# Complexity and Approximation of the Minimum Recombination Haplotype Configuration Problem

Lan Liu[1], Xi Chen[3], Jing Xiao[3], and Tao Jiang[1,2]

[1] Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA
[2] Center for Advanced Study, Tsinghua University, Beijing, China
{lliu, jiang}@cs.ucr.edu
[3] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P.R. China
{xichen00, xiaojing00}@mails.tsinghua.edu.cn

**Abstract.** We study the complexity and approximation of the problem of reconstructing haplotypes from genotypes on pedigrees under the Mendelian Law of Inheritance and the minimum recombinant principle (MRHC). First, we show that MRHC for simple pedigrees where each member has at most one mate and at most one child (*i.e.* binary-tree pedigrees) is NP-hard. Second, we present some approximation results for the MRHC problem, which are the first approximation results in the literature to the best of our knowledge. We prove that MRHC on two-locus pedigrees or binary-tree pedigrees with missing data cannot be approximated (the formal definition is given in section 1.2) unless P=NP. Next we show that MRHC on two-locus pedigrees without missing data cannot be approximated within any constant ratio under the Unique Games Conjecture and can be approximated within ratio $O(\sqrt{\log(n)})$. Our L-reduction for the approximation hardness gives a simple alternative proof that MRHC on two-locus pedigrees is NP-hard, which is much easier to understand than the original proof. We also show that MRHC for tree pedigrees without missing data cannot be approximated within any constant ratio under the Unique Games Conjecture, too. Finally, we explore the hardness and approximation of MRHC on pedigrees where each member has a bounded number of children and mates mirroring real pedigrees.

**Keywords:** Haplotyping, pedigree, recombinant, SNP, complexity, approximation, L-reduction, positive result, negative result, bounded number, children, mates.

## 1 Introduction and Definitions

The secret mechanism behind phenotypic variation and inheritance has intrigued the study of genetic markers. With the discovery of genetic markers such as microsatellite DNA sequences and Single Nucleotide Polymorphisms (SNPs), it is now possible to provide a unique genetic map to track the variation and inheritance of genetic markers. The international HapMap project launched in October 2002, aims to discover the haplotype structure of human beings and examine the common haplotypes among populations [17].

Homologous recombination, the combination of genetic material between chromosome pairs during meiosis, is essential in diploid organisms such as humans [7]. Unfortunately, the diploid structure of humans makes it very expensive to collect haplotype data directly to display the recombination events. In a large-scale sequencing project, genotype data instead of haplotype data are collected. However, haplotype data are required in many genetic marker applications, such as linkage disequilibrium analysis and disease association mapping to name a few [12,13]. Therefore, combinatorial algorithms and statistical methods to reconstruct haplotypes from genotypes (*i.e.* the haplotype phasing or inference problem) are urgently needed.

The input data for this problem can be SNP fragments from an individual, genotype data in a population or genotype data in a family [8,9,10,11,15]. There are many combinatorial [1,2,14,16] and statistical ways [11,19] of tackling the phasing problem. They are usually quite computationally demanding.

Some of the commonly used combinatorial methods [1,2,14,16] take advantage of the availability of pedigree data. In other words, given a pedigree and the genotype information, they reconstruct a haplotype configuration for each individual in the pedigree by trying to solve the Minimum Recombinant Haplotype Configuration (MRHC) problem [1]. During the process of reconstruction, the minimum recombinant criterion is used as the objective function. Because this objective attempts to reduce the number of candidate haplotype configurations, it naturally preserves common haplotype structures.

All the existing methods to the MRHC problem are time and space consuming for realistic applications. For example, a Pentium IV computer with 256MB RAM is used to solve MRHC on an input pedigree with 29 members and 51 SNP markers. An effective combinatorial algorithm ILP takes about 5 hours to find an exact solution, whereas a well-known statistical approach SimWalk2 takes even more than 6 days to find a haplotype configuration with the maximum likelihood [21]. While over 5 millions of SNPs have been identified in the public database dbSNP [17], there is a great need for efficient algorithms that could scale up to the whole genome level. This difficulty motivates us to analyze the hardness and approximability of MRHC problems from a theoretical point of view.

