On the Accuracy of Ancestral Sequence Reconstruction for Ultrametric Trees with Parsimony

Lina Herbst · Mareike Fischer

Abstract We examine a mathematical question concerning the reconstruction accuracy of the Fitch algorithm for reconstructing the ancestral sequence of the most recent common ancestor given a phylogenetic tree and sequence data for all taxa under consideration. In particular, for the symmetric four-state substitution model which is also known as Jukes–Cantor model, we answer affirmatively a conjecture of Li, Steel and Zhang which states that for any ultrametric phylogenetic tree and a symmetric model, the Fitch parsimony method using all terminal taxa is more accurate, or at least as accurate, for ancestral state reconstruction than using any particular terminal taxon or any particular pair of taxa. This conjecture had so far only been answered for two-state data by Fischer and Thatte. Here, we focus on answering the biologically more relevant case with four states, which corresponds to ancestral sequence reconstruction from DNA or RNA data.

Keywords Maximum parsimony · Ancestral sequence reconstruction · Reconstruction accuracy · Symmetric four-state model

1 Introduction

The reconstruction of ancestral sequences, e.g., DNA sequences of common ancestors of present-day species, is an important approach in understanding the evolution and
origin of these species (Li et al. 2008; Liberles 2007; Yang et al. 2011). There exist various methods to do such reconstructions, e.g., the Fitch algorithm (Tuffley and Steel 1997; Semple and Steel 2003; Fitch 1971), which is based on the maximum parsimony criterion. However, how reliable is such a reconstruction?

Several studies analyzed the reliability, the so-called reconstruction accuracy, of the Fitch algorithm for reconstructing ancestral sequence data of the most recent common ancestor given a phylogenetic tree and sequences for all taxa under consideration (Li et al. 2008; Fischer and Thatte 2009; Zhang et al. 2010). It seems intuitive that the root state is more likely to be conserved for taxa that are closer to the root, since over time more sequence changes can occur. Moreover, one might expect that the reconstruction accuracy is highest when all taxa are taken into account, which was also suggested by earlier simulation studies (Salisbury and Kim 2001). However, it can be shown that there are cases in which the reconstruction accuracy improves when only a subset of taxa is considered (Li et al. 2008; Fischer and Thatte 2009). In particular, the reconstruction accuracy can even improve when a taxon close to the root is ignored (Fischer and Thatte 2009). Despite these counterintuitive results, in 2008 Li et al. conjectured that for any rooted binary ultrametric phylogenetic tree (i.e., a tree in which all branches have the same distance to the root) and a simple model of evolution, the Fitch algorithm using all taxa for ancestral state reconstruction is at least as accurate as using a single taxon (Li et al. 2008). Note that ultrametric trees are also often referred to as clocklike trees or molecular clocks. So the conjecture by Li et al. means that under a molecular clock, the reconstruction accuracy is at least as good as the conservation probability of any taxon. Note that under a molecular clock all taxa have the same conversation probability, and that this conjecture provides a lower bound on the reconstruction accuracy for any rooted binary ultrametric phylogenetic tree under a simple model of evolution. Ignoring all data besides the data of one species displays the extreme case of throwing information away. Thus, showing that the conjecture holds is good news for maximum parsimony as a criterion for ancestral state reconstruction. Fischer and Thatte (2009) proved the conjecture for two-state characters, but it remained unclear if it also holds for four-state data like DNA or RNA. Thus, the aim of this paper is to consider this biologically relevant case with four states. In particular, we answer the conjecture affirmatively. Additionally, we also prove that the conjecture holds for three-state characters. Along the way, we also prove that the Fitch parsimony method applied to all taxa is always at least as good as applied to any pair of taxa if the underlying tree is clocklike. However, we also show that this does not improve the lower bound induced by single leaves.

2 Preliminaries

Before we can present our results, we first have to introduce some basic concepts. Recall that a rooted binary phylogenetic tree on the leaf set $X$ ($|X| = n \geq 2$) is a connected, acyclic graph in which the vertices of degree 1 are called leaves, and in which there is exactly one node $\rho$ of degree 2, which is referred to as root, and all other non-leaf nodes have degree 3. Moreover, in a rooted binary phylogenetic $X$-tree the leaves are bijectively labeled by the elements of $X$. Let each vertex of the tree
be assigned a state element of a finite state set $A$ with $|A| \geq 2$. In particular, we are interested in the biologically relevant case with four states, e.g., $A = \{\alpha, \beta, \gamma, \delta\}$, which corresponds for instance to DNA or RNA data.

