

Haplotype Inference for Pedigrees with Few Recombinations

B. Kirkpatrick

Intrepid Net Computing

June 8, 2016

ISBRA 2016

Slides:

www.intrepidnetcomputing.com/isbra2016.pdf

Outline

- 1 Genetics
- 2 Min. Recombination (MR) Graph
- 3 Parity Constraint Sets
- 4 Contributions

Contribution

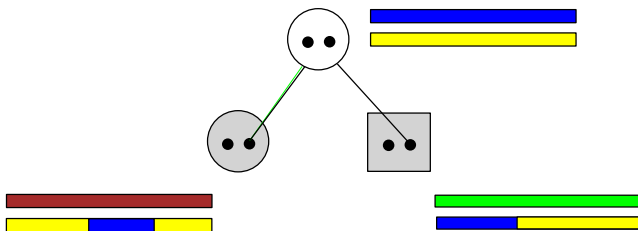
Before this work there were two relevant results for the minimum recombination haplotype configuration (MRHC) problem:

- 1 Li and Jiang showed that MRHC is NP-complete
- 2 Doan and Evans gave an algorithm parameterized by the number of recombinations

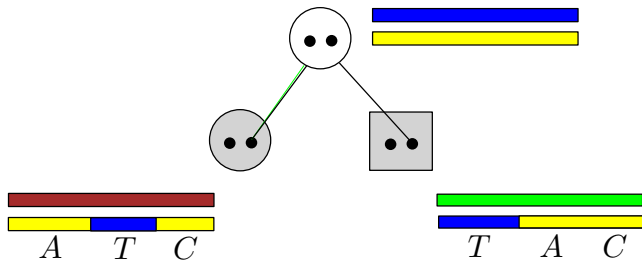
Due to this work we know

- 1 The algorithm given by Doan and Evans was correct in design and incorrect in detail
- 2 The algorithm is corrected in this work
- 3 Open Problem: is there is a FPT algorithm?

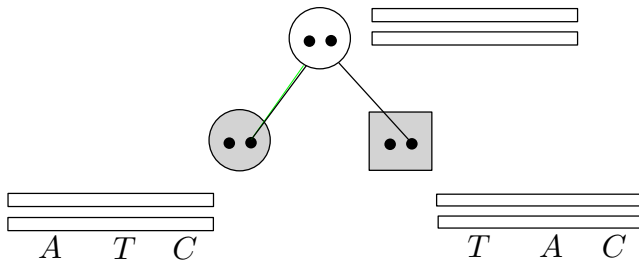
Human Genetics



Human Genetics



Human Genetics



Pedigrees or Family 'Trees'