## 1.1   Formal Definition of the MRHC Problem

In this subsection, we give a formal definition of the MRHC problem as well as the issue of pedigree representation and biological background. We follow the conventions in [1].

**Definition 1.** *A pedigree graph is a connected directed acyclic graph (DAG) $G=\{V, E\}$, where $V= M \cup F \cup N$, M represents the male nodes, F represents the female nodes, N represents the matting nodes, and $E= \{ (u, v): u \in M \cup F$ and $v \in N$ or $u \in N$ and $v \in M \cup F\}$. M $\cup F$ is called individual nodes. The in-degree of each individual node is at most one. The in-degree of a mating node must be two, with one edge starting from a male (called the father) node and the other edge from a female node (called the mother) and the out-degree of a mating node must be larger than zero.*

In a pedigree, the individual nodes outgoing from a mating node are called the *children*. The individual nodes with zero in-degree are called the *founders*. The induced

subgraph by the father, the mother and one child adjacent to the same mating node is called a *family trio*. If there are two node-disjoint paths between two mating nodes in the pedigree graph, this pedigree has a *mating loop*. A pedigree without mating loops is called a *tree pedigree*. A pedigree where each member has at most one mate and at most one child looks like a binary tree, so this kind of pedigree is called a *binary-tree pedigree*. Fig. 1 demonstrates an example pedigree drawn in both the formal and conventional ways. In the conventional way, the mating nodes are omitted. For convenience, we use conventional drawings of pedigrees throughout this paper.



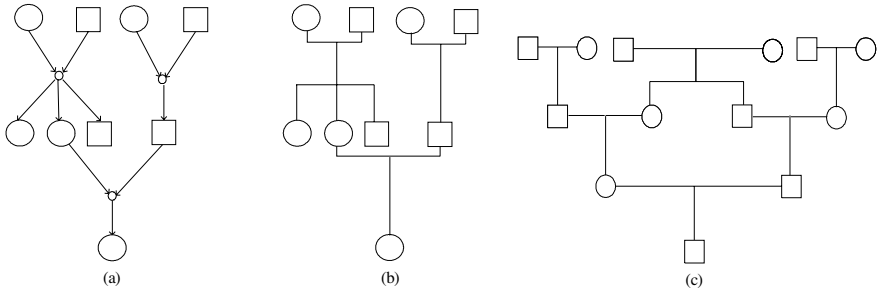(a)                              (b)                              (c)

**Fig. 1.** (a) A pedigree drawn in the formal way. (b) The pedigree drawn in the conventional way. (c) A pedigree with a mating loop.

A *genetic marker* is a short non-redundant discriminative DNA sequence that can be used to trace inheritance. Some common genetic markers are microsatellite DNA sequences or SNP data. Each polymorphism state of a genetic marker is called an *allele*. Different kinds of markers have different numbers of alleles. For instance, a microsatellite marker has multiple possible alleles occurring at a locus, which is called *multi-allelic*. An SNP marker commonly has only two possible alleles occurring at a locus, which is called *bi-allelic*. We will mostly be interested in bi-allelic markers because they are becoming the most popular markers in practice. Bi-alleles can be in exactly one of the two alternative states, such as 1 or 2. If an allele is missing at some locus, it is denoted as a "*".

In diploid organisms, because chromosomes come in pairs, at each locus there is a pair of alleles, which is referred to the *genotype* of this locus. If these alleles are the same, the genotype at this locus is *homozygous*; otherwise, the genotype is *heterozygous*. The alleles on the same chromosome form a *haplotype*. Each individual has a pair of haplotypes.