The states evolve from $\rho$ by the well-known symmetric $r$-state model $N_r$ with alphabet $A = \{\alpha_1, \ldots, \alpha_r\}$ (Tuffley and Steel 1997). In this model, a state of $A$ is selected as the root state with probability $\frac{1}{|A|}$. Assume that $e = (u, v)$ is an edge of the tree, and node $u$ is closer to the root than $v$. Then, in this model, $p_e$ is the substitution probability on edge $e$: It is the probability that $v$ is in some state $\alpha$ under the condition that $u$ is in a distinct state, say, $\beta$. This is denoted by $P(v = \alpha | u = \beta)$. The model is supposed to be symmetric; thus $p_e = P(v = \alpha | u = \beta) = P(v = \beta | u = \alpha) = P(v = \gamma | u = \alpha) = P(v = \alpha | u = \gamma) = P(v = \alpha | u = \delta) = P(v = \delta | u = \alpha)$. Furthermore, we assume that $0 \leq p_e \leq \frac{1}{|A|}$, in particular for four states we have $0 \leq p_e \leq \frac{1}{4}$. The biologically relevant case with four states, namely the $N_4$-model, is also often referred to as Jukes–Cantor model (Jukes and Cantor 1969).

Similar as in Fischer and Thatte (2009) and Regner (2012), we consider ultrametric trees, often known as clocklike trees or molecular clocks by biologists. It means that the expected number of substitutions from the root to any leaf is the same (Semple and Steel 2003).

In this manuscript, we reconstruct ancestral states by the maximum parsimony criterion with the Fitch algorithm, which we briefly explain now. Assume that we have a rooted binary tree with leaf set $X$. To introduce the Fitch algorithm, we first consider the kind of data we will map onto the leaves of the tree. The data are given by a character on a leaf set $X$, which is a function $f : X \rightarrow A$. Thus, each leaf is assigned a character state. Note that as we consider $X = \{1, \ldots, n\}$, we often write $f = f(1) f(2) \ldots f(n)$ instead of listing $f(1), \ldots, f(n)$ explicitly. Moreover, real data usually are given in terms of an alignment, i.e., a sequence of characters. As we see in the following, the Fitch algorithm considers each character separately (i.e., one after the other). This means that for each site, its ancestral state is estimated before parsimony moves on to the next site (in particular, root state estimates of different sites are independent of each other). This is why we focus on single characters in this manuscript.

Given a character, the Fitch algorithm (Fitch 1971) assigns a set of states to all interior vertices by minimizing the number of changes. The algorithm is based on Fitch’s parsimony operation. Therefore, let $A$ be a non-empty finite alphabet and let $A, B \subseteq A$. Then, Fitch’s parsimony operation $*$ is defined by

$$A * B := \begin{cases} A \cap B, & \text{if } A \cap B \neq \emptyset, \\ A \cup B, & \text{otherwise.} \end{cases}$$

Using this operation, the Fitch algorithm works as follows. Consider all vertices $v$, whose two direct descendants have already been assigned a set, say $A$ and $B$. Then, $v$ is assigned $A * B$. This step is continued upwards along the tree until the root $\rho$ is assigned a set, which is denoted by $\text{MP}(f, T)$. An example can be seen in Fig. 1.

Note that what we call the Fitch algorithm is in fact only one phase of the algorithm, but it is the only part we require to estimate potential root states. For more details, we refer to (Fitch 1971).
Fig. 1 Example for the Fitch algorithm for a rooted binary tree and the character $f : \gamma \gamma \beta \delta \beta \alpha \delta \delta$. At first, each leaf is assigned the state specified by the character. Then, all other vertices whose direct descendants have already been assigned a set are assigned a set by applying the parsimony operation. This step is continued until the root is assigned a set; here $MP(f, T) = \{\delta\}$.

For a four-state character, there are $2^4 - 1 = 15$ possible sets for each interior vertex, since 16 is the cardinality of the power set of an alphabet with four elements minus one for the empty set, i.e.: $\{\alpha\}, \{\beta\}, \{\gamma\}, \{\delta\}, \{\alpha, \beta\}, \ldots, \{\alpha, \beta, \gamma, \delta\}$. We say that the Fitch algorithm unambiguously reconstructs the root state if $|MP(f, T)| = 1$. Otherwise, the root state is reconstructed ambiguously, i.e., the method cannot decide between different states and therefore $|MP(f, T)| > 1$.

As stated above, real data usually come in the form of an alignment, i.e., a sequence of characters, rather than in the form of an individual character. In this case, the Fitch algorithm would consider each character, i.e., each column (“site”) of the alignment, separately. This is why we focus on the case of a single character and its reconstruction accuracy.