Directed acyclic graph++

nodes individuals

boxes male

circles female

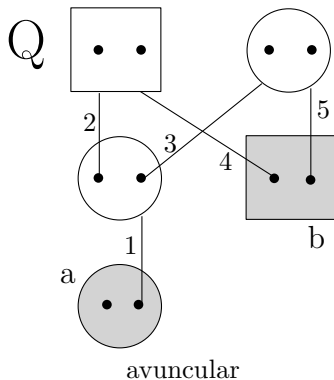
shaded has data

edges from parent to child

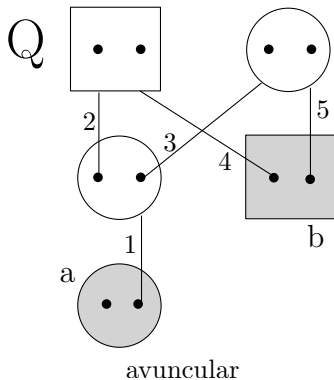
Probability model

dots diploid alleles

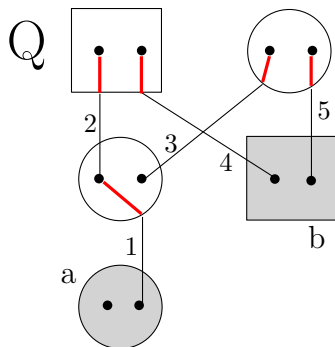
edges inheritance random
variable



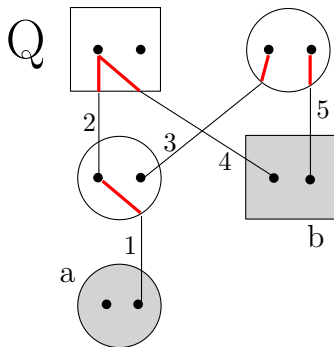
Inheritance Path



Inheritance Path



Inheritance Path

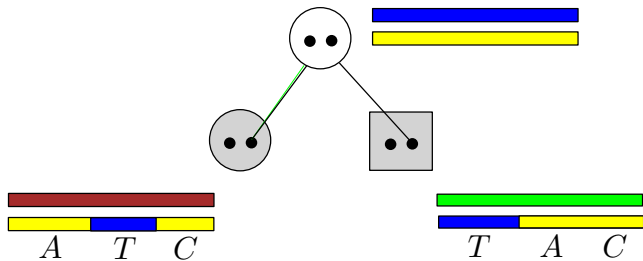


$O(2^{2n})$ inheritance paths

Why use families?

- 1 Relationship inference - the pedigree that best generates the data
- 2 Pedigree correction - which edges of pedigree are incorrect
- 3 Forensics - infer a relationship to decide which individuals are suspicious
- 4 Disease gene-finding - use a pedigree to find genes that correlate to a disease
- 5 Recombination rate inference - use a pedigree to predict recombinations

Data



Max Genomic Interval

```
ATAAAAA      AAAAAAAA  
AAAAAAAT     AAAAAAAA
```

```
ATAAAAAT  
AAAAAAA
```

Max Genomic Interval

ATAAAAAA
AAAAAAAT

AAAAAAA
AAAAAAA

ATAAAAT
AAAAAAA

Max Genomic Interval

ATAAAAAA
AAAAAAAT

AAAAAAA
AAAAAAA

ATAAAAAA
AAAAAAA

Max Genomic Interval

ATAAAAAA
AAAAAAAT

AAAAAAA
AAAAAAA

ATAA|AAAT
AAAAAAA

Max Genomic Interval

ATAAAAAA
AAAAAAAT

AAAAAA
AAAAAA

ATAAA AAT
AAAAAA

Max Genomic Interval

ATAAAAAA
AAAAAAAT

AAAAAAA
AAAAAAA

ATAAAAAT
AAAAAAA

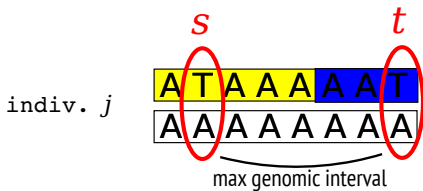
Max Genomic Interval

ATAAAAAA
AAAAAAAT

AAAAAA
AAAAAA

ATAAAAAT
AAAAAA

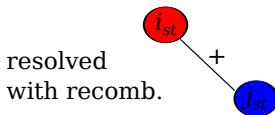
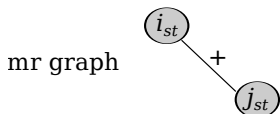
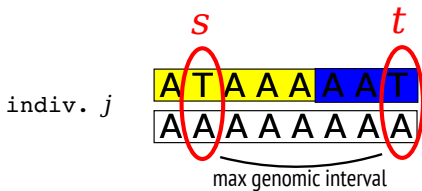
Min Recombination Graph



Min Recombination Graph

ATAAAAAA
AAAAAAAT

AAAAAA
AAAAAA



Colorings

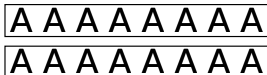
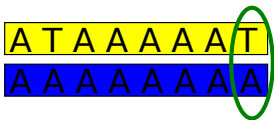
Any gray node in an MR graph can be colored either blue (if the minor alleles inherited in the same haplotype or red (otherwise) as long as the parity constraints are satisfied.

