Mapping-based analysis of insertions and rearrangements in chromosomes

Lipovich Maria
Sergei Nurk

Structual variation

- Copy-number variation (CNV)
- Inversion
- Others (cryptic translocations, segmental uniparental disonomy etc.)

Database of structual variation:

http://projects.tcag.ca/variation/

Mapping-based analysis tools

- Mapping-based tools mostly infer structural variants like deletions and insertions from mate-pair reads alignment data

- Tools can possibly be classified according to

  - use of reference data

  - deterministic or probabilistic approach

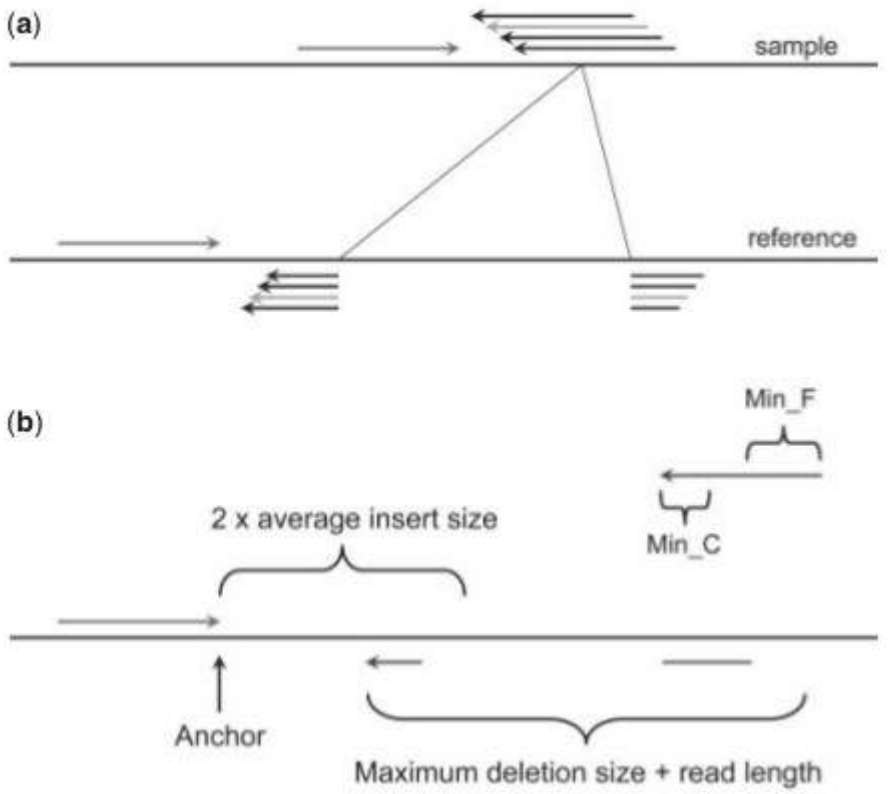
Mapping-based analysis tools

- Dindel
- Pindel
- SVDetect
- GASV
- GASVPro
- etc.

Pindel

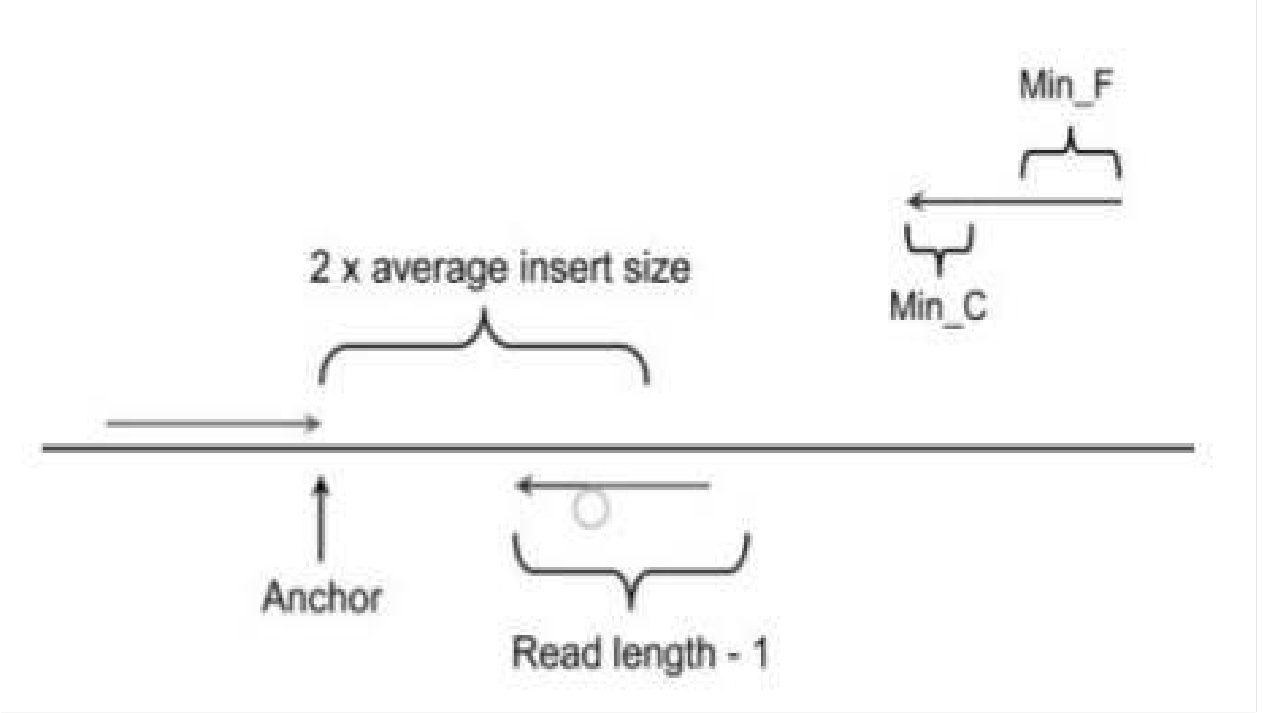
Detecting deletion events:
- uses pattern-growth algorithm to search min and max
unique substrings