3 The Accuracy of Ancestral Sequence Reconstruction with Four-state Characters

Similar to Li et al., we now define the reconstruction accuracy for all $|A| \geq 2$ (Li et al. 2008). Therefore, let $MP(f, T)$ denote the set of character states chosen by the Fitch algorithm as possible root states when applied to character $f$ on tree $T$. Let $R \subseteq A$, $\alpha \in R$ and $|R| \geq 1$. The probability that the root state $\alpha$ evolves on $T$ to a character $f$ for which the Fitch algorithm assigns $R$ as possible root state set is given by $\mathbb{P}(MP(f, T) = R | \rho = \alpha)$. The reconstruction accuracy is then defined by

$$RA(X) := \sum_{R : R \subseteq A \text{ and } \alpha \in R} \frac{1}{|R|} \cdot \mathbb{P}(MP(f, T) = R | \rho = \alpha).$$

This definition is motivated by the fact that when the Fitch algorithm reconstructs the state at the root ambiguously, we select one of the states with equal probability (Fischer and Thatte 2009).
To illustrate this definition, consider the case with $\mathcal{A} = \{\alpha, \beta, \gamma, \delta\}$. In this case, the reconstruction accuracy for the Fitch algorithm for ancestral state reconstruction is given by

$$RA(X) = P_\alpha(X) + \frac{1}{2} \cdot (P_\alpha\beta(X) + P_\alpha\gamma(X) + P_\alpha\delta(X))$$

$$+ \frac{1}{3} \cdot (P_\alpha\beta\gamma(X) + P_\alpha\beta\delta(X) + P_\alpha\gamma\delta(X)) + \frac{1}{4} \cdot P_\alpha\beta\gamma\delta(X),$$

(2)

where we define

$$P_\alpha(X) := \Pr(\mathcal{M}(f, T) = \{\alpha\}|\rho = \alpha),$$

$$P_{\alpha\beta}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \beta\}|\rho = \alpha),$$

$$P_{\alpha\gamma}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \gamma\}|\rho = \alpha),$$

$$P_{\alpha\delta}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \delta\}|\rho = \alpha),$$

$$P_{\alpha\beta\gamma}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \beta, \gamma\}|\rho = \alpha),$$

$$P_{\alpha\beta\delta}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \beta, \delta\}|\rho = \alpha),$$

$$P_{\alpha\gamma\delta}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \gamma, \delta\}|\rho = \alpha),$$

$$P_{\alpha\beta\gamma\delta}(X) := \Pr(\mathcal{M}(f, T) = \{\alpha, \beta, \gamma, \delta\}|\rho = \alpha).$$

The main aim of this manuscript is to show that the reconstruction accuracy for a rooted binary ultrametric phylogenetic tree under the $N_4$-model using all terminal taxa is more accurate, or at least as accurate, for ancestral state reconstruction than using any particular terminal taxon. This provides a lower bound on $RA(X)$ and is stated in the following theorem, where $p$ denotes the probability for one specific change, e.g., from $\alpha$ to $\beta$, from the root to any of the leaves.

**Theorem 1** For any rooted binary phylogenetic ultrametric tree and the $N_4$-model, the Fitch algorithm using all terminal taxa is more accurate, or at least as accurate, for ancestral state reconstruction than using any particular terminal taxon, that is,

$$RA(X) \geq 1 - 3p.$$ 

The proof of Theorem 1 requires some more general properties. Therefore, we first turn our attention to the following. If not stated otherwise, we always consider rooted binary ultrametric phylogenetic trees under the $N_4$-model. Due to the symmetry of the model, we can assume without loss of generality that the root is in state $\alpha$, so $\alpha$ evolves along the tree to a character $f$ on $X$. Thus, $p$ is the probability that from the root to one leaf the state changes from $\alpha$ to one specific state in $\mathcal{A} \setminus \{\alpha\} = \{\beta, \gamma, \delta\}$, i.e., $3p$ is the probability that a given leaf is not in state $\alpha$.

