Bridge types

Unicycler makes SPAdes contig bridges from paths in SPAdes’ contigs paths file. These bridges connect two single-copy contigs with the contigs in the path.

Example

SPAdes contig path: 1, 17, 16, 18, 3, 1, 1

Bridge: 2-24=15=5

Score functions

Depth agreement: \( d \)

- Applies to all bridge types
- \( a, y = \frac{x}{1+10^d(x-y)} \)
- \( x = \) contig 1 depth
- \( y = \) contig 2 depth
- \( d \) is the depth difference for the two contigs

Good case

\( 0 \leq d < 0.3 \)

Bad case

\( 0.3 \leq d < 0.6 \)

Single-copy contigs have the same depth (1.8x).

Depth consistency: \( c \)

- Applies to SPAdes contig bridges
- \( \sum_{i=1}^{n} a_i \cdot y_i \cdot z_i \)
  - \( a_i \): set of alignment identities to contig 1
  - \( y_i \): set of alignment identities to contig 2
  - \( z_i \): set of alignment identities to contig 3
- \( d = \) length of contig 1
- \( e = \) length of contig 2
- \( f = \) length of contig 3
- \( k = \) minimum alignment length
- \( m = \) minimum alignment length
- \( n = \) total number of dead ends
- \( o = \) total number of dead-end contigs
- \( p = \) number of contig bridges
- \( q = \) number of contig bridges with path
- \( r = \) number of reads supporting the bridge
- \( s = \) number of reads supporting the bridge
- \( t = \) number of reads supporting the bridge
- \( u = \) number of reads supporting the bridge
- \( v = \) number of reads supporting the bridge
- \( w = \) number of reads supporting the bridge
- \( x = \) total number of reads
- \( y = \) expected number of reads in bridge
- \( z = \) expected number of reads in bridge

Good case

\( 1.05 \leq c \leq 1.1 \)

Bad case

\( c < 1.05 \)

Bridge length: \( l \)

- Applies to SPAdes contig bridges
- \( (a_1, p_1, p_2) = \begin{cases} \text{if } h > p_1, & (a_1, p_1, p_2) \\ \text{otherwise,} & (a_1, p_1, p_2) \end{cases} \)
- \( h = \) bridge length
- \( p_1 = \) paired-end mean insert size
- \( p_2 = \) paired-end insert size standard deviation

Good case

\( 1.1x \leq l \leq 2x \)

Bad case

\( l < 1.1x \)

Loop unrolling bridges

Loop unrolling bridges are a special case of SPAdes contig bridge for when a SPAdes contig bridge connects to a single-copy contig to the middle contig of a loop. In such cases, Unicycler concludes that the loop is contiguous with the contig and uses the contig depths to determine the number of times to traverse the loop.

Score: \( 100 \sqrt[4]{44d_0} \)

Good case

\( 1.6x \leq l \leq 1.8x \)

Bad case

\( l < 1.6x \)

Long-read bridge with path

Unicycler finds long reads which contains two single-copy contigs and uses them to form a long-read bridge. It then searches for a graph path corresponding to the long-read consensus sequence. If a graph path is found, that sequence is used for the bridge instead of the long-read consensus sequence.

Score: \( 100 \sqrt[4]{46 rms} \)

Good case

\( 1.3x \leq l \leq 1.5x \)

Bad case

\( l < 1.3x \)

Long-read bridge without path

When Unicycler cannot find a graph path corresponding to a bridge’s long-read consensus sequence (either due to poor homology or the absence of a path), it uses the consensus sequence directly.

This approach is less desirable, as the long-read consensus is likely to contain more errors than the short-read graph. However, it is necessary in cases when the short-read graph is incomplete and contains dead ends.

Score: \( 100 \sqrt[4]{46 rms} \)

Good case

\( 1.3x \leq l \leq 1.5x \)

Bad case

\( l < 1.3x \)

Databases

Bridge types

Graph is incomplete and contains dead ends. This approach is less desirable, as the long-read consensus is likely to contain more sequence (either due to poor homology or the absence of a path), it uses the long-read consensus instead of the long-read consensus sequence.

When Unicycler cannot find a graph path corresponding to a bridge’s long-read consensus sequence, the long-read consensus may not be homologous with the reads. The function penalizes bridges between long-read consensus and path, and this function penalizes cases where the agreement is poor.

Bridge: 1   read

Example

Bridge: 1-4.5=17=5

Depth count penalty: \( p \)

- Applies to loop unrolling bridges
- \( a = \) loop count (not rounded)
- \( b = \) loop count (rounded to nearest integer)
- \( c = \) loop count
- \( d = \) total number of dead ends at the end of contig 1
- \( e = \) start of contig 2 (0, 1 or 2)
- \( f = \) end of contig 1 and the closest dead-end contig

Good case

\( 0.7x \leq p \leq 1.0x \)

Bad case

\( p < 0.7x \)

Long loop

If a bridge’s read consensus aligns poorly to the graph path, this suggests the graph path may not be homologous with the reads. The function penalizes bridges where the read consensus has a low alignment identity to the graph path.

Alignments

Bridge connecting long contigs.

Good case

\( 1.05 \leq g \leq 1.2x \)

Bad case

\( g < 1.05 \)

Contig length: \( X \)

- Applies to all long-read bridges
- \( \sum_{i=1}^{n} a_i \cdot x_i \)
  - \( a_i \): set of alignment identities to contig 1
  - \( x_i \): length of contig 1
- \( m = \) minimum alignment length

Good case

\( 1.8x \leq X \leq 2x \)

Bad case

\( X < 1.8x \)

Dead ends: \( q \)

- Applies to long-read bridges without path
- \( d = \) total number of dead ends at the end of contig 1 and the start of contig 2

Good case

\( 0.7x \leq q \leq 0.8x \)

Bad case

\( q < 0.7x \)