

# Consensus Methods for Reconstruction of Sibling Relationships from Genetic Data

Saad I. Sheikh and Tanya Y. Berger-Wolf and Ashfaq A. Khokhar and Bhaskar DasGupta

{ssheikh, tanyabw, ashfaq, dasgupta}@cs.uic.edu

Department of Computer Science

University of Illinois Chicago

851 S. Morgan St (M/C 152),

Room 1120 SEO,

Chicago, IL 60607.

## Abstract

Reconstruction Sibling Relationships is an important problem with applications in conservational biology. A number of methods have been proposed for this problem, and biologists today find it challenging to consolidate different reconstructions into one solution. Towards this end, consensus based methodology has been proposed recently to combine different results. In this paper we study the use of different consensus techniques, including strict consensus, voting consensus, majority consensus, to realize a single solution. We also discuss the relative merits of different consensus techniques and extend their use to data sets with genotyping errors. We explain the implications of Mirkin's impossibility results in context of the siblings reconstruction problem.

## Introduction

Siblings Reconstruction is an important problem in Population Biology, with applications in a number of areas. In recent years there has been a boost in the genotyping methods and the cost has reduced considerably. This opens the possibilities of investigating many fundamental biological phenomena, including behavior, mating systems, heritabilities of adaptive traits, kin selection, and dispersal patterns. There are a number of methods (Almudevar 2003; Wang 2004; Beyer and May 2003; Smith, Herbinger, and Merry 2001; C.Thomas and G.Hill 2002; Berger-Wolf et al. 2005; 2007) for Sibship Reconstruction. With the number of methods growing it is becoming harder for the biologists to come up with a unified view of the population. There seem to be no existing methods (Blouin 2003) that are able to combine different results into one representative solution. We recently proposed a distance-based solution (Sheikh et al. ). In this paper we present different approaches to consensus for reconstructing sibling relationships and discuss their effectiveness for combining different solutions. We also discuss the implications of Mirkin's impossibility results

in the context of the siblings reconstruction problem. We conclude with how these consensus methods can be used to reconstruct siblings relationships in presence of genotyping errors.

## Consensus Methods

The idea behind consensus methods is to combine different solutions to the same problem into one solution, *i.e.*, group decision making. Group decision making is as old and as ubiquitous as human societies. The formal theory of voting and social choice dates back to the eighteenth century members of the French Academy of Sciences, Marquis de Condorcet (de Caritat marquis de Condorcet 1785) and de Borda (de Borda 1784). The modern developments in the field date back to Kenneth J. Arrow's seminal doctoral thesis (Arrow 1963) in 1951.

In the past fifteen years the mathematical and computational techniques developed in the context of group choice and consensus decisions have started to be applied to biological problems, mainly in systematics, taxonomy, and phylogenetics (Bininda-Emonds 2001). Many computational approaches to biological problems result in multiple answers either from the same or different methods. In absence of a verifiable true answer, as is common in biological problems, one may apply a consensus method to combine these solutions into one representative answer. The mathematical and computational field of BIOCONSENSUS explores different options for combining biological data (*e.g.* phylogenetic trees), establishes a formal framework to develop new consensus methods that stress various aspects of the data, compare their relative merits, and evaluate their practical performance. Our solution is based on using such methods to tolerate genotyping errors.

## Definitions

**Siblings:** a group of individuals that share at least one parent. When they share both parents they are called *full siblings*, and when they share exactly one of the

parents they are called *half siblings*. In this paper when we refer to *siblings* we mean *full siblings*.

**Locus:** the location of a gene on a chromosome.

**Allele:** one of the different versions of the same gene found at the same locus but in homologous chromosomes or in different individuals.

**Genetic marker:** a set of alleles of genes used as experimental probes to keep track of an individual.

**Diploid individual** is one having two alleles (not necessarily different) for each locus.

**Homozygous individual** is one having two identical alleles at a particular genetic locus.

**Heterozygous individual** is one having two different alleles at a particular genetic locus.

**Allele frequency:** the fraction of all the alleles of a gene in a population that are of one type.

**Genotype:** the actual alleles present in an individual; the genetic makeup of an organism.

## Problem Statement

We now restate the sibling reconstruction problem as defined in (Berger-Wolf et al. 2007). Given a genetic (microsatellite) sample from a population of  $n$  diploid individuals of the same generation,  $U$ , the goal is to reconstruct the full sibling groups (groups of individuals with the same parents). We assume no knowledge of parental information.