Therefore, in the case of the $N_4$-model, $1 - 3p$ is the probability that the root is in the same state as one leaf, since three different changes ($\alpha \to \beta, \alpha \to \gamma, \alpha \to \delta$) can occur. This is at the same time the reconstruction accuracy when only one leaf is taken into account. The main aim of this paper is to show that $1 - 3p$ is a lower bound for $RA(X)$, that is, considering all taxa under a molecular clock is always better, or as good as, considering just one taxon.
As shown in Fig. 2, every binary tree $T$ can be decomposed into two maximal pending subtrees $T_1$ and $T_2$ with leaf sets $Y_1$ and $Y_2$ ($X = Y_1 \cup Y_2$, $Y_1 \cap Y_2 = \emptyset$). This is the so-called standard decomposition (Semple and Steel 2003). We denote the children of $\rho$ by $y_1$ and $y_2$, and with probability $p_i$ one specific change occurs from $\rho$ to $y_i$ ($i \in \{1, 2\}$). Analogously, one specific change occurs from $y_i$ to any leaf with probability $p'_i$ ($i \in \{1, 2\}$). Note that $p$ can then be calculated by all possibilities given for one specific change from $\rho$ to any leaf. Suppose that the root is in state $\alpha$ and leaf $l$ in state $\beta$ (without loss of generality we have $l \in Y_1$). Then, there are four different possibilities for a change from $\rho = \alpha$ to $l = \beta$:

\[
\begin{align*}
\rho = \alpha & \to y_1 = \alpha \to l = \beta, \\
\rho = \alpha & \to y_1 = \beta \to l = \beta, \\
\rho = \alpha & \to y_1 = \gamma \to l = \beta, \\
\rho = \alpha & \to y_1 = \delta \to l = \beta.
\end{align*}
\]

Thus,

\[
p = (1 - 3p_i)p'_i + p_i(1 - 3p'_i) + p_i p'_i + p_i p'_i
= p_i + p'_i - 4p_i p'_i.
\] (3)

Therefore, the probability $p$ for a particular change from $\rho = \alpha$ to any leaf can be calculated using (3), regardless of the fact if this leaf is in state $\beta$, $\gamma$ or $\delta$. This is due to the symmetry of the model as introduced in Preliminaries (e.g., $P(v = \beta | u = \alpha) = P(v = \gamma | u = \alpha)$).

Under the model assumptions of the $N_4$-model, due to the symmetry (e.g., $P(v = \beta | \rho = \alpha) = P(v = \gamma | \rho = \alpha)$), we have that
\[ P_{\alpha\beta}(X) = P_{\alpha\gamma}(X) = P_{\alpha\delta}(X), \] (4)
\[ P_{\alpha\beta\gamma}(X) = P_{\alpha\beta\delta}(X) = P_{\alpha\gamma\delta}(X), \] (5)
since, for example,
\[ P_{\alpha\beta}(X) = \mathbb{P}(\mathcal{M}(f, T) = \{\alpha, \beta\} | \rho = \alpha) = \mathbb{P}(\mathcal{M}(f, T) = \{\alpha, \gamma\} | \rho = \alpha) = P_{\alpha\gamma}(X). \]

Therefore, by (2), (4) and (5), \( RA(X) \) can be simplified and becomes
\[ RA(X) = P_{\alpha}(X) + \frac{3}{2} P_{\alpha\beta}(X) + P_{\alpha\beta\gamma}(X) + \frac{1}{4} P_{\alpha\beta\gamma\delta}(X). \] (6)

Moreover, we define
\[ P_{\beta}(X) := \mathbb{P}(\mathcal{M}(f, T) = \{\beta\} | \rho = \alpha), \]
\[ P_{\gamma}(X) := \mathbb{P}(\mathcal{M}(f, T) = \{\gamma\} | \rho = \alpha), \]
\[ P_{\delta}(X) := \mathbb{P}(\mathcal{M}(f, T) = \{\delta\} | \rho = \alpha), \]
\[ P_{\beta\gamma}(X) := \mathbb{P}(\mathcal{M}(f, T) = \{\beta, \gamma\} | \rho = \alpha), \]
\[ P_{\beta\gamma\delta}(X) := \mathbb{P}(\mathcal{M}(f, T) = \{\beta, \gamma, \delta\} | \rho = \alpha). \]

Again, by the symmetry of the \( N_4 \)-model, we obtain
\[ P_{\beta}(X) = P_{\gamma}(X) = P_{\delta}(X). \] (7)

Biologically, this means that under the assumption that \( \alpha \) is the true root state, the probability that \( \alpha \) evolves to a character for which the Fitch algorithm assigns \( \{\beta\} \) to the root is the same as for \( \{\gamma\} \) and \( \{\delta\} \), since each specific change occurs with probability \( p \).

This brings us to our next result, where \( P_{\alpha}(X), P_{\beta}(X), P_{\alpha\beta}(X), P_{\beta\gamma}(X), P_{\alpha\beta\gamma}(X) \) and \( P_{\beta\gamma\delta}(X) \) are linked to each other, and the root is again without loss of generality assumed to be in state \( \alpha \).