If there is no genetic mutation in a meiosis process, the child inherits one haplotype from the mother and the other one from the farther. This is the well-known *Mendelian law of inheritance*. The haplotype inherited from the mother is called the *maternal haplotype* while the one from the father is called the *paternal haplotype*. Given a pair of haplotypes of an individual, if it is known which one was inherited from his (or her) father and which was from his (or her) mother, the haplotypes and the inheritance information together are called a *haplotype configuration* (*i.e.* a *configuration* in short); otherwise, the haplotypes without inheritance information form a *haplotype grouping* (*i.e.* a *grouping* in short).

Usually, an entire haplotype of the mother's (or father's) haplotype pair is passed onto the child during meiosis. However, crossover between the haplotype pair might occur, where the haplotype pair gets shuffled and one of the mixed haplotypes is passed onto the child. This crossover is called a recombinant.

A PS (or phase) value represents the paternal or maternal information about the alleles at a locus. The PS value can take the values 0 or 1, where 1 means that the allele with the smaller identification number is from the mother and the allele with the larger identification number is from the father, and 0 otherwise. Thus, the reconstruction of haplotype configuration for an input pedigree can be viewed as assigning PS values to each locus of every member of the pedigree.

Now, the MRHC problem is defined as follows:

**Definition 2** (MRHC [1]). *Given a pedigree and genotype information for each member of the pedigree, find a haplotype configuration of the pedigree that obeys the Mendelian law of inheritance and requires the minimum number of recombinants.*

## 1.2   Variants of MRHC and Some Related Problems

We give the definitions of the variants of MRHC and list the related problems that are going to be discussed later in the paper.

**Definition 3.** *MRHC(k, j) is defined the same as MRHC except that each member in the pedigree has at most k mates and at most j children with each mate. Binary-tree-MRHC is defined as MRHC on a binary-tree pedigree. Binary-tree -MRHC\* is defined the same as binary-tree-MRHC except it is allowed to have missing alleles. 2-locus-MRHC is MRHC on a two-locus pedigree without missing data. 2-locus-MRHC\* is defined the same as 2-locus-MRHC except it is allowed to have missing data. Tree-MRHC is MRHC on a pedigree without mating loops or missing data.*

In order to discuss the hardness and approximation of the variants, we are going to make use of some related problems or properties, such as the Min UnCut [5] (*i.e.* 2-Linear-Equations Mod 2 [4]), Min UnCut($k$) (the same definition as Min UnCut except that each variable occurs at most $k$ times), Min 2CNF Deletion [4, 5] problems, consistency and satisfiablility property. The Min UnCut and Min 2CNF Deletion problems are known to be NP-hard [5]. We will show that the Min UnCut($k$) problem is also NP-hard in this paper.

For any NP-hard minimization (or maximization) problem, if there is some polynomial time algorithm to give a solution with the objective value no more (or less, respectively) than $f(n) \cdot OPT$ (or $OPT/f(n)$, respectively), where $f(n)$ can be any function of the input size $n$, the problem can be approximated within ratio $f(n)$; otherwise, the problem cannot be approximated.

## 1.3   Previous Complexity Results on MRHC

Qian and Bechmann proposed a ruled-based algorithm to reconstruct haplotype configurations based on six rules [16]. Their algorithm is a heuristic without theoretical

analysis. Li and Jiang first proved that MRHC on two-locus pedigree is NP-hard [1]. Doi, Li and Jiang further proved that MRHC on tree pedigrees is also NP-hard in the general case [2], even though MRHC can be solved by dynamic programming algorithms when the number of members or loci in the input pedigree is bounded by a constant. However, the NP-hardness proof requires pedigrees containing individuals with an unbounded number of mates or children. It was left as an open question if the proof can be improved to work for tree pedigrees where every individual has a bounded number of mates and children.

Consistency checking of the Mendelian law of inheritance (*i.e.* the Mendelian law checking problem) is closely related to the MRHC problem. The purpose of Mendelian law checking is to determine whether the given genotype data obey the classic Mendelian law of inheritance. Mendelian law checking usually needs to be done ahead of phasing haplotype configurations. Aceto *et al.* showed that the Mendelian law checking problem is NP-hard in general, although checking the consistency on pedigrees with bi-allelic data or with no mating loops [3] can be done in polynomial time.