Formally, we are given a set  $U$  of  $n$  individual microsatellite samples from  $l$  genetic loci

$$U = \{X_1, \dots, X_n\}, \text{ where } X_i = (\langle a_{i1}, b_{i1} \rangle, \dots, \langle a_{il}, b_{il} \rangle)$$

and  $a_{ij}$  and  $b_{ij}$  are the two alleles of the individual  $i$  at locus  $j$ .

The goal is to find a partition of individuals  $P_1, \dots, P_m$  such that

$$\forall 1 \leq k \leq m, \forall X_u, X_v \in P_k : \text{Parents}(X_u) = \text{Parents}(X_v)$$

Notice, here that we have not defined the function  $\text{Parents}(x)$ . This is biological objective. We will discuss computational approaches to achieve a good estimate of the biological sibling relationship.

## 2-Allele and 4-Allele Properties

Mendelian genetics lay down a very simple rule for inheritance of diploid organisms: *a child inherits one allele from each of its parents for each gene*. This introduces two overlapping necessary (but not sufficient) constraints on full siblings groups: 4-allele property and 2-allele property (Berger-Wolf et al. 2005).

**4-Allele Property:** The total number of distinct alleles occurring at any locus in a sibling group may not exceed 4.

Formally, a set  $S \subseteq U$  has the 4-allele property if

$$\forall 1 \leq j \leq l : \left| \bigcup_{i \in S} \{a_{ij}, b_{ij}\} \right| \leq 4. \quad (1)$$

The 4-allele property is effective for identifying sibling groups where the data are mostly heterozygous. Generally, as in Table 1, a set consisting of any two individuals satisfies the 4-allele property. The set of individuals 1, 3 and 4 from Table 1 satisfies the 4-allele property. However, the set of individuals 2, 3 and 5 fails to satisfy it as the alleles occurring at the first locus are  $\{12, 31, 56, 44, 51\}$ .

**2-Allele Property:** In every sibling group there exists a swapping of individual alleles within a locus such that the number of distinct alleles on each side at this locus does not exceed 2.

2-Allele property is clearly more strict than 4-allele property. Looking at the Table 1, our previous 4-allele set of individuals 1, 3 and 4 fails to satisfy the stricter 2-allele property as the alleles appearing on the left side at locus 1  $\{44, 31, 13\}$  are more than two. Moreover, there is no swapping of alleles that will bring down the number of alleles on each side to two: the 1st and 4th individuals with alleles 44/44 and 13/13 already fill the capacity.

Individual	Alleles ( $a/b$ ) at Locus 1	Locus 2
Radish 1	44/44	55/23
Radish 2	12/56	14/31
Radish 3	31/44	55/14
Radish 4	13/13	31/23
Radish 5	31/51	14/31

Table 1: An example of input data for the sibling reconstruction problem. The five individuals have been sampled at two genetic loci. Each allele is represented by a number. Same numbers represent the same alleles.

## Consensus Methods for Siblings Reconstruction

Recall that for a population of individuals  $U = \{X_1 \dots X_n\}$  the goal of a siblings reconstruction problem is to find a partition of the population into sibling groups  $S = \{P_1 \dots P_m\}$  where and all individuals are covered

$$\bigcup_{1 \leq j \leq m} P_j = U$$

with no overlap

$$\forall j, k : P_j \cap P_k = \emptyset$$

It is well-known that a partition defines an equivalence relationship. Two individuals are equivalent if they are in the same partition of the solution  $S$ .

$$X_i \equiv_S X_j \iff \exists P_k \in S \text{ s.t. } X_i \in P_k \wedge X_j \in P_k$$

We are now ready to give the definition of a consensus method.

**Definition 1** A consensus method for sibling groups is a computable function  $f$  that takes  $k$  solutions  $\mathcal{S} = \{S_1, \dots, S_k\}$  as input and computes one final solution.

$$f : S^* \rightarrow S$$

### Strict Consensus

**Definition 2** A strict consensus (McMorris, Meronik, and Neumann 1983)  $\mathcal{C}$  is a partitioning of sibling groups where two individuals are together only if they were in the same partition for all solutions:

$$\mathcal{C} = \{P_{\mathcal{C},1} \dots P_{\mathcal{C},m}\}$$

where:

$$X_j \equiv_{\mathcal{C}} X_k \iff \forall S_i \in \mathcal{S} \ X_j \equiv_{S_i} X_k$$

Note that the strict consensus defines a true equivalence relation and, thus, is a transitive function:

$$X_i \equiv_{\mathcal{C}} X_j \text{ and } X_j \equiv_{\mathcal{C}} X_k \Rightarrow X_i \equiv_{\mathcal{C}} X_k$$

Any individual that is not consistently placed into a partition in all solutions will be added as a singleton. Such a consensus solution is reliable for the individuals that have been placed together in a group, but there may be a lot of singleton sibling groups.

