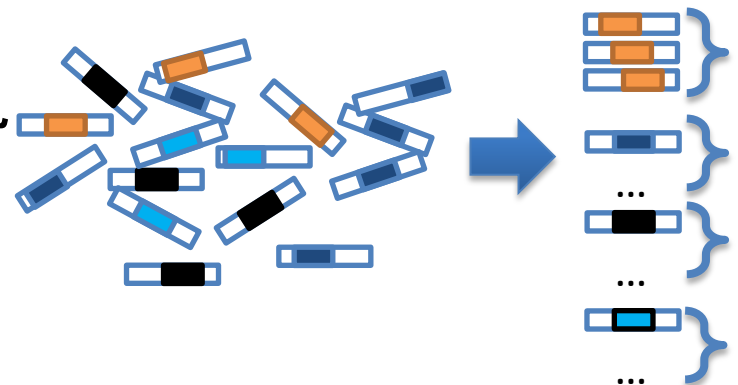


P5-35: SlideSort: 次世代シーケンサデータの ための高速全ペア類似度検索法

清水 佳奈、○津田 宏治 (産総研CBRC)

- 次世代シーケンサー: Illumina HiSeq 2000
 - リード長: 約100bp、固定
 - 2500億bpのリードを8日で生産
 - ヒトゲノム(30億bp)よりも遥かに大きい
- リード間の全ペア近傍探索
 - 編集距離 d 以内のペアを全発見
 - ゲノムアセンブリに必要
 - **SlideSort**: BWA (Suffix Array)に比して1000倍以上高速



• アルゴリズム

- Multiple Sorting Method (Uno 2008)を編集距離に拡張
- 配列をブロックに分割し、オフセットをつけてソート

