JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 10, Numbers 5, 2003
© Mary Ann Liebert, Inc.
Pp. 669–676

On the Impossibility of Reconstructing Ancestral Data and Phylogenies

ELCHANAN MOSSEL

ABSTRACT

We prove that it is impossible to reconstruct ancestral data at the root of "deep" phylogenetic trees with high mutation rates. Moreover, we prove that it is impossible to reconstruct the topology of "deep" trees with high mutation rates from a number of characters smaller than a low-degree polynomial in the number of leaves. Our impossibility results hold for all reconstruction methods. The proofs apply tools from information theory and percolation theory.

Key words: phylogeny, phase transitions, trees, ancestral data.

1. INTRODUCTION

Trees are a useful tool in representing relationships between species. The reconstruction of these trees, named *phylogenies*, from genetic data is an important task of computational biology. A related problem is that of reconstructing *ancestral* data from genetic data at the leaves of a given tree with known topology.

In this paper, we discuss both problems under the standard assumption that mutations evolve independently and identically according to a Markov model. We focus on the case where the mutation rates along edges are high.

In the informal discussion below, it is helpful to think of "high mutation" as a setting where the number of species at time t is at most e^{qt}, where q is a positive number defined in terms of the mutation rate matrix Q. See bounds 1.1 and 1.2 below for exact formulations.

We derive a general bound on the mutual information between the data present at the leaves of a character and the ancestral data for this character. This bound in turn implies the impossibility of reconstructing ancestral data for high mutation rates.

Later, using the bound above, we derive a bound on the number of characters needed in order to reconstruct phylogenies.

When the mutation rates are high, the second bound is polynomial in the number of species, thus improving the well-known bound (for all mutation rates) which is logarithmic in the number of species.

The bounds derived in this paper give a positive answer to a question raised by David Penny and Mike Steel in the annual New Zealand phylogenetics meeting of 2002 (thus winning the Penny ante). The bounds we prove imply that it may be hard to test the hypothesis of common ancestry of all life on earth. More

---

Department of Statistics, University of California, Berkeley, CA 94720-3860. Supported by a Miller Fellowship. Most of the research reported here was conducted while the author was a PostDoc in theory group, Microsoft Research.

generally, it may be hard to determine "deep divergences" in the tree of life (see Sober and Steel [2002] and Bound 1.2).

On first sight, it may seem surprising that qualitatively different bounds are obtained for high low mutation rates.

In fact, there is an extensive body of work in statistical physics (formulated in a different language) which indicates that the high low mutation rate phases behave quite differently.

For low mutation rates, the system is "ordered" and data at a node v is correlated with functions of data at nodes arbitrarily far from that node. In contrast, for high mutation rates, the system is "unordered," and the correlation between data at a node v and a function of data at nodes F far from v decays exponentially in the distance between v and F. See Bleher *et al.* (1995), Evans *et al.* (2000), Ioffe (1996), and Mossel (2001) for some related results in statistical physics.

Mossel (2003) exploits this relationship to obtain both upper bounds and lower bounds for the CFN (Cavender–Farris–Neyman) model (Cavender, 1978; Farris, 1973; Neyman, 1971). Subsequent to a previous paper by this author (Mossel, 2003) and the results reported here, in a joint paper with M. Steel (Mossel and Steel, 2003), this relationship is also exploited to derive a phase transition for random cluster models on phylogenetic trees.

In this note, we omit a background survey of phylogeny. We refer the reader to Cavender (1978), Farris (1973), Neyman (1971), Steel (1994), Erdös *et al.* (1999), and Swofford *et al.* (1996) for background.

1.1. *Basic definitions*

We denote the root of a tree T by ρ and direct all edges away from the root. We let $\mathcal{C}$ be the set of character states (such that $\mathcal{C} = \{A, C, G, T\}$ or $\mathcal{C} = \{20 \text{ amino acids}\}$).

We assume that sites evolve identically and independently. We will often refer to sites as "characters." To each directed edge e of T, we associate a Markov matrix $M(e)$ which encodes the mutations along edge e (see Erdös *et al.* [1999] or Steel [1994] for a more formal description of the model).