As we will see later, strict consensus is good as a baseline as it ensures Pareto optimality. However it results in too many singletons and scattered sibgroups, therefore has limited application on its own.

### Majority Consensus

**Definition 3** A majority consensus  $\mathcal{C}$  is a partitioning of sibling groups where two individuals are together only if they were in the same partition for all solutions.

Majority consensus may lead to violation of the transitive property of equivalence relationships. Violation of this property also means that there is no partitioning of individuals and consequently no siblings reconstruction. Therefore some refinement of the basic definition is need to produce a partitioning of individuals. Such a refinement may be useful in practice for combining solutions from different algorithms.

### Voting Consensus

One form of Majority Consensus would be to have the solutions vote on all pairs of individuals. If a majority of votes put two individuals together, then the sibgroups containing those individuals should be merged. While this does produce a partition of individuals, it does not account for the other individuals in the sibgroups being merged.

### Distance-based consensus

For a distance based consensus, we start with a strict consensus of the solutions and search for the *nearest good solution*. In order to search for such a solution we need quantitative measures to 1) assess quality of a solution,  $f_q$ , and 2) calculate the pairwise distance between solutions,  $f_d$ . Assume that we have the two functions  $f_q$  and  $f_d$ .

$$f_q : S \rightarrow \mathbf{R}$$

$$f_d : S \times S \rightarrow \mathbf{R}$$

Since we start with a strict consensus  $\mathcal{C}$  the partitions in the solution cannot be refined any further. Therefore to improve the solution, we use the operations of merging two sets. The following monotonic property must be obeyed by any improved solution  $\mathcal{C}'$ :

$$\forall X_i, X_j \in U \ X_i \equiv_{\mathcal{C}} X_j \implies X_i \equiv_{\mathcal{C}'} X_j. \quad (2)$$

Thus, given a solution  $\mathcal{C}$ , we look for an improved solution  $\mathcal{C}'$  that minimizes  $f_d(\mathcal{C}, \mathcal{C}')$  and maximizes  $f_q(\mathcal{C}')$ . To combine the two objectives we can formulate the following optimization problems:

1. Maximize  $f_q$  with an upper bound on  $f_d$
2. Minimize  $f_d$  with a lower bound on  $f_q$
3. Maximize/Minimize some (linear) combination of  $f_d$  and  $f_q$

We have shown all of these problems to be NP-Hard in general for arbitrary  $f_q$  and  $f_d$  (Sheikh et al. ).

**Theorem 1** Let  $\mathcal{C}$  be a collection of sibling groups and  $k \in \mathbf{R}$ . Let  $\mathcal{S}$  be the set of all solutions that are an improvement of  $\mathcal{C}$  and are obtainable from  $\mathcal{C}$  by merging sibling sets. The problem of finding an improved solution  $\mathcal{C}' \in \mathcal{S}$  such that

$$f_q(\mathcal{C}') = \max_{\substack{\mathcal{S} \in \mathcal{S} \\ f_d(\mathcal{C}, \mathcal{S}) \leq k}} f_q(\mathcal{S})$$

is NP-hard.

**Theorem 2** Let  $\mathcal{C}$  be a collection of sibling groups and  $k \in \mathbf{R}$ . Let  $\mathcal{S}$  be the set of all solutions that are an improvement of  $\mathcal{C}$  and are obtainable from  $\mathcal{C}$  by merging sibling sets. The problem of finding an improved solution  $\mathcal{C}' \in \mathcal{S}$  such that

$$f_d(\mathcal{C}, \mathcal{C}') = \min_{\substack{\mathcal{S} \in \mathcal{S} \\ f_q(\mathcal{S}) \geq k}} f_d(\mathcal{C}, \mathcal{S})$$

is NP-hard.

Lastly, if no exact combination of  $f_q$  and  $f_d$  is specified, objective 3 is unattainable as well.