Every gray node must be assigned either red or blue.

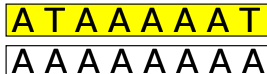
Plus Edge: no recombination when the assigned colors of the nodes are identical, one recombination otherwise

Minus Edge: no recombination when the assigned colors of the nodes are opposite, one recombination otherwise

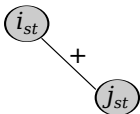
Min Recombination Graph



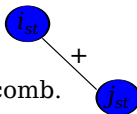
individ. j



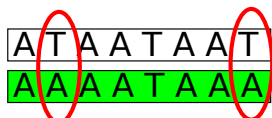
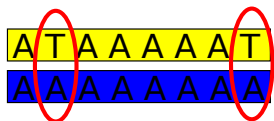
mr graph



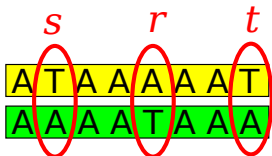
resolved
with no recomb.



Parity Constraints



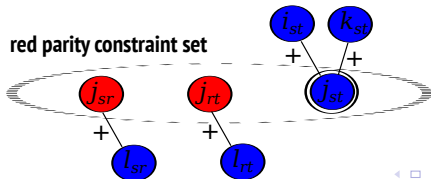
child j



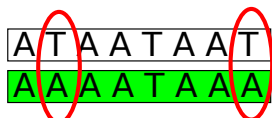
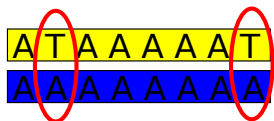
grandchild l



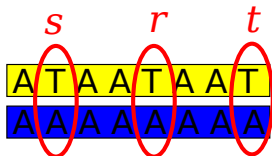
mr graph



Parity Constraints



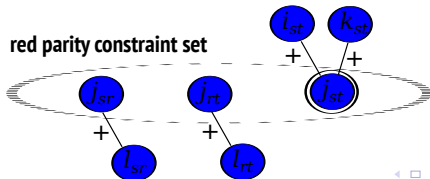
child j



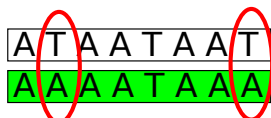
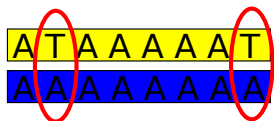
grandchild l



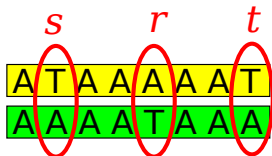
mr graph



Parity Constraints



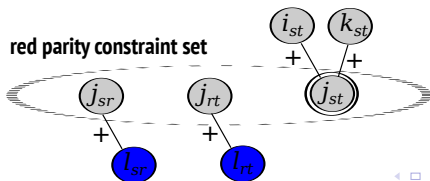
child j



grandchild l



mr graph



Bipartization

As suggested by Doan and Evans, coloring the MR graph can be reduced to an **edge bipartization** problem.

Edge bipartization: given a graph, break the minimum number of edges such that adjacent vertices belong to opposite colors groups (blue/red)

The reduction involves replacing all positive edges in the MR graph with two negative edges.

Algorithm

The algorithm to find the edge bipartization is exponential with the number of recombinations as an exponent.

Open problem: is there a fixed-parameter tractable algorithm with the parameter being the number of recombinations.

Contributions

- 1 Enumerated the cases used to build the MR graph.
- 2 Identified impossible cases.
- 3 Corrected parity constraint sets.
- 4 Proofs of equivalence between haplotype phasings and MR graph colorings.
- 5 MRHC algorithm after reducing to a simplified graph coloring problem.

Thanks to....

- Mathias Mnich
- Intrepid Net Computing