In this paper, we consider a simple variant of MRHC, which involves pedigrees with members that has at most one mate and one child (*i.e.* binary-tree-MRHC). It is an open question if binary-tree MRHC is NP-hard. A polynomial-time algorithm for it, if exists, could be useful for solving the general-case MRHC problem. Another important question is whether a good approximation algorithm exists for MRHC. Here, in terms of computing the minimum-recombinant haplotype, the accuracy is sacrificed to improve the efficiency. Previously, there is no known polynomial-time approximation algorithm for MRHC with guaranteed ratio.

**Table 1.** The known hardness results of the Mendelian law checking and MRHC problems

| Pedigree / Problem | Loop? | Multi-allelic? | Unbounded number of loci? | Unbouned number of members? | Hardness |
|---|---|---|---|---|---|
| Mendelian law checking | Yes | Yes | | | NP-hard [3] |
| | | No | | | P [3] |
| | No | | | | P [3] |
| M R H C | Yes | No | No | Yes | NP-hard [1] |
| | No | No | No | Yes | P [2] |
| | No | No | Yes | No | P [2] |
| | No | No | Yes | Yes | NP-hard [2] |

## 1.4   Our Results

We will consider pedigrees with bi-allelic genotype data throughout this paper. First, we reduce ≠3SAT to the binary-tree-MRHC problem and show that this problem is NP-hard, which answers an open question in [2]. Second, we study the approximability of MRHC on pedigree data with the following restrictions: (*I*) 2-locus genotype data with missing alleles, (*II*) binary tree pedigrees with missing alleles, (*III*) 2-locus genotype data without missing alleles, and (*IV*) tree pedigrees without missing alleles. These four restricted cases of MRHC are NP-hard problems shown either in the literature [1,2] or in this paper. We demonstrate that for MRHC in the former two cases *I* and *II* cannot be approximated unless P = NP. We also prove that it is NP-hard to approximate problems *III* and *IV* within any constant ratio under the Unique Games

Conjecture [4]. Moreover, we show that problem *III* can be approximated with ratio $O(\sqrt{\log(n)})$ in polynomial time by reducing it to the Min 2CNF Deletion problem, Finally, we discuss the approximation of MRHC on pedigrees where each member has a bounded number of children and mates, mirroring pedigrees in real applications.

## 1.5  Organization of the Paper

The paper is organized as follows. We briefly give definitions of the MRHC problem and other closely related problems, introduce the related biological background in section 1. We prove binary-tree-MRHC is NP-hard and state the approximatability of MRHC on pedigrees with missing data in section 2. We show the approximation lower bound of MRHC on pedigrees without missing data and the approximation upper bound of 2-locus-MRHC in section 3. In section 4, we tentatively explore the approximation hardness of MRHC on the pedigrees where each member has a bounded number of mates and children. We organize our hardness results and conclude this paper with a few remarks in section 5. Due to space limitations, the proofs are omitted in the main text and are given in the full version [22].

## 2  Approximation of MRHC on Pedigrees with Missing Data

In this section, we prove the hardness of approximating MRHC on pedigree data with missing alleles. Two variants are considered.

**Lemma 1.** *If it is NP-hard to decide whether OPT(R)=0 for a minimization problem R, R cannot be approximated unless P=NP.*

### 2.1  Hardness and Approximation of Binary-Tree-MHRC(*)

**Theorem 2.** *Binary-tree-MRHC is NP-hard.*

**Theorem 3.** *It is NP-hard to decide whether OPT(binary-tree-MRHC*)=0.*

**Corollary 4.** *Binary-tree- MRHC* cannot be approximated unless P=NP.*

### 2.2  Approximation of 2-Loop-MHRC*

**Theorem 5.** *It is NP-hard to decide whether OPT(2-locus-MRHC*)=0*

**Corollary 6.** *2-locus-MRHC* cannot be approximated unless P=NP.*

## 3  Approximation of MRHC on Pedigrees Without Missing Data

In this section, we consider the approximability of the same variants of MRHC without missing data. In order to show the negative result, we need to use some gap-introducing reduction (or gap-preserving reduction) for MRHC. We will use the concept of *L-reduction* proposed by Papadimitriou and Yannakakis [18].