Many of the standard models in biology satisfy $M(e) = \exp(t(e)Q)$, where $Q = (q_{i,j})_{i\in\mathcal{C}, j\in\mathcal{C}}$ is a mutation rates matrix and $t(e)$ represents the length of the edge e. We will consider only such models here.

Two examples are the Cavender–Farris–Neyman model,

$$Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix},$$

and Tajima and Nei's "equal input" model

$$Q = \begin{pmatrix} -(a+b+c) & a & b & c \\ d & -(b+c+d) & b & c \\ d & a & -(a+c+d) & c \\ d & a & b & -(a+b+d) \end{pmatrix}.$$

See Swofford *et al.* (1996) for various mutation rates Q appearing in biology.

The mutation process is therefore defined by a mutation rate matrix Q, a rooted tree $T = (V, E, \rho)$, a collection of edge lengths $t : E \to \mathbb{R}_+$, and a distribution π of the value of the character at the root.

A character $\sigma : V \to \mathcal{C}$ is a labeling of the vertices of T by $\mathcal{C}$. We will write σ_v for the value of σ at the vertex $v \in V$. The distribution of σ_ρ is given by π. If u is v's parent, then the conditional distribution of σ_v given σ_u is given by the matrix $M(e) = \exp(t(e)Q)$, where $e = (u, v)$. Furthermore, the assumption is that mutations events across different edges are independent and that different characters are independent and identically distributed.

We denote the characters by $\sigma^1, \ldots, \sigma^k$. We denote the collection of leaves of the tree T by ∂T and the values of character σ^i at the leaves by σ^i_∂. Similarly, we denote by σ^i_ρ the value of the character σ^i at the root ρ.

For a rooted tree $T = (V, E, \rho)$, its *topology* is the unrooted tree $T = (V, E)$.

1.2. Information bound for a single character and reconstruction of ancestral data

It is helpful to measure “information” using the notions of *entropy*, *conditional entropy*, and *mutual information* (see Cover *et al.* [1991] for background). For a random variable X, the *entropy* operator is defined by

$$H(X) = -\sum_x \mathbf{P}[X = x] \log_2 \mathbf{P}[X = x].$$

The *conditional entropy* of X given Y is defined by $H(X|Y) = H(X, Y) - H(Y)$, and the *mutual information* between X and Y is defined by

$$I(X, Y) = H(X) + H(Y) - H(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

For a leaf v, we let $\mathtt{path}(v)$ be the set of edges on the path connecting v to the root ρ, and $t(v) = \sum_{e \in \mathtt{path}(v)} t(e)$.

Bound 1.1. *Let σ be a character evolving according to a Markov model on a tree T, where $M(e) = \exp(t(e)Q)$ for all e. Let*

$$q_j = \min\nolimits_{i \neq j} q_{i,j}, \quad q = \sum\nolimits_j q_j. \tag{1}$$

Let σ_ρ be the value of the character at the root of the tree, and σ_∂ be the values of the character at the leaves of the tree.

Then

$$H(\sigma_\rho | \sigma_\partial) \geq H(\sigma_\rho) \left(1 - \sum_{v \in \partial T} e^{-qt(v)} \right), \tag{2}$$

and therefore,

$$I(\sigma_\rho, \sigma_\partial) \leq H(\sigma_\rho) \sum_{v \in \partial T} e^{-qt(v)} \leq \log_2 |\mathcal{C}| \sum_{v \in \partial T} e^{-qt(v)}. \tag{3}$$

Moreover, letting π be the prior distribution of the root character, and

$$\Delta = \sup_f \mathbf{P}[f(\sigma_\partial) = \sigma_\rho] \tag{4}$$

be the probability of reconstructing the value of σ_ρ given σ_∂ (the sup *is taken over all functions and is obtained for $f =$ maximum likelihood), it holds that*

$$\Delta \leq \max_i \pi[\sigma_\rho = i] + \sum_{v \in \partial T} e^{-qt(v)}. \tag{5}$$

A similar bound to (3) for general models is proved by Mossel (2001) in the discrete time setting. The results of Mossel (2001) imply that for general mutation matrices Q, it is impossible to obtain bounds of the form (3) with $e^{-cqt(v)}$ unless $c \leq 1$. Therefore, our results are sharp for general mutation models. However, for the CFN model, a better bound is proved by Evans *et al.* (2000), where the term $e^{-qt(v)}$ is replaced by $e^{-2qt(v)}$.

