Workflow to Resolve Misassembly Caused by Collapsed Repeats

Assess project status with Assembly View

- Identify clone ends
- Reorient contigs relative to scaffold
- Run crossmatch

Total size of real digests consistent with the size of major contigs + vector?

Inconsistent mate pairs?

No

Inconsistent mate pairs?

Yes

Inconsistent mate pairs?

No

Inconsistent mate pairs?

Yes

Regions with multiple high quality discrepancies?

No

Identify regions with missing data

Yes

Retrieve partners of unpaired reads from Trace Archive

Tear apart collapsed regions

BLASTN with sequence near gaps against the Trace Archive

Retrieve reads and their mate pairs from the Trace Archive

Identify regions with missing data

• Substantial decrease in consistent mate pairs density (M shape)
• Extra or larger bands in the real digests
• Repetitive regions with unusually high Q20 read depth

No major misassembly

Both inconsistent mate pairs placed in repeats?

Yes

No

Pull both mate pairs out of assembly

Pull the mate pair within repeat out of assembly

Add “Tell Phrap No Overlap” tags to real discrepancies

Retrieve partners of unpaired reads from Trace Archive

Tear apart collapsed regions

BLASTN with sequence near gaps against the Trace Archive

Retrieve reads and their mate pairs from the Trace Archive

Inconsistent mate pairs?

Yes

No

Run Miniassembly

Incorporate new reads (phredPhrap or Add New Reads)

Tear contig at the most distal high quality discrepancy

No

Real high quality discrepancies in the entire aligned region?

Yes

Join contigs

Discrepancies caused by SNPs or unclipped vector?

Yes

No

Compare assembled contigs with main contigs using crossmatch or Search for String

Alignment has high quality discrepancies?

Yes

No

Signs that a project has either missing or misplaced data