**Theorem 3** Let  $\mathcal{C}$  be a collection of sibling groups. Let  $\mathcal{S}$  be the set of all solutions that are an improvement of  $\mathcal{C}$  and are obtainable from  $\mathcal{C}$  by merging sibling sets and let  $g(f_q, f_d)$  be a (linear) combination of the functions  $f_q$  and  $f_d$ . The problem of finding an improved solution  $\mathcal{C}' \in \mathcal{S}$  such that

$$g(f_d(\mathcal{C}, \mathcal{C}'), f_q(\mathcal{C}')) = \underset{\mathcal{S} \in \mathcal{S}}{\text{OPT}}\{g(f_d(\mathcal{C}, S), f_q(S))\}$$

is NP-hard.

Distance based consensus seems to be an ideal ground ensuring Pareto optimality, and parsimony can be enforced using the quality measures. However, the problem is computationally intractable, therefore we propose greedy heuristics.

### Pairwise Greedy Consensus

Pairwise Greedy Consensus is a heuristic for distance-based consensus, it works iteratively by merging the *closest* pair of sibling groups. Some *editing* costs associated with different types of genotyping errors are needed and we assume it is available to us in a table *costs*. We define two functions to calculate the distance  $f_d$ : one calculates the alleles that needs to be removed to add an individual to a group; and the other calculates the shared alleles and allele pairs if no changes are needed. The former is used when an individual cannot be assigned without violating 2-allele property. The latter uses the same costs for calculating the “new” alleles/allele pairs brought by a new individual in a sibgroup, and gives a negative value where a higher value means more restrictions introduced to the sibgroup. Also, we assume that we know what is the maximum editing cost (*maxedit*) we can allow for an individual to be assigned to a sibgroup. If it costs more than that for all sibgroups, we would rather create a new sibgroup.

This is done by calculating the  $f_d$  distance for all pairs of sibling groups at every iteration. The pair that gives the smallest distance is merged and then all the pairs are compared again. This continues until no group of individuals can be merged without exceeding *maximum editing cost per individual* for some individual. Both of these costs are input parameters. The quality

function  $f_q$  is based on the parsimony assumption: the number of sibgroups. The objective is to maximize:

$$f_q = |U| - |\mathcal{C}|$$

This method can perform well depending upon the exact distance function, but it fails to maintain a control on how groups are evolving over time and may allow too much distance overall in both the solution and the groups.

### Sibgroup Greedy Consensus

Sibgroup Greedy Consensus is also a heuristics for distance-based consensus, with a different use of the

distance function. It is similar to the Pairwise Greedy Consensus, works by iteratively merging closest groups (see (Sheikh et al. ) for details and performance analysis). Instead of just making a purely local decision, a total merge cost is maintained for every sibgroup, and is added to  $f_d$  when comparing with another sibgroup. The pair that gives the least *total* merging cost is merged, and the total cost for the merged group is updated. This continues until the minimum distance is greater than either the *maximum editing cost per sibling group* or the average per individual distance exceeds *maximum average editing cost per sibling group*. Both of these costs are input parameters.

Even though this method is *greedy*, it maintains a control on both inter-sibgroup and intra-sibgroup distance.

### Impossibility Results

We now discuss the known impossibility results for equivalence relationships, as they automatically apply to siblings reconstruction. We first present the axioms for rules on equivalence relations.

All of these are defined on consensus rules of the form  $C : S^k \rightarrow S$  on the set of equivalence relations  $\mathcal{S} = \{S_1, \dots, S_k\}$  over elements of  $U$ .

**Definition 4 Independence:**  $\forall X \subseteq U \wedge \forall P, P' \in S^k : [P|_X = P'|_X] \implies [C(P)|_X = C(P')|_X]$ .

The independence property implies that for any subset  $X$  of individuals, for any pair of input profiles  $P, P'$ . If the restricted input profiles are same when restricted to  $X$ , then the restriction of the consensus must also produce the same equivalence relations when restricted to  $X$ . This is a very desirable property for siblings reconstruction as sibling relationship between a set of individuals should not change with the context in which they are observed.

**Definition 5 Pareto Optimality:**  $\forall x, y \in U \wedge \forall P = (S_1, \dots, S_k) \in S^k : [\forall i \in K : xE_i y] \implies xC(P)y$ .

In context of siblings reconstruction, Pareto optimality means that if all solutions pair two individuals together, they **MUST** be together. In other words, the solution is obtained by merging groups in the strict consensus.

**Definition 6 Oligarchy:** A  $V \subseteq K$  exists such that  $\forall P = (S_1 \dots S_k) \in S^k : C(P) = \bigcap_{i \in V} S_i$ .