**Lemma 1** For any rooted binary phylogenetic tree and the \( N_4 \)-model, we have that
\[ P_{\alpha}(X) \geq P_{\beta}(X), \]
\[ P_{\alpha\beta}(X) \geq P_{\beta\gamma}(X), \]
\[ P_{\alpha\beta\gamma}(X) \geq P_{\beta\gamma\delta}(X). \]

Note that Lemma 1 does not require the underlying tree to be ultrametric. The proof of Lemma 1 is by induction on \( n \) and is presented in the online supplementary material.\(^1\) For this proof and also for the proof of Theorem 1, we state some recursions required for the induction. Therefore, we define \( f_{Y_i} \) as a restriction of \( f \) to \( Y_i \subseteq X \)

\(^1\) The online supplementary material is available from: http://www.mareikefischer.de/SupplementaryMaterial/Herbst_Fischer_Supplementary_Material.pdf.
for \( i \in \{1, 2\} \): \( f_Y := f|_{Y_i} \). For \( i \in \{1, 2\} \), the probability \( P(A)(Y_i) \) to obtain a set \( A \in \{\{\alpha\}, \{\beta\}, \{\alpha, \beta\}, \{\beta, \gamma\}, \{\alpha, \beta, \gamma\}, \{\beta, \gamma, \delta\}, \{\alpha, \beta, \gamma, \delta\}\} \) as estimate state for \( Y_i \) with the Fitch algorithm under the assumption that \( \rho \) is in state \( \alpha \) can be defined using the law of total probability:

\[
P(A)(Y_i) := \mathbb{P}(\text{MP}(f_Y, T_i) = A)
\]

\[
= (1 - 3p_i)\mathbb{P}(\text{MP}(f_Y, T_i) = A|\gamma_i = \alpha) + p_i\mathbb{P}(\text{MP}(f_Y, T_i) = A|\gamma_i = \beta)
\]

\[
+ p_i\mathbb{P}(\text{MP}(f_Y, T_i) = A|\gamma_i = \delta).
\]

Then, with (4), (5) and (7) we have:

\[
P(\alpha)(Y_i) = (1 - 3p_i)P_\alpha(Y_i) + 3p_iP_\beta(Y_i),
\]

(8)

\[
P(\beta)(Y_i) = (1 - p_i)P_\beta(Y_i) + p_iP_\alpha(Y_i) = P_\gamma(Y_i) = P_\delta(Y_i),
\]

(9)

\[
P(\alpha\beta)(Y_i) = (1 - 2p_i)P_{\alpha\beta}(Y_i) + 2p_iP_{\beta\gamma}(Y_i) = P_{(\alpha\beta\gamma)}(Y_i) = P_{(\alpha\beta\delta)}(Y_i),
\]

(10)

\[
P_{(\beta\gamma)}(Y_i) = (1 - 2p_i)P_{\beta\gamma}(Y_i) + 2p_iP_{\beta\delta}(Y_i) = P_{(\beta\gamma\delta)}(Y_i) = P_{(\beta\gamma\delta)}(Y_i),
\]

(11)

\[
P_{(\alpha\beta\gamma\delta)}(Y_i) = (1 - 3p_i)P_{(\alpha\beta\gamma\delta)}(Y_i) + 3p_iP_{\alpha\beta\gamma}(Y_i),
\]

(12)

\[
P_{(\alpha\beta\gamma\delta)}(Y_i) = P_{\alpha\beta\gamma\delta}(Y_i).
\]

(13)

With (8), (9), (10), (11), (12), (13) and (14), we therefore have

\[
P_\alpha(X) = P(\alpha)(Y_1)P(\alpha)(Y_2) + 3P(\alpha)(Y_1)P(\alpha\beta)(Y_2) + 3P(\alpha\beta)(Y_1)P(\alpha)(Y_2)
\]

\[
+ 3P(\alpha)(Y_1)P(\alpha\beta\gamma)(Y_2) + 3P(\alpha\beta\gamma)(Y_1)P(\alpha)(Y_2)
\]

\[
+ 6P(\alpha)(Y_1)P(\alpha\beta)(Y_2) + 3P(\alpha\beta)(Y_1)P(\alpha\beta\gamma)(Y_2)
\]

\[
+ 3P(\alpha\beta\gamma)(Y_1)P(\alpha\beta)(Y_2) + P(\alpha)(Y_1)P(\alpha\beta\gamma\delta)(Y_2)
\]

\[
+ P(\alpha\beta\gamma\delta)(Y_1)P(\alpha)(Y_2),
\]

