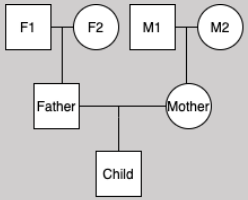
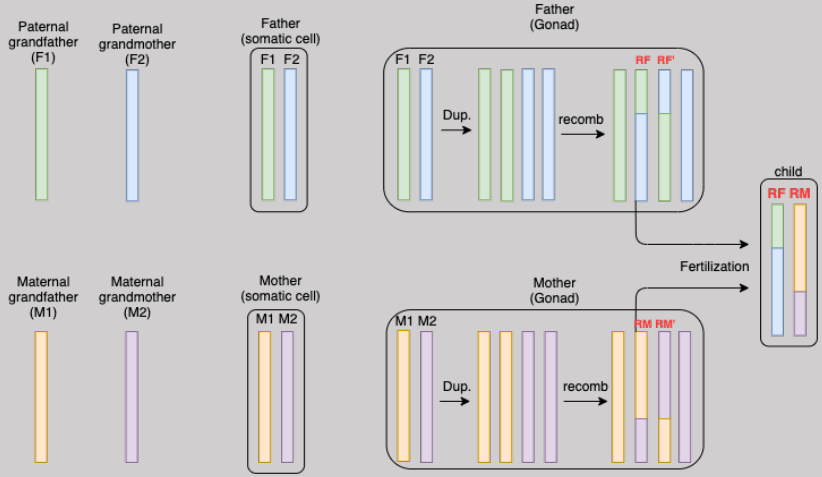
@SinaMajidian

PhaseRE: Identifying recombination location per sample using long reads

Homologous recombination is vital for chromosomal disjunction and has a significant contribution to genomic stability during the meiosis process. The process is associated with repair of double strand breaks (DSBs) on loci among the genome promoting precise chromosome segregation. The DSB repair leads to either crossing over (reciprocal exchange of chromosomal regions) or gene conversion (short DNA tract exchange). Such exchanges in genetic material are a source of genetic diversity. Recombination has also an important role in evolution, specifically speciation and gene flow in the population. The rate of recombination across chromosomal regions is affected by several factors including gender, age, distance to the telomeres/centromeres and enzymes. **Goal of project** is to estimate positions of recombination events from a trio of genomes.

Illustration of a recombination event:



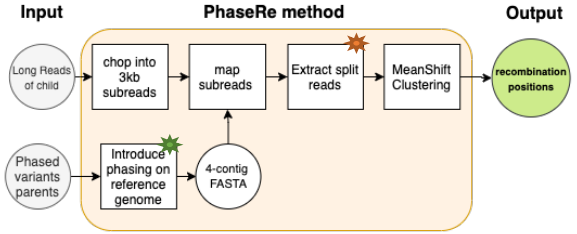


Current state of the art includes methods based on population, pedigree and gamete. However, detecting recombination in individual genomes has proven challenging due to the short span of NGS reads. However, advances in long-read technology and phasing algorithms offer an opportunity to tackle this problem. Detecting recombination potentially deepen our understanding of selection, aneuploidies, infertility, and carcinogenesis.

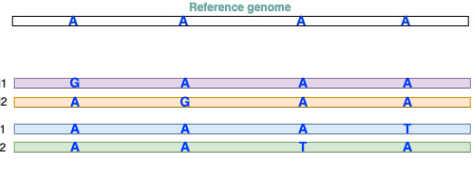
Sina Majidian[1,2], Chen-Shan Chin[3], Christophe Dessimoz[1,2,4], Fritz J Sedlazeck[5]

[1] Department of Computational Biology, University of Lausanne. [2] SIB Swiss Institute of Bioinformatics. [3] DNAnexus. [4] Department of Computer Science, University College London. [5] Human Genome Sequencing Center, Baylor College of Medicine.

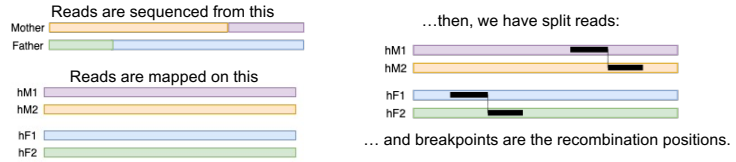
Here we present a cost-efficient method (PhaseRE) to identify recombination on a single genome level and accurately pinpoint the number and positions of recombination events.



Introducing phasings of parents on the reference genome:

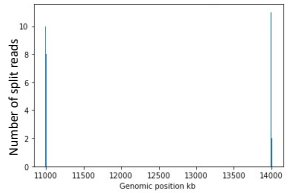


Extracting split reads:



Preliminary results (simulation):

- Genome simulated on Chr22 using SURVIVOR with 10 INDELS & Duplications.
- Simulated HiFi reads with depth of 30 using PBSIM.
- Recombination events were correctly detected at 11/14mb.



Challenges: Switch errors, Duplications, Evaluation on real data.