

An overview of the Chinese version [11]

De Bruijn graph-based genome assembly algorithm

De Bruijn 图基因组组装算法的要点:

1. 采用短序列技术 (short-read sequencing)，将长的 DNA 序列随机切断成短片段，每个片段被称为 read。
2. 利用 De Bruijn 图对 read 进行拼接和组装。De Bruijn 图是由节点和边构成的图，节点表示长度为 k 的 DNA 片段，边表示相邻的 k-1 个核苷酸相同。
3. 根据 read 的长度和覆盖度信息，在 De Bruijn 图中寻找通路 (path) 并将它们拼接成连续的序列。
4. 在拼接过程中，需要解决 read 的错误和重复问题。错误可以通过纠错算法或覆盖度过滤来解决，重复可以通过去重算法或重叠计算来解决。
5. 组装后的序列可能存在缺失或错误，需要经过后续的验证和纠错。

CNVs (Copy Number Variations)

CNVs 是指人类基因组中存在的大片段 (通常大于 1kb) 的 DNA 序列变异，可以包括基因重复、缺失、插入和倒位等类型。CNVs 可能会影响基因的表达、蛋白质的结构和功能等，从而对个体表现出的特征和易感性产生影响。例如，一些致病基因在某些人中可能存在 CNVs，这些 CNVs 可能与某些疾病的风险有关。因此，CNVs 在人类基因组学、医学遗传学等领域都具有重要的意义。CNVs 的检测通常基于高通量测序数据，包括全基因组测序和基因组重测序等技术。

eQTL mapping

eQTL mapping (表观遗传量性状基因型关联映射) 是一种基因组学方法，用于识别基因表达和遗传变异之间的关系。eQTL 是表达量性状基因型关联的简称。它是一种关联分析技术，用于确定表达型 (例如基因表达水平) 与基因型之间的相关性。eQTL mapping 可以用来识别影响基因表达的遗传变异，从而有助于我们更好地理解基因表达调控的遗传机制。这项技术在研究人类疾病和基因功能方面发挥了重要作用。