### 3.1  Approximation of Tree-MRHC

**Lemma 7.** *There is an L-reduction from Min UnCut to tree-MRHC that transforms a set of Boolean constraints $\varphi$ to a tree pedigree $\xi$ such that:*
  *(i) $OPT_{Min\ UnCut}(\varphi) = OPT_{tree\text{-}MRHC}(\xi)$, and*
  *(ii) Given a haplotype solution for $\xi$ with $k$ recombinants, we can construct a solution for $\varphi$ with at most $k$ unsatisfied clauses.*

**Theorem 8.** *It is NP-hard to approximate tree-MRHC within any constant ratio under the Unique Games Conjecture* [4].

### 3.2  Approximation of 2-Locus-MRHC

We will present a lower bound and an upper bound on the approximation ratio for the 2-locus-MRHC problem.

#### 3.2.1  Negative Result for Approximating 2-Locus-MRHC
**Lemma 9.** *There is a polynomial-time L-reduction from Min UnCut to 2-locus-MRHC that transforms a Boolean constraints set $\varphi$ to a pedigree $\xi$ such that*
  *(i) $OPT_{Min\ UnCut}(\varphi) = OPT_{2\text{-}locus\text{-}MRHC}(\xi)$, and*
  *(ii) Given any haplotype solution for $\xi$ with $k$ recombinants, we can find in polynomial time a truth assignment for $\varphi$ with at most $k$ unsatisfied constraints.*

**Theorem 10.** *It is NP-hard to approximate 2-locus-MRHC within any constant ratio under the Unique Games Conjecture* [4].

#### 3.2.2  Positive Result for Approximating 2-Locus-MRHC
We first would like to reduce an instance of 2-locus-MRHC so that each member of the pedigree can be described by one Boolean variable. Since only two loci are involved, there are three types of members in a pedigree: (*I*) both loci are homozygous, (*II*) one locus is homozygous, and (*III*) both loci are heterozygous. A type *I* (or *II*) member has a fixed haplotype grouping. A type *III* member has a variable haplotype grouping.

Agarwal and Charikar recently presented a randomized polynomial-time $O(\sqrt{\log(n)})$ approximation algorithm for the Min 2CNF Deletion problem [5], where $n$ is the number of variables in the input 2CNF constraints.

**Theorem 11.** *There is a randomized polynomial-time $O(\sqrt{\log(n)})$ approximation algorithm for 2-locus-MRHC, where n is the number of members in the input pedigree.*

Observe that the results in this section show that, in terms of approximability, the 2-locus-MRHC problem is easier than the Min 2CNF Deletion problem and harder than the Min UnCut problem. Also, Lemma 9 presents an alternative proof that 2-locus-MRHC is NP-hard, which is much easier to understand than the original proof in [1].

## 4  Approximation of MRHC$(k, j)$

The proof of Lemma 7 uses a pedigree that contains members with a variable number of children, although every member in the pedigree has only one mate. Can we get the

same hardness result for tree-MRHC if we bound the number of mates instead of the number of children? In addition, the pedigrees in the proofs of Theorem 5 and Lemma 9 contain members with a variable number of children or mates. Another question is whether MRHC on two-locus pedigrees with a bounded number of children and mates leads to the same hardness result. In this section, we discuss the approximation of MRHC on pedigrees with bounded number of children and mates. For the convenience of comparison, we state strengthened versions of the previous theorems in the order they appear in this paper. We use $u$ to present an integer variable.