Note that it is trivial to achieve $\Delta = \max_i \pi[\sigma_\rho = i]$ in (5), by applying the constant function $f(\sigma_\partial) \equiv j$, where j maximizes $\max_i \pi[\sigma_\rho = i]$ over i. Bound (5) measures how much better than trivial reconstruction can be achieved.

1.3. Bounds on reconstruction of phylogenetic trees

Bound 1.1 implies a bound on the number of characters needed for the reconstruction of a phylogenetic tree from a collection of characters.

We let $T(s)$ denote the topology of the tree T up to time s and $T^c(s)$ denote the topology of the tree from time s. In other words, $T(s)$ contains all divergences up to time s, while $T^c(s)$ contains all divergences from time s.

Bound 1.2. *Consider the problem of phylogenetic reconstruction for an i.i.d. model on a rooted tree on n leaves, where the prior distribution on trees is given by a measure μ. Suppose, furthermore, that the characters evolve as in Bound 1.1 with mutation matrix Q; $q_j = \min_{i \neq j} q_{i,j}$ and $q = \sum_j q_j$.*

For all s, given $T^c(s)$ and k independent characters at the n leaves of the tree, $(\sigma_\partial^j)_{j=1}^k$,

$$I(T(s); (\sigma_\partial^j)_{j=1}^k | T^c(s)) \le k \sum_{v \in \partial T} e^{-q(t(v)-s)}. \tag{6}$$

Let $\Delta^T(s)$ be the probability of reconstructing, given $T^c(s)$, the tree topology up to time s,

$$\Delta^T(s) = \sup_f \mathbf{P}[f(\sigma_\partial^j)_{j=1}^k) = \textit{ Tree topology at time } s], \tag{7}$$

(the sup *is taken over all functions, and is obtained for $f =$ maximum likelihood). Similarly, let Δ^T be the probability of reconstructing T from $(\sigma_\partial^j)_{j=1}^k$.*

Then

$$\Delta^T \le \Delta^T(s) \le \max_T \mu[T(s) = T] + k \sum_{v \in \partial T} e^{-q(t(v)-s)}. \tag{8}$$

Assume, furthermore, that $T(s)$ has a uniform distribution among all topologies on r leaves conditioned on $T^c(s)$. Then

$$\Delta^T \le \Delta^T(s) \le \max \left\{ \frac{e}{2^{H(T(s)|T^c(s))/2}}, \frac{2k \sum_{v \in \partial T} e^{-q(t(v)-s)}}{H(T(s)|T^c(s))} \right\} \tag{9}$$

$$= \max \left\{ \frac{e}{\sqrt{(2r-5)!!}}, \frac{2k \sum_{v \in \partial T} e^{-q(t(v)-s)}}{\log_2 (2r-5)!!} \right\}.$$

Note that Bound (8) is applicable when $k \sum_{v \in \partial T} e^{-q(t(v)-s)} << 1$, while Bound (9) requires uniformity and is applicable when $H(T(s)|T^c(s))$ is large compared to $k \sum_{v \in \partial T} e^{-q(t(v)-s)}$.

Remark 1.1. *In order to get a feeling for the asymptotic behavior of (8), fix s and assume that the tree has $n = e^{\beta t}$ leaves, all at time t. Here we take the asymptotics where $t \to \infty$ (and therefore $n \to \infty$), while s, q, and β are all constants.*

Then

$$\sum_{v \in \partial T} e^{-q(t(v)-s)} = \exp(sq)\exp(-t(q-\beta)).$$

Therefore, if $q > \beta$, then by (8) if we want to reconstruct the topology up to time s with high probability, i.e., $\Delta^T(s) \ge \max_T \mu[T(s) = T] + n^{-o(1)}$, then we need that

$$k \ge \exp(t(q - \beta - o(1))) = n^{q/\beta - 1 - o(1)}.$$

So the number of characters needed is polynomial in n.

