Statistic	Formula	Description
Coverage breadth genome 1*	$\frac{LU_{AB1} + LU_{BA1}}{\lambda(g1)}$	Proportion of genome 1 covered by alignments prior to alignment trimming
Coverage breadth genome 2*	$\frac{LU_{AB2} + LU_{BA2}}{\lambda(g2)}$	Proportion of genome 2 covered by alignments prior to alignment trimming
Coverage breadth	$\frac{L_{AB} + L_{BA}}{\lambda(A,B)}$	Same as d0 below but not rescaled to a distance
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 alignment length
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	Total number of alignments per genome
Alignment pairs**	NA	For each aligned sequence, alignment pairs = alignments – 1 (assuming linear sequence topology). e.g. given 4 alignments A,B,C,D; the alignment pairs are A,B; B,C; C,D
Breakpoint distance d0**	$\frac{breakpoints}{alignment\ pairs}$	If the denominator is 0 (1 alignment), breakpoint distance is assigned 0
Breakpoint distance d1**	$\frac{breakpoints}{((L_{AB} + L_{BA})/2\ )/1000}$	Breakpoints expressed per kilobase of aligned sequence
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

*Coverage breadth of genome 1/genome 2 is calculated from the set of untrimmed alignments (analogous to the 'query coverage' statistic reported by NCBI web BLAST). *All other statistics are calculated from trimmed alignments.*

**To calculate breakpoints, alignments, alignment pairs, breakpoint distances and alignment length statistics, the *mean* of the values in both BLAST directions (genome A vs genome B and genome B vs genome A) is calculated. For all distance scores, as well as percent identity, coverage breadth and coverage breadth of the smaller genome, a combined total (alignment length and/or number of identical positions) is calculated from the query genome alignments of each BLAST direction. For coverage breadth of genome 1/genome 2; for each genome, a combined total alignment length is calculated from query and subject genome alignments of each BLAST direction.

$LU_{AB1} + LU_{BA1}$ = combined total length of untrimmed query and subject alignments from one genome ('genome 1') of a pairwise bi-directional BLAST comparison

$LU_{AB1} + LU_{BA1}$ = as above but for the other genome in the pairwise comparison ('genome 2')

$\lambda(g1)$ = twice the length of genome 1; **$\lambda(g2)$** = twice the length of genome 2

$L_{AB} + L_{BA}$ = combined total length of query genome alignments from BLAST of genome A against genome B, in both directions

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

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

$ID_{AB} + ID_{BA}$ = combined total number of identical base pairs across query genome alignments from the BLAST of genome A against genome B, in both directions

The different distance scores reflect different distance concepts. For example, d6 and d7 represent resemblance and containment respectively.