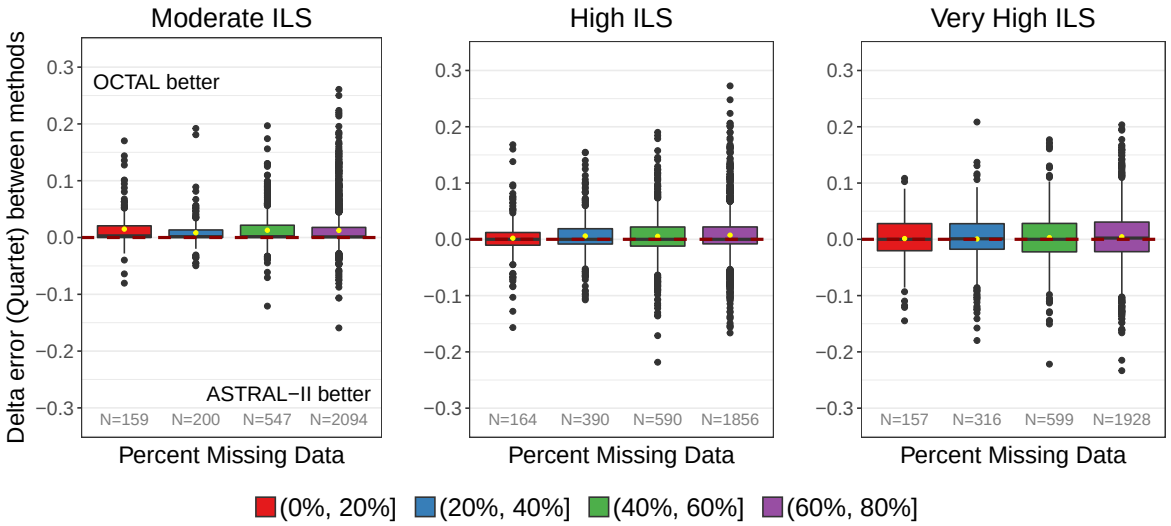

Figure S4: **The impact of degree of missing data on the relative performance of OCTAL and ASTRAL-II with the quartet distance evaluation criterion.** The y-axis shows the difference in the matching distance between trees completed using OCTAL (using ASTRID as the reference tree) and ASTRAL-II. Positive values indicate that OCTAL is better than ASTRAL-II, and negative values indicate that ASTRAL-II is better. Each boxplot includes genes with the specified percent of missing data (e.g., red indicates genes are missing 0-20% of the species) for a given level of ILS. The number N of genes in each plot is also provided on the x-axis.

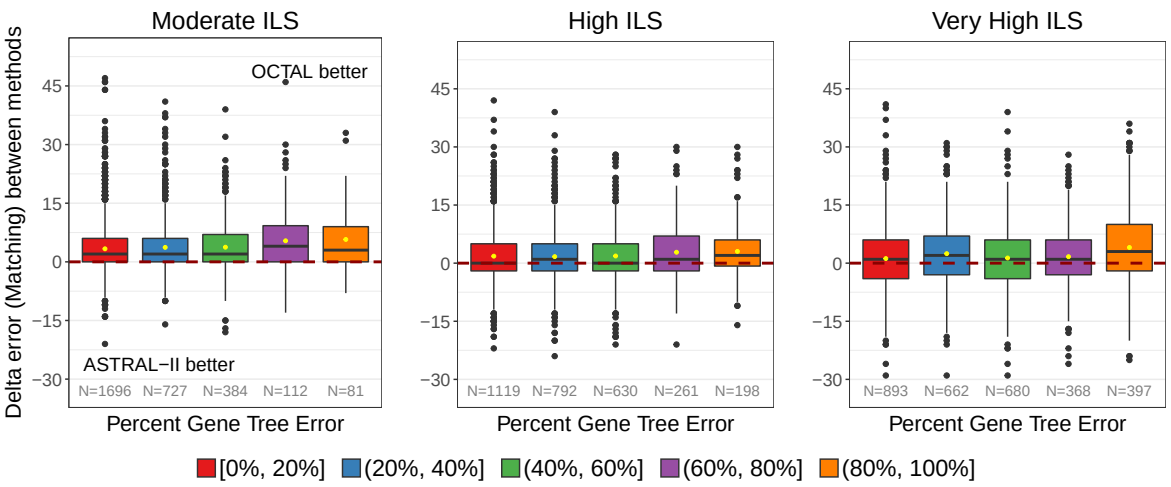

Figure S5: **The impact of gene tree estimation error on the relative performance of OCTAL and ASTRAL-II with the matching distance evaluation criterion.** The y-axis shows the difference in the matching distance between trees completed using OCTAL (using ASTRID as the reference tree) and ASTRAL-II. Positive values indicate that OCTAL is better than ASTRAL-II, and negative values indicate that ASTRAL-II is better. Each boxplot includes genes with the specified percent of gene tree estimation error (e.g., red indicates genes have 0-20% RF error) for a given level of ILS. The number N of genes in each plot is also provided on the x-axis.

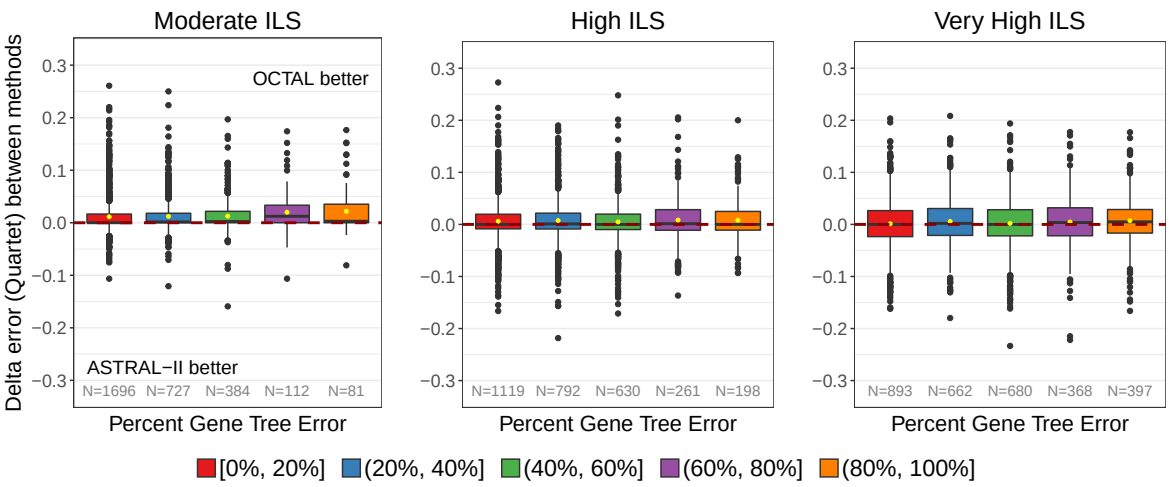

Figure S6: **The impact of gene tree estimation error on the relative performance of OCTAL and ASTRAL-II with the quartet distance evaluation criterion.** The y-axis shows the difference in the quartet distance between trees completed using OCTAL (using ASTRID as the reference tree) and ASTRAL-II. Positive values indicate that OCTAL is better than ASTRAL-II, and negative values indicate that ASTRAL-II is better. Each boxplot includes genes with the specified percent of gene tree estimation error (e.g., red indicates genes have 0-20% RF error) for a given level of ILS. The number N of genes in each plot is also provided on the x-axis.