

Statistic	Formula	Description
Coverage breadth of smaller genome	$\frac{L_{AB} + L_{BA}}{\lambda_{\min}(A, B)}$	Same as d1 below but not rescaled to a distance
Distance score d0	$1 - \frac{L_{AB} + L_{BA}}{\lambda(A, B)}$	Proportion of genomes covered by alignments
Distance score d1	$1 - \frac{L_{AB} + L_{BA}}{\lambda_{\min}(A, B)}$	Proportion of smaller genome covered by alignments
Distance score d2	$-\log \frac{L_{AB} + L_{BA}}{\lambda(A, B)}$	Rescaled variant of d0
Distance score d3	$-\log \frac{L_{AB} + L_{BA}}{\lambda_{\min}(A, B)}$	Rescaled variant of d1
Percent identity	$\frac{ID_{AB} + ID_{BA}}{L_{AB} + L_{BA}}$	Same as d4 below but not rescaled to a distance
Distance score d4	$1 - \frac{ID_{AB} + ID_{BA}}{L_{AB} + L_{BA}}$	Total number of identical base pairs across alignments relative to total coverage of alignments
Distance score d5	$-\log \frac{ID_{AB} + ID_{BA}}{L_{AB} + L_{BA}}$	Rescaled variant of d4
Distance score d6	$1 - \frac{ID_{AB} + ID_{BA}}{\lambda(A, B)}$	Total number of identical base pairs across alignments relative to combined genome size
Distance score d7	$1 - \frac{ID_{AB} + ID_{BA}}{\lambda_{\min}(A, B)}$	Total number of identical base pairs across alignments relative to twice the length of the smaller genome
Distance score d8	$-\log \frac{ID_{AB} + ID_{BA}}{\lambda(A, B)}$	Rescaled variant of d6
Distance score d9	$-\log \frac{ID_{AB} + ID_{BA}}{\lambda_{\min}(A, B)}$	Rescaled variant of d7
Breakpoints*	NA, see source code for calculation	Number of cases where an adjacent pair of alignments in one genome is not adjacent in the same relative order in the other genome
Alignments*	NA	
Breakpoint distance	$\frac{\text{breakpoints}}{\text{alignments} - 1}$	If the denominator is 0 (1 alignment), breakpoint distance is assigned 0
l10, l20..., l90*	NA	After ordering alignments from large to small, l10 is the number of alignments that must be cumulatively summed to reach 10% of total alignment length. Calculations for other %s through to 90% are conducted
n10, n20..., n90*	NA	As above, but the size of the alignment is given

*To calculate overall breakpoints, alignments, and alignment length statistics, the mean of the values in both BLAST directions (A vs B, B vs A) is calculated.

All statistics are calculated from trimmed alignments.

L_{AB} = total length of alignments from the BLAST of genome A (query) against genome B (subject)

L_{BA} = total length of alignments from the BLAST of genome B (query) against genome A (subject)

$\lambda(A, B)$ = combined size of genome A and genome B

$\lambda_{\min}(A, B)$ = twice the length of the smaller genome

ID_{AB} = total number of identical base pairs across alignments from the BLAST of genome A against genome B

ID_{BA} = total number of identical base pairs across alignments from the BLAST of genome B against genome A

The different distance scores reflect different distance concepts. For example, d6 and d7 represent [resemblance and containment](#) respectively.