(15)

\[
P_{\alpha\beta}(X) = P(\alpha)(Y_1)P(\beta)(Y_2) + P(\beta)(Y_1)P(\alpha)(Y_2) + P(\alpha)(Y_1)P(\alpha\beta)(Y_2)
\]

\[
+ 2P(\alpha)(Y_1)P(\alpha\beta\gamma)(Y_2) + 2P(\alpha\beta\gamma)(Y_1)P(\alpha\beta)(Y_2)
\]

\[
+ P(\alpha\beta)(Y_1)P(\alpha\beta\gamma\delta)(Y_2) + P(\alpha\beta\gamma\delta)(Y_1)P(\alpha\beta)(Y_2)
\]

\[
+ 2P(\alpha)(Y_1)P(\alpha\beta\gamma\delta)(Y_2),
\]

(16)

\[
P_{\alpha\beta\gamma}(X) = P(\alpha)(Y_1)P(\beta\gamma)(Y_2) + P(\beta\gamma)(Y_1)P(\alpha)(Y_2) + 2P(\beta\gamma)(Y_1)P(\alpha\beta)(Y_2)
\]

\[
+ 2P(\alpha\beta)(Y_1)P(\beta\gamma)(Y_2) + P(\alpha\beta\gamma)(Y_1)P(\alpha\beta)(Y_2)
\]

\[
+ P(\alpha\beta\gamma)(Y_1)P(\alpha\beta\gamma\delta)(Y_2) + P(\alpha\beta\gamma\delta)(Y_1)P(\alpha\beta\gamma)(Y_2),
\]

(17)

\[
P_{\alpha\beta\gamma\delta}(X) = P(\alpha)(Y_1)P(\beta\gamma\delta)(Y_2) + P(\beta\gamma\delta)(Y_1)P(\alpha)(Y_2) + 3P(\beta\gamma)(Y_1)P(\alpha\beta\gamma)(Y_2)
\]

\[
+ 3P(\alpha\beta\gamma)(Y_1)P(\beta\gamma)(Y_2) + 3P(\alpha\beta)(Y_1)P(\beta\gamma)(Y_2)
\]

\[
+ 3P(\beta\gamma)(Y_1)P(\alpha\beta)(Y_2) + P(\alpha\beta\gamma\delta)(Y_1)P(\alpha\beta\gamma)(Y_2).\]

(18)

As stated before, all these recursions are needed for the proof of Lemma 1 and Theorem 1. Now, we are in the position to prove Theorem 1, our main result, which states a lower bound on \( RA(X) \).
Proof The proof is by induction on $n$. In order to show $RA(X) \geq 1 - 3p$, we define $D(X) := RA(X) - (1 - 3p)$ and show that $D(X)$ is nonnegative. For $n = 2$, the subtrees $Y_1$ and $Y_2$ both contain one leaf, and thus

$$D(X) = P_\alpha(X) + \frac{3}{2} P_{\alpha\beta}(X) + P_{\alpha\beta\gamma}(X) + \frac{1}{4} P_{\alpha\beta\gamma\delta}(X) - 1 + 3p \quad \text{by (6)}$$

$$= P_\alpha(X) + \frac{3}{2} P_{\alpha\beta}(X) - 1 + 3p$$

since $P_{\alpha\beta\gamma}(X) = P_{\alpha\beta\gamma\delta}(X) = 0$ for $n = 2$

$$= (1 - 3p)^2 + \frac{3}{2} 2(1 - 3p)p - 1 + 3p$$

$$= 1 - 6p + 9p^2 + 3p - 9p^2 - 1 + 3p$$

$$= 0.$$  

This shows that $D(X) = RA(X) - (1 - 3p) = 0$ is nonnegative and thus $RA(X) = 1 - 3p$, which completes the base case of the induction.

Now, we show by induction that $D(X)$ is nonnegative. Suppose that $T$ has $n$ taxa and that $D(X)$ is nonnegative for all trees having fewer than $n$ taxa. We define $D_i := D(Y_i) = RA(Y_i) - (1 - 3p_i)$ for $i \in \{1, 2\}$. Thus, $D_1$ and $D_2$ are nonnegative since $Y_1$ and $Y_2$ contain both fewer than $n$ taxa. Furthermore, for $i \in \{1, 2\}$ we define $P_i := 1 - 4p_i$, and similarly $P := 1 - 4p$.