Oligarchy means that only a subset of solutions determines the partitioning, not all input solutions may be necessary. For siblings reconstruction, it means that some “chosen” input solutions always determine the output. In our formulation for genotyping errors, there cannot be an oligarchy as any two input solutions are based on only slightly different data. Generally, an oligarchy is not acceptable as it means that some of the

input solutions may not have any say in the final solution.

**Definition 7** *Symmetry*:  $\forall P \in S^k \wedge \forall$  permutations  $\sigma$  of  $K$ :  $P = (S_1, \dots, S_k) \implies [C(P) = C(S_{\sigma(1)}, \dots, S_{\sigma(k)})]$ .

Symmetry implies that it does not matter how the solutions are obtained, the output solution depends only on the inputs and not their order or source.

The following impossibility theorem was presented by Mirkin (Mirkin 1975).

**Theorem 4** *Consensus rule*  $C : S^k \rightarrow S$  is independent and Pareto optimal if and only if it is oligarchic.

Which easily yields to:

**Corollary 1** *Consensus rule*  $C : S^k \rightarrow S$  is rule by unanimity if and only if it is independent, Pareto optimal and symmetric.

Let's consider what this result means for reconstruction of sibling relationships. If a consensus rule can guarantee independence regarding subsets of individuals and also guarantees that if all input solutions identify a set of individuals as siblings, then there is an oligarchy of solutions determining the output. Both independence and Pareto optimality are extremely important, but if they apply then there is a dictatorial subgroup of solutions which decide which individuals can be siblings. The corollary shows that if we desire for all the inputs to be treated equally, then they must always agree.

### Consensus based approach for error-tolerant siblings reconstruction

With the exception of COLONY (Wang 2004), none of the existing kinship reconstruction methods is designed to tolerate genotyping errors or mutation. Yet, both errors and mutation cannot be avoided in practice and identifying these errors without any prior kinship information is a challenging task. We now describe our approach (Sheikh et al. ) to reconstructing sibling relationships in presence of genotyping errors using consensus. Consider an individual  $X_i$  which has some genotyping error(s). Any error that is affecting siblings reconstruction must be preventing  $X_i$ 's sibling relationship with at least one other individual  $X_j$ , who in reality is a sibling. It is possible that there is more than one error in an individual's genotype, yet it is unlikely that all errors will bias the solution in the same direction.

Thus, we can discard one locus at a time, considering it to be erroneous, and obtain a sibling reconstruction solution based on the remaining loci. If all such solutions put the individuals  $X_i$  and  $X_j$  in the same sibling group (*i.e.*, there is a consensus among those solutions), we consider them to be siblings. The bulk of our error-tolerant approach design is concerned with pairs of individuals that do not consistently end up in the same

sibling group during this process, that is, there is no consensus about their sibling relationship.

We now discuss how the approaches defined above perform for this input.

### Majority and Voting Consensus

Such a consensus is highly prone to errors when used with our input solutions which are based on dropping one locus at a time. Errors will not be out-voted as long as we are using subsets where each locus is present in a majority of subsets.

**Theorem 5** *Majority Consensus for sibship reconstruction or any partitioning problem using "drop-one-locus/column" approach will always prefer in favor of the error.*

*Proof.* Let's consider the population of individuals as  $A$  a  $n \times k$  matrix. When a locus is dropped,  $t^{th}$  column vector from this matrix is dropped and the remaining  $n \times (k - 1)$  matrix  $A_t$  is used to compute a sibship reconstruction. Consider an error at row  $i$ , column  $j$ . Decisions made on  $A_1, \dots, A_{j-1}, A_{j+1}, \dots, A_k$  are based on data with error. Therefore, any majority rule will be in favor of the error with overwhelming majority.  $\square$

Even though such a technique may be useful in general, but not effective to handle errors in our framework.

Voting consensus is not suitable for the "drop-one-locus" approach as shown in the above theorem.

### Distance based consensus

Distance based consensus is well suited for this approach as a strict consensus should enforce an erroneous individual to be classified differently in at least one solution. Both pair-wise greedy and Sibgroup greedy consensus make intuitive sense for identifying and correcting errors. Pair-wise greedy consensus can allow too many genotyping errors in a sibgroup, leading to large sibgroups of individuals that share alleles but are otherwise unrelated. Sibgroup Greedy consensus maintains a control on errors at all levels and thus only allow errors that are allowed by the relative costs.

