**A. Assembly strategy**

- Read quality filters: Scythe, Sickle
- Digital normalization: khmer
- de novo assembly (multiple k-mers): CLC Genomics
- Clustering of assemblies: CD-HIT-EST
- Repeat masking: BLAST, TREP
- Protein-guided assembly: blast2cap3

**B. Annotation**

- ORF prediction: BLASTX, findorf
- Pseudogenes analyses: findorf
- Identification of artificially fused transcripts: BLASTX, findorf
- Functional annotation: HMMER3.0, Pfam-A
- Gene Models: BLASTN, exonerate

**C. Post-assembly phasing**

- Re-mapping the reads: Novoalign
- Identification of SNPs: FreeBayes
- SNP phasing: HapCUT
- Read phasing: readphaser
- Contig sub-assembly: MIRA, CAP3

**D. Read quality filters**

<table>
<thead>
<tr>
<th>contig 5</th>
<th>A</th>
<th>G</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>contig 16</td>
<td>C</td>
<td>T</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Example reads:

- >PHASE0
  - GCAAAGATGCAATCTCGAGATCACCCA
- >PHASE1
  - GCAAGGATGCAATCTTGAGATCACCCA