Degree of contiguity (coverage gaps)

Higher coverage (>50 x) (HGAP, PacBioToCA, short read assemblers)

Long Reads

Use short range paired reads to make pre-assembled UNITIGS (PBcR + Celera) – still hybrid, though 'self correcting'

Shorter SMRT reads within longer ones (HGAP). High coverage needed.

Short Reads

Test range of k-mers (Velvet)

Adapt for low coverage regions – use a smaller value of K (ALLPATHS-LG)

Extend contigs using mate pair information (ABySS)

Generate supercontigs (gapped contigs) using OLC method (Forge)

Correctness of assembly (base calling errors, low coverage resolution, repeats)

Remove low quality reads, ones with N's, short reads (<100 bp), ones with high error probability bases, 'chastity filter'

Long Reads

No GC bias (doesn't require PCR amplification)

Read Length > Repeat Length

Short Reads

Remove dead end paths (Base calling errors)

remove or merge divergent bubbles (polymorphisms, inserts)

Collapse repeats, compare 24-mer stacks in reads (ALLPATHS-LG)

Legend:
- Poisson distribution
- Schematic experimental k-mer coverage
- Schematic average contig coverage

Probability (y-axis)

Coverage depth (x-axis)