2. PROOFS

Proof of Bound 1.1. The proof utilizes the method of coupling (see, e.g., Aldous and Fill [2003] for background on coupling for Markov chains) and arguments from the theory of percolation (see Mossel [2001], Peres [1997], and Athreya and Ney [1972] for background).

The mutation rate from state i to state j is given by $q_{i,j}$. Recalling (1), we may define the chain equivalently as follows. Given the current state i,

I. Jump to state j with rate q_j.
II. Jump to state j with rate $q_{i,j} - q_j$.

The point here is that Step I is performed independently of the state i.

For edge $e = (u, v)$, let $D(e)$ be the event that transition of type I occurs along the edge e. Note that the events $D(e)$ are independent for different edges and that $\mathbf{P}[D(e)^c] = \exp(-qt(e))$. Moreover, conditioned on $D(e)$, σ_v is independent of σ_ρ.

For a leaf v, let $D(v)$ be the event that transition of type I occurs along an edge $e \in \mathtt{path}(v)$. Then

$$\mathbf{P}[D(v)^c] = \prod_{e \in \mathtt{path}(v)} \mathbf{P}[D(e)^c] = \prod_{e \in \mathtt{path}(v)} e^{-qt(e)} = e^{-qt(v)}.$$

Note that conditioned on $D(v)$, σ_v is independent of σ_ρ. Finally, let D be the event that $D(v)$ holds for all leaves $v \in \partial T$. Then

$$\mathbf{P}[D^c] \le \sum_{v \in \partial T} \mathbf{P}[D(v)^c] = \sum_{v \in \partial T} e^{-qt(v)}. \tag{10}$$

Note that conditioned on D, σ_∂ and σ_ρ are independent.

We now use the following standard identities in information theory:

$$I(\sigma_\rho, \sigma_\partial) = H(\sigma_\rho) - H(\sigma_\rho | \sigma_\partial), \tag{11}$$

and

$$H(\sigma_\rho | \sigma_\partial) = \sum_{\sigma} \mathbf{P}[\sigma_\partial = \sigma] H(\sigma_\rho | \sigma_\partial = \sigma). \tag{12}$$

Moreover,

$$H(\sigma_\rho | \sigma_\partial = \sigma) \ge \mathbf{P}[D | \sigma_\partial = \sigma] H(\sigma_\rho | \sigma_\partial = \sigma, D) + \mathbf{P}[D^c | \sigma_\partial = \sigma] H(\sigma_\rho | \sigma_\partial = \sigma, D^c). \tag{13}$$

Recall that conditioned on D, σ_ρ and σ_∂ are independent. Therefore, $H(\sigma_\rho | \sigma_\partial = \sigma, D) = H(\sigma_\rho | D)$. Since σ_ρ and D are clearly independent, it follows that $H(\sigma_\rho | D) = H(\sigma_\rho)$.

It therefore follows from (13) that

$$H(\sigma_\rho | \sigma_\partial = \sigma) \ge H(\sigma_\rho) \mathbf{P}[D | \sigma_\partial = \sigma]. \tag{14}$$

Plugging (14) into (12) and using (10), we obtain that

$$\begin{aligned} H(\sigma_\rho | \sigma_\partial) &\ge H(\sigma_\rho) \sum_{\sigma} \mathbf{P}[\sigma_\partial = \sigma] \mathbf{P}[D | \sigma_\partial = \sigma] = H(\sigma_\rho) \mathbf{P}[D] \qquad (15) \\ &= H(\sigma_\rho)(1 - \mathbf{P}[D^c]) \ge H(\sigma_\rho) \left(1 - \sum_{v \in \partial T} e^{-qt(v)} \right), \end{aligned}$$

so (2) follows. Combining (11) and (2), we see that (3) follows (recall that $H(\sigma_\rho) \le \log_2 |\mathcal{C}|$).