By elementary term conversion, we can show that

$$8D(X) = \left(4P_{(\alpha)}(Y_1) + 10P_{(\alpha\beta)}(Y_1) + 6P_{(\alpha\beta\gamma)}(Y_1)\right)\left(P_{\alpha\beta}(Y_2) - P_{\beta\gamma}(Y_2)\right)P_2$$

$$+ \left(4P_{(\alpha)}(Y_2) + 10P_{(\alpha\beta)}(Y_2) + 6P_{(\alpha\beta\gamma)}(Y_2)\right)\left(P_{\alpha\beta}(Y_1) - P_{\beta\gamma}(Y_1)\right)P_1$$

$$+ \left(2P_{(\alpha)}(Y_1) + \frac{16}{3} P_{(\alpha\beta)}(Y_1) + 2P_{(\alpha\beta\gamma)}(Y_1)\right)\left(P_{\alpha\beta\gamma}(Y_2) - P_{\beta\gamma\delta}(Y_2)\right)P_2$$

$$+ \left(2P_{(\alpha)}(Y_2) + \frac{16}{3} P_{(\alpha\beta)}(Y_2) + 2P_{(\alpha\beta\gamma)}(Y_2)\right)\left(P_{\alpha\beta\gamma}(Y_1) - P_{\beta\gamma\delta}(Y_1)\right)P_1$$

$$+ 4P_1D_1 + 4P_2D_2$$

$$+ \left(\frac{2}{3} P_{(\alpha\beta)}(Y_1) + 2P_{(\alpha\beta\gamma)}(Y_1) + P_{(\alpha\beta\gamma\delta)}(Y_1)\right)\left(3P + 4P_2D_2\right)$$

$$+ \left(\frac{2}{3} P_{(\alpha\beta)}(Y_2) + 2P_{(\alpha\beta\gamma)}(Y_2) + P_{(\alpha\beta\gamma\delta)}(Y_2)\right)\left(3P + 4P_1D_1\right) \quad \text{(19)}$$

The exact conversions are provided in the online supplementary material (see footnote 1). Moreover, note that $P_1$, $P_{(\alpha)}(Y_i)$, $P_{(\alpha\beta)}(Y_i)$, $P_{(\alpha\beta\gamma)}(Y_i)$, $P_{(\alpha\beta\gamma\delta)}(Y_i)$ are all probabilities and therefore are all nonnegative for $i \in \{1, 2\}$. By Lemma 1, we have that (for $i \in \{1, 2\}$) $P_{\alpha\beta}(Y_i) - P_{\beta\gamma}(Y_i)$ and $P_{\alpha\beta\gamma}(Y_i) - P_{\beta\gamma\delta}(Y_i)$ are nonnegative, resulting in (19) being nonnegative. This implies $D(X) \geq 0$ and thus $RA(X) \geq 1 - 3p$. This completes the proof.  

\[ \square \]
We have shown that the reconstruction accuracy using all terminal taxa is always greater or equal than the conservation probability of one single taxon. Moreover, the base case of the proof of Theorem 1 provides more insight into the reconstruction accuracy of using 2-taxon trees under the $N_4$-model.

**Corollary 1** Let $T$ be a rooted binary ultrametric phylogenetic tree on taxon set $X$ with $|X| = 2$. Let $p$ denote the probability of change from the root to any leaf under the $N_4$-model. Then, the reconstruction accuracy for ancestral state reconstruction using the Fitch algorithm is given by

$$RA(X) = 1 - 3p.$$  

Corollary 1 states the reconstruction accuracy for ancestral state reconstruction with the Fitch algorithm using ultrametric 2-taxon trees, which is the same probability when using one terminal taxon. In the following proposition, we show that the reconstruction accuracy with the Fitch algorithm using any two terminal taxa of a taxa set $X$ is also $1 - 3p$.

**Proposition 1** For any rooted binary phylogenetic ultrametric tree and the $N_4$-model, the reconstruction accuracy for the Fitch algorithm using any two terminal taxa $x_1, x_2 \in X$ for ancestral state reconstruction is given by

$$RA(\{x_1, x_2\}) = 1 - 3p.$$  

**Proof** Let $x_1, x_2 \in X$ be two terminal taxa of any rooted binary ultrametric phylogenetic tree $T$. Moreover, we consider the standard decomposition of $T$ into its two maximal pending subtrees $T_1$ and $T_2$ as depicted in Fig. 2. Thus, the proof is divided into two cases. In the first case, we have without loss of generality $x_1 \in Y_1$ and $x_2 \in Y_2$. By Corollary 1, the reconstruction accuracy using $x_1$ and $x_2$ is then $RA(\{x_1, x_2\}) = 1 - 3p$.