First, we refine Theorem 5. The hardness result in this theorem holds for 2-locus-MRHC($u$,1), because some member might appear in every clause gadget and every member has at most one child in the proof of Theorem 5.

**Theorem 12.** *2-locus-MRHC\*(4,1) cannot be approximated unless P=NP.*

Next, let us look at Lemma 7. This lemma actually works for tree-MRHC(1,$u$). It is natural to consider tree-MRHC on pedigrees where members have a bounded number of children with each mate. In order to decrease the number of children and mates in the pedigree, we need a bounded version of Min UnCut like the one for Max 3SAT.

In fact, there is an L-reduction from Min UnCut to Min UnCut(15) that transforms a Boolean constraints set $\varphi$ to another Boolean constraints set $\psi$ such that

(*i*) $OPT_{UnCut}(\varphi) = OPT_{UnCut(15)}(\psi)$, and

(*ii*) Given any truth assignment for $\psi$ with $k$ unsatisfied constraints, we can find in polynomial time a truth assignment for $\varphi$ with at most $k$ unsatisfied constraints.

This L-reduction from Min UnCut to Min UnCut(15) can be constructed using the same idea as the L-reduction that transforms Max 3SAT to Max 3SAT(29) in [6] with just a few minor modifications. The details of this L-reduction are omitted here. Based on the property of this L-reduction, we know that it is NP-hard to approximate Min UnCut(15) within any constant ratio under the Unique Games Conjecture [4].

**Theorem 13.** *It is NP-hard to approximate tree-MRHC(u,1) within any constant ratio under the Unique Games Conjecture [4].*

Finally, we consider Lemma 9. The hardness result actually holds for 2-locus-MRHC($u$, $u$), because neither the number of mates nor the number of children for a member is bounded by any constant.

**Theorem 14.** *It is NP-hard to approximate 2-locus-MRHC(16,15) within any constant ratio under the Unique Games Conjecture [4].*

## 5   Discussion and Conclusion

The results presented in this paper are organized in Table 2. First, we showed that binary-tree-MRHC is NP-hard. Binary-tree-MRHC is a simplest variant of MRHC because one mate and one child are the minimum requirement to express the inheritance of human beings. Second, we showed some approximability results concerning the MRHC problem. With the presence of missing data, it is NP-hard to tell if an instance of 2-locus-MRHC\* and binary-tree-MRHC\* requires any recombinant.

This gives an interesting contrast to the results in [1] where the problem of finding a zero-recombinant haplotype solution for MRHC was shown to be solvable in polynomial time. This result also implies that 2-locus-MRHC* and binary-tree-MRHC* is not approximable in polynomial time. Without the presence of missing data, 2-locus-MRHC can be approximated with the ratio $O(\sqrt{\log(n)})$. In addition it is NP-hard to approximate 2-locus-MRHC and tree-MRHC within any constant ratio under the Unique Games Conjecture [4]. Our final results concern the inapproximability of MRHC on pedigrees where each member has a bounded number of mates and/or a bounded number of children with each mate.

**Table 2.** Our hardness and approximation results for MRHC with bi-alleles

| | Loop? | Miss-ing data? | Unbounded Number of loci? | Unbounded number of members? | Hardness | Lower bound of approx. ratio | Assumption | The lower bound holds for | Upper bound of approx. ratio |
|---|---|---|---|---|---|---|---|---|---|
| Binary-tree-MRHC | No | No | Yes | Yes | NP | | | | |
| 2-locus-MRHC* | Yes | Yes | No | Yes | | Any f(n) | P≠ NP | 2-locus-MRHC* (4,1) | |
| Binary-tree-MRHC* | No | Yes | Yes | Yes | | Any f(n) | P≠ NP | Binary-tree-MRHC* | |
| 2-locus-MRHC | Yes | No | No | Yes | | Any constant | P≠ NP, the Unique Games Conjecture | 2-locus-MRHC (16,15) | $O(\sqrt{\log(n)})$ |
| Tree-MRHC | No | No | Yes | Yes | | Any constant | P≠ NP, the Unique Games Conjecture | Tree-MRHC(1,u) Tree-MRHC(u,1) | |