### Results

We tested these approaches on some real datasets using the "drop-one-locus" approach. Pair-wise greedy algorithm performs reasonably, better than both voting and strict consensus. Sibgroup greedy algorithm performs considerably better than all the other approaches. In fact, it outperforms all the known sibling reconstruction methods for few loci and high allele frequencies (Sheikh et al. ). We show the performance of three main approaches on a real dataset in Table 2. The dataset of tiger shrimp *Penaeus monodon* dataset (Jerry et al. 2006) consists of 59 individuals from 13 families with 7 loci. There are 16 missing alleles. The parentage is known and was used to identify errors.

Sibgroup Greedy	Pairwise Greedy	Voting
91.52	89.8	88.1

Table 2: Solution accuracy of consensus algorithms on Shrimp data.

## Conclusion

We have formulated a consensus-based approach for error-tolerant reconstruction of sibling relationships from genetic data. We have formulated and investigated various consensus based approaches. Strict Consensus ensures Pareto optimality but produces too many singleton groups. Majority Consensus may not produce a partition, and we have shown that it will not work with our error-tolerant approach. A distance-based consensus achieves the desired balance between Pareto optimality and parsimony, however, it is computationally intractable. Therefore, we propose greedy heuristics to approximate it.

We have also shown that it is not possible to have a fair consensus method that is both independent and Pareto optimal. This result is not unusual in social choice theory and we have shown that it holds in the domain of siblings reconstruction.

In future we intend to design an approximation algorithm with provable performance guarantees for distance based consensus methods for siblings reconstruction. Currently there are no consensus methods for hierarchical kinship analysis, we also intend to address this issue.

## References

- Almudevar, A. 2003. A simulated annealing algorithm for maximum likelihood pedigree reconstruction. *Theoretical Population Biology* 63.
- Arrow, K. J. 1963. *Social Choice and Individual Values*. John Wiley, New York, second edition.
- Berger-Wolf, T. Y.; DasGupta, B.; Chaovalitwongse, W.; and Ashley, M. V. 2005. Combinatorial reconstruction of sibling relationships. In *Proceedings of the 6th International Symposium on Computational Biology and Genome Informatics (CBGI 05)*, 1252–1255.
- Berger-Wolf, T. Y.; Sheikh, S. I.; Dasgupta, B.; Caballero, M. V. A. I. C.; Chaovalitwongse, W.; and Lahari, S. P. 2007. Reconstructing sibling relationships in wild populations. *Bioinformatics* 23(13).
- Beyer, J., and May, B. 2003. A graph-theoretic approach to the partition of individuals into full-sib families. *Molecular Ecology* 12:2243–2250.
- Bininda-Emonds, O. R. P. 2001. MRP supertree construction in the consensus setting. In Janowitz, M.; Lapointe, F.; McMorris, F.; Mirkin, B.; and Roberts, F., eds., *Bioconsensus*, DIMACS Series in Discrete Math-

ematics and Theoretical Computer Science. DIMACS-AMS.

Blouin, M. S. 2003. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *TRENDS in Ecology and Evolution* 18(10):503–511.

C.Thomas, S., and G.Hill, W. 2002. Sibship reconstruction in hierarchical population structures using markov chain monte carlo techniques. *Genet. Res., Camb.* 79:227–234.

de Borda, J.-C. 1784. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*.

de Caritat marquis de Condorcet, M. J. A. N. 1785. Essay on the application of analysis to the probability of majority decisions.

Jerry, D. R.; Evans, B. S.; Kenway, M.; and Wilson, K. 2006. Development of a microsatellite dna parentage marker suite for black tiger shrimp *penaeus monodon*. *Aquaculture* 542–547.

McMorris, F. R.; Meronik, D. B.; and Neumann, D. A. 1983. A view of some consensus methods for trees. In Felsenstein, J., ed., *Numerical Taxonomy*. Springer-Verlag. 122–125.

Mirkin, B. G. 1975. On the problem of reconciling partitions. In *Quantitative Sociology: Internations Perspectives on Mathematical and Statistical Modeling*. Academic Press, New York. 441–449.

Sheikh, S. I.; Berger-Wolf, T. T.; Ashley, M. V.; Caballero, I. C.; Chaovalitwongse, W.; and DasGupta, B. Error-tolerant sibship reconstruction. In *Submitted*.

Smith, B. R.; Herbinger, C. M.; and Merry, H. R. 2001. Accurate partition of individuals into full-sib families from genetic data without parental information. *Genetics* 158(3):1329–1338.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. *Genetics* 166:1968–1979.