A Heuristic to Infer ARGs

Summarizing the algorithm of “Mapping Trait Loci by Use of Inferred Ancestral Recombination Graphs” by Minichiello & Durbin
The Hypothesis

- Assuming any base in the genome mutates at most one time,
- if we can figure out the family tree, there will be one branch that all the sick people fall under
- because the disease mutation occurred on that branch
The Difficulty

Given a set of people, there are many possible trees to explain their genotypes!
The Ideal Solution

Find the tree with the minimum number of recombinations and mutations
A Feasible Solution

The algorithm presented in this paper
<table>
<thead>
<tr>
<th>100</th>
<th>001</th>
<th>010</th>
<th>011</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>U</td>
<td>D</td>
<td>D</td>
</tr>
</tbody>
</table>
Rule #1

If you see one haplotype differs from all the others in one spot, assume a mutation happened there
Rule #2

When nothing else can be done, find a shared segment and split one haplotype
Rule #3

When two haplotypes are identical, "coalesce" them
Rule #1

If you see one haplotype differs from all the others in one spot, assume a mutation happened there.
Rule #3

When two haplotypes are identical, ”coalesce” them
Rule #1

If you see one haplotype differs from all the others in one spot, assume a mutation happened there.
Rule #3

When two haplotypes are identical, ”coalesce” them
Restrictions

➲ Perform a recombination only if no mutations or coalescences are possible
➲ If multiple mutations and/or coalescences are possible, the order is chosen arbitrarily
Restrictions

- Coalesce only when the overlap has “known” (not '.') sequence
- Recombine longer segments first
- After recombination, the first coalescence must be based on the shared segment