To prove the bound on reconstruction (5), note that if we are not given σ_∂ (or any other information on σ_ρ), then the best reconstruction function f satisfies $f \equiv j$, where j maximized $\pi[\sigma_\rho = i]$ over all i, and this function has success probability $\max_i \pi[\sigma_\rho = i]$. Now let f be any reconstruction procedure and note that on the event D, σ_ρ is independent of σ_∂ and therefore

$$\begin{aligned}\mathbf{P}[f(\sigma_\partial) = \sigma_\rho] &\leq \mathbf{P}[D^c] + \mathbf{P}[D]\mathbf{P}[f(\sigma_\partial) = \sigma_\rho | D] \\ &\leq \mathbf{P}[D^c] + \mathbf{P}[D] \max_i \pi[\sigma_\rho = i] \leq \mathbf{P}[D^c] + \max_i \pi[\sigma_\rho = i],\end{aligned}$$

and we obtain (5) by (10) as needed. ■

Proof of Bound 1.2. We state below some basic properties of I and Fano's inequality which roughly says that if $I(X, Y)$ is small, then it is hard to reconstruct X given Y (see, e.g., Cover *et al.* [1991] for background).

Lemma 2.1. *Let X, Y, and Z be random variables such that X and Z are independent given Y, then*

$$I(X, Z) \leq \min\{I(X, Y), I(Y, Z)\} \quad \text{(“Data Processing Lemma”)}, \tag{16}$$

$$I((X, Y), Z) = I(Y, Z), \tag{17}$$

$$I((X, Z), Y) \leq I(X, Y) + I(Z, Y). \tag{18}$$

Lemma 2.2 (Fano's inequality). *Let X and Y be random variables s.t. X has values in a set A of size m, Y has values in a set B, and*

$$\Delta = \Delta(X, Y) = \sup_{f : B \to A} \mathbf{P}[f(Y) = X] \tag{19}$$

is the probability of reconstructing the value of X given Y (the sup *is taken over all functions). Then*

$$H(\Delta) + (1 - \Delta)\log_2(m - 1) \geq H(X|Y), \tag{20}$$

where $H(\Delta) = -\Delta \log_2 \Delta - (1 - \Delta)\log_2(1 - \Delta)$.

We start by proving Bound (6) for $T(s)$. All the probabilities, entropies, and mutual informations in the proof of (6) are conditioned on $T^c(s)$. Recall that $\sigma_\partial^i = (\sigma_v^i : v \in \partial T)$. Let $\sigma_s^i = (\sigma_v^i : d(v, \rho) = s)$ be the i'th character at time s. Clearly $T(s)$ and $(\sigma_\partial^i)_{i=1}^k$ are independent given $(\sigma_s^i)_{i=1}^k$.

In particular, by the data processing Lemma (16),

$$I\left(T(s), (\sigma_\partial^i)_{i=1}^k\right) \leq I\left((\sigma_s^i)_{i=1}^k, (\sigma_\partial^i)_{i=1}^k\right). \tag{21}$$

Since σ^i and $\sigma^{i'}$ are independent when $i \neq i'$,

$$I\left((\sigma_s^i)_{i=1}^k, (\sigma_\partial^i)_{i=1}^k\right) = \sum_{i=1}^k I(\sigma_s^i, \sigma_\partial^i). \tag{22}$$

For v at time s, let $\partial(v)$ denote all the vertices at ∂T that are descendants of v. Let L_s be the collection of species at time s. Note that $(\sigma_{\partial(v)}^i : v \in L_s)$ are conditionally independent given σ_s^i. Therefore, by (18),

$$I(\sigma_s^i, \sigma_\partial^i) \leq \sum_{v \in L_s} I(\sigma_s^i, \sigma_{\partial(v)}^i). \tag{23}$$