- provides with deletions from small to large



Pindel

Detecting insertion events:
- uses pattern-growth algorithm too
- provides with medium sized insertions



Dindel

Outline of an algorithm

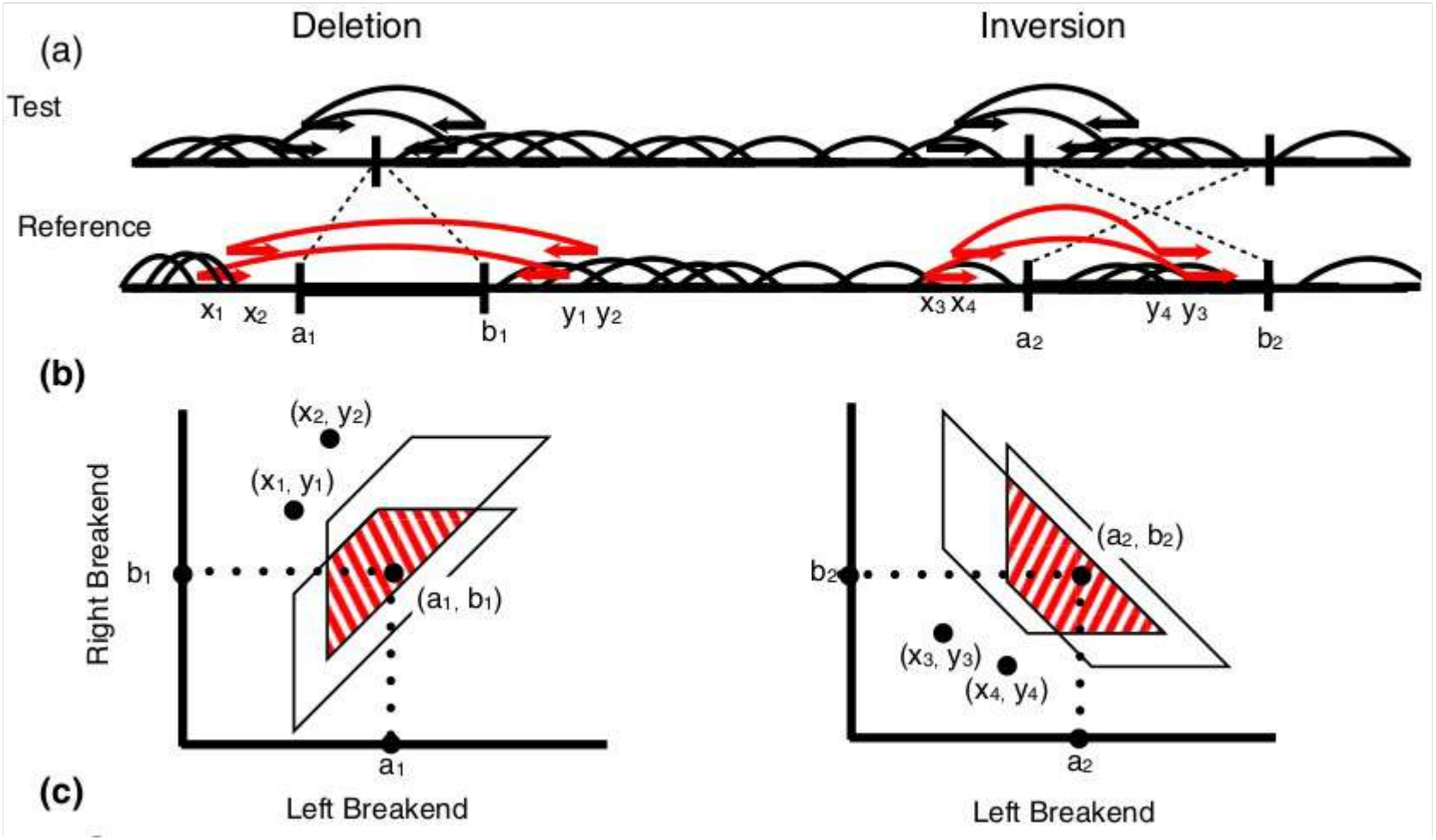
- Read mapping
- Collect candidate indels
- Define realignment windows

For every realignment window:

- Generate candidates haplotypes
- Realign reads to candidate haplotypes
- Estimate posterior probability of candidate indels

Gasv and GasvPro

Detecting deletion and inversion coordinates via geometric approach (deterministic and probabilistic)



Aims for mapping-based analysis

- *Method for detecting large insertions*
- Introducing inexact matching to the algorithms due to errors in reads
- Incorporate breakpoint uncertainty into databases of known structural variants
- Define optimal fragment size to use for studies of human structural variation