# Acknowledgement

# References

1. J. Li and T. Jiang: Efficient rule-based haplotyping algorithm for pedigree data. Proc. of the 7th Annual Conference on Research in Computational Molecular Biology (RECOMB'03), pages 197-206, 2003.
2. K. Doi, J. Li and T. Jiang: Minimum recombinant haplotype configuration on tree pedigrees. Proc. of the 3rd Annual Workshop on Algorithms in Bioinformatics (WABI'03), pages 339-353, 2003.
3. L. Aceto *et al.*: The complexity of checking consistency of pedigree information and related problems. J. Comp. Sci. Tech., 19(1): 42–59, 2004.
4. S. Khot: On the power of 2-Prover 1-Round Games. Proc. of the 34th ACM Symposium on Theory of Computing (STOC'02), pages 767-775, 2002.
5. A. Agarwal, M. Charikar: $O(\sqrt{\log(n)})$ approximation algorithms for min UnCut, min 2CNF deletion, and directed cut problems. Proc. STOC'05, pages 573-581, 2005.
6. G. Ausiello *et al.*: Complexity and approximation: combinatorial optimization problems and their approximability properties, pages 276-279, Springer, 1999.
7. L. Jorde: Where we are hot, they are not. Science, Volume 308, pages 60-62, 2005.

8. L. Li, J.H. Kim, and M. S. Waterman: Haplotype reconstruction from SNP alignment. Proc. RECOMB'03, pages 207-216, 2003.

9. R. Lippert, *et al.*: Algorithmic strategies for the single nucleotide polymorphism haplotype assembly problem. Briefings in Bioinformatics 3(1): 23-31, 2002.

10. E. Eskin E. Halperin, and R.M. Karp: Large scale reconstruction of haplotypes from genotype data. Proc. RECOMB'03, pages 104-113, 2003.

11. L. Excoffier and M. Slatkin: Maximum–likelihood estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol., 12: 921-927, 1995.

12. H. Seltman, K. Roeder, and B. Devlin: Transmission/disequilibrium test meets measured haplotype analysis: family-based association analysis guided by evolution of haplotypes. Am. J . Hum. Genet., 68(5): 1250-1263, 2001.

13. S. Zhang *et al.*: Transmission/ disequilibrium test based on haplotype sharing for tightly linked markers. Am. J. Hum. Genet., 73(3): 556-579, 2003.

14. J. Li and T. Jiang: An exact solution for finding minimum recombinant haplotype configurations on pedigrees with missing data by integer linear programming. Proc. RECOMB'04, pages 20-29, 2004.

15. J.R. O'Connell: Zero-recombinant haplotyping: applications to fine mapping using SNPs. Genet. Epidemiol., 19 Suppl. 1: S64-70, 2000.

16. D. Qian and L. Beckmann: Minimum-recombinant haplotyping in pedigrees. Am. J. Hum. Genet., 70(6): 1434-1445, 2002.

17. The International HapMap Consortium: The International HapMap Project. Nature, Volume 426, pages 789-796, December 2003.

18. C.H. Papadimitriou and M. Yannakakis: Optimization, Approximation, and Complexity Classes. J. Comp. System Sci., pages 425-440, 1991.

19. M. Stephens, N. J. Smith, and P. Donnelly: A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet., 68(4):978–989, 2001.

20. T.J. Schaefer: The complexity of satisfiability problems. Proc. of the 10th STOC, pages 216-226, 1978.

21. J. Li and T. Jiang: Computing the Minimum Recombinant Haplotype Configuration from incomplete genotype data on a pedigree by integer linear programming. Proc. RECOMB'04, pages 20-29, 2004.

22. Available at http://www.cs.ucr.edu/~lliu/App_MRHC.pdf.