Finally, note that $(\sigma_w^i : w \in L_s, w \neq v)$ are independent of $\sigma_{\partial(v)}^i$ given σ_v^i, and therefore for all $v \in L_s$,

$$I(\sigma_s^i, \sigma_{\partial(v)}^i) = I(\sigma_v^i, \sigma_{\partial(v)}^i). \tag{24}$$

Combining (21) with (22), (23), and (24), we obtain

$$I\left(T(s), (\sigma_\partial^i)_{i=1}^k\right) \le \sum_{i=1}^k \sum_{v \in L_s} I(\sigma_v^i, \sigma_{\partial(v)}^i).$$

By Bound (1.1), $I(\sigma_v^i, \sigma_{\partial(v)}^i) \le \Gamma(v)$, where

$$\Gamma(v) = \sum_{w \in \partial(v)} e^{-q(t(w)-s)}.$$

Therefore,

$$I\left(T(s), (\sigma_\partial^i)_{i=1}^k\right) \le k\Gamma,$$

where

$$\Gamma = \sum_{v \in \partial T} e^{-q(t(v)-s)},$$

and we obtain (6) as needed.

We now prove (8) and (9). First, note that reconstruction probabilities increase given additional information. Therefore, Δ^T is bounded by the probability of reconstructing $T(s)$ given $T^c(S)$. In other words, $\Delta^T \le \Delta^T(s)$ for all s.

In order to obtain Bound (8), we repeat the proof of Bound 1.1. For character i, we say that the event D_i occurs if for all $v \in \partial T$, there exists a time $t \ge s$ at which transition of type I. (of Bound 1.1) occurs along $\mathtt{path}(v)$. By the proof of Bound 1.1, it follows that $\mathbf{P}[D_i^c] \le \Gamma$, and therefore $\mathbf{P}[(\cap_{i=1}^k D_i)^c] \le k\Gamma$. Conditioned on $\cap_{i=1}^k D_i$, $(\sigma_s^i)_{i=1}^k$ and $(\sigma_\partial^i)_{i=1}^k$ are independent, and therefore, $T(s)$ and $(\sigma_\partial^i)_{i=1}^k$ are independent. As in Bound 1.1, we conclude that

$$\Delta^T(s) \le \mathbf{P}[(\cap_{i=1}^k D_i)^c] + \mathbf{P}[(\cap_{i=1}^k D_i)] \max_T \mu[T(s) = T] \le k \sum_{v \in \partial T} e^{-q(t(v)-s)} + \max_T \mu[T(s) = T],$$

and (8) follows.

From (6) we conclude that

$$H\left(T(s) \mid (\sigma_\partial^i)_{i=1}^k, T^c(S)\right) = H(T(s)|T^c(S)) - I\left(T(s), (\sigma_\partial^i)_{i=1}^k | T^c(S)\right) \ge H(T(s)|T^c(S)) - k\Gamma.$$

From now on we assume that $T(s)$ has a uniform distribution given $T^c(s)$. Again, all probabilities, entropies, etc. are conditioned on $T^c(s)$.

By Lemma 2.2, we conclude that the probability $\Delta = \Delta\left(T(s), (\sigma_\partial^i)_{i=1}^k\right)$ of reconstructing $T(s)$ given $(\sigma_\partial^i)_{i=1}^k$ satisfies

$$H(\Delta) + (1-\Delta)H(T(s)) \ge H(T(s)) - k\Gamma. \tag{25}$$

The rest of the proof consists of calculations showing how to derive (9) from (25).

First, note that we can rewrite (25) as

$$H(\Delta) + k\Gamma \ge \Delta H(T(s)),$$

from which we conclude that

$$\Delta \le \max\left\{\frac{2H(\Delta)}{H(T(s))}, \frac{2\Gamma}{H(T(s))}\right\}. \tag{26}$$

Note that $-(1-x)\log(1-x) \le x$ for $x \in [0,1]$, and therefore, $H(\Delta) \le -\Delta\log_2\Delta + \Delta\log_2(e)$. Thus, if $\Delta \le 2H(\Delta)/H(T(s))$, then $0.5\Delta H(T(s)) \le -\Delta\log_2\Delta + \Delta\log_2(e)$, or $\Delta \le e/2^{H(T(s))/2}$. So by (26), we obtain

$$\Delta \le \max\left\{\frac{e}{2^{H(T(s))/2}}, \frac{2\Gamma}{H(T(s))}\right\}. \tag{27}$$

Since the number of binary trees on r labeled leaves is $(2r-5)!!$, we obtain (9) as needed. ■

ACKNOWLEDGMENTS

I wish to thank Mike Steel and David Penny for proposing the ante that led to these bounds. David Penny is partially responsible for a whiff of Glenmorangie which the may be detected by the reader. I thank Geoff Nichols and Yuval Peres for helpful discussions. I thank Lea Popovic, Roded Sharan, and the anonymous referee for helpful comments on a draft of this paper.

REFERENCES

Aldous D., and Fill, J. A. 2003. *Reversible Markov Chains and Random Walks on Graphs*. Book in preparation. Current version available at *www.berkeley.edu/users/aldous/book.html.*

Athreya, K.B., and Ney, P.E. 1972. *Branching Processes*, Springer-Verlag, Berlin.

Bleher, P.M., Ruiz, J., and Zagrebnov, V.A. 1995. On the purity of limiting Gibbs state for the Ising model on the Bethe lattice. *J. Stat. Phys.* 79, 473–482.

Cavender, J.A. 1978. Taxonomy with confidence. *Math. Biosci.* 40, 271–280.

Cover, T.M., and Thomas, J.A. 1991. *Elements of Information Theory*, John Wiley, New York.

Erdös, P.L., Steel, M.A., Székely, L.A., and Warnow, T.A. 1999. A few logs suffice to build (almost) all trees (Part 1). *Rand. Struct. Alg.* 14(2), 153–184.

Evans, W., Kenyon, C., Peres, Y., and Schulman, L.J. 2000. Broadcasting on trees and the Ising Model. *Ann. Appl. Probab.* 10, 2, 410–433.

Farris, J.S. 1973. A probability model for inferring evolutionary trees. *Syst. Zool.* 22, 250–256.

Grimmett, G. 1999. *Percolation*, 2nd ed., Springer-Verlag, Berlin.

Ioffe, D. 1996. A note on the extremality of the disordered state for the Ising model on the Bethe lattice. *Lett. Math. Phys.* 37, 137–143.

Mossel, E. 2001. Reconstruction on trees: Beating the second eigenvalue. *Ann. Appl. Probab.* 11, 285–300.

Mossel, E. 2003. Phase transitions in phylogeny. *Trans. AMS*. To appear.

Mossel, E., and Steel, M.A. 2003. A phase transition for a random cluster model on phylogenetic trees. Submitted to *Math. Biosci.*

Neyman, J. 1971. Molecular studies of evolution: A source of novel statistical problems, *in* S.S. Gupta, and J. Yackel, eds., *Statistical Desicion Theory and Related Topics* 1–27. **AU1**

Peres, Y. 1997. Probability on trees: An introductory climb. *Lectures on probability theory and statistics (Saint-Flour, 1997)*. Lecture Notes in Math., 1717, 193–280, Springer, Berlin.

Steel, M.A. 1994. Recovering a tree from the leaf colourations it generates under a Markov model. *App. Math. Lett.* 7 (2), 19–24.

Sober, E., and Steel, M.A. 2002. Testing the hypothesis of common ancestry. *J. Theor. Biol.* 218, 395–408.

Swofford, D.L., Olsen, G.J., Waddell, P.J., and Hillis, D.M. 1996. Phylogenetic inference, *in* D.M. Hilles, C. Moritz, and B.K. Marble, eds., *Molecular Systematics* (2nd ed.), 407–514, Sinauer, Sunderland, U.S.A.

Address correspondence to:
Elchanan Mossel
Department of Statistics
Evans Hall
University of California
Berkeley, CA 94720-3860

E-mail: mossel@stat.berkeley.edu

Pub:
Please supply recto runninghead. Verso okay as set?

AU1
Publisher and city?