In the second case, we have either $x_1, x_2 \in Y_1$ or $x_1, x_2 \in Y_2$. Thus, without loss of generality we consider $x_1, x_2 \in Y_1$ as depicted in Fig. 3. Let $y$ be the last common ancestor of $x_1$ and $x_2$, i.e., the first node that occurs both on the path from $x_1$ to $\rho$ as well as on the path from $x_2$ to $\rho$. Let $\hat{T}$ be the subtree of $T_1$ that consists of the paths from $y$ to $x_1$ and $x_2$, respectively, as well as all vertices which lie on one of these paths. $\hat{T}$ is depicted with dotted lines in Fig. 3. Thus, the root of $\hat{T}$ is $y$. In addition, let $\hat{p}$ be the probability for one specific change from $\rho$ to $y$, and let $\hat{p}_x$ be the probability for one specific change from $y$ to $x_1$ or $x_2$.

By (6) we have

$$RA(\{x_1, x_2\}) = P_\alpha(\{x_1, x_2\}) + 3P_\alpha\beta(\{x_1, x_2\}).$$  

Note that $P_{\alpha\beta\gamma}(\{x_1, x_2\}) = P_{\alpha\beta\gamma\delta}(\{x_1, x_2\}) = 0$ since we cannot obtain sets with more than two elements with the Fitch algorithm when only $x_1$ and $x_2$ are used for the reconstruction.

In the following, we use the notation $f|_{\{x_1, x_2\}}$ for the restriction of character $f$ on taxa $x_1$ and $x_2$.  

© Springer
Fig. 3  Illustration of a rooted binary ultrametric phylogenetic tree and the standard decomposition into its two maximal pending subtrees $T_1$ and $T_2$ (Fischer and Thatte 2009). The subtree $\hat{T}$ of $T_1$ is represented by the dotted lines, and the root of $\hat{T}$ is denoted by $y$

Furthermore, we have

$$P_{\alpha}(\{x_1, x_2\}) = P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha\}|\rho = \alpha)$$
$$= (1 - 3\bar{p}) P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha\}|y = \alpha, \rho = \alpha)$$
$$+ \bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha\}|y = \beta, \rho = \alpha)$$
$$+ \bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha\}|y = \gamma, \rho = \alpha)$$
$$+ \bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha\}|y = \delta, \rho = \alpha)$$
$$= (1 - 3\bar{p}) P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha\}|y = \alpha, \rho = \alpha)$$
$$+ 3\bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\beta\}|y = \alpha, \rho = \alpha)$$
by the symmetry of the $N_4$-model
$$= (1 - 3\bar{p})(1 - 3\bar{p})^2 + 3\bar{p}\bar{p}^2. \quad (21)$$

Moreover,

$$P_{\alpha\beta}(\{x_1, x_2\}) = P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha, \beta\}|\rho = \alpha)$$
$$= (1 - 2\bar{p}) P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha, \beta\}|y = \alpha, \rho = \alpha)$$
$$+ \bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha, \beta\}|y = \gamma, \rho = \alpha)$$
$$+ \bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha, \beta\}|y = \delta, \rho = \alpha)$$
$$= (1 - 2\bar{p}) P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\alpha, \beta\}|y = \alpha, \rho = \alpha)$$
$$+ 2\bar{p} P(\text{MP}(f |_{\{x_1, x_2\}}, \hat{T}) = \{\beta, \gamma\}|y = \alpha, \rho = \alpha)$$
by the symmetry of the $N_4$-model
$$= (1 - 2\bar{p}) \left(1 - 3\bar{p}\right) \bar{p} + 2 \bar{p} \bar{p}^2. \quad (22)$$
Thus, by (21) and (22), (20) becomes

\[
RA(\{x_1, x_2\}) = (1 - 3p)(1 - 3\hat{p})^2 + 3p\hat{p}^2 + \frac{3}{2}\left((1 - 2p)(1 - 3\hat{p})\hat{p} + 2\hat{p}\right)2\\
= 1 - 3p - 3\hat{p} + 12p\hat{p}
\]

since similar to (3) we have that \( p = \bar{p} + \hat{p} - 4p\hat{p} \).

Therefore, in both cases \( RA(\{x_1, x_2\}) = 1 - 3p \) which completes the proof. 

This proposition provides us the reconstruction accuracy for the Fitch algorithm when any two terminal taxa are considered. Note that this reconstruction accuracy is the same as when only one terminal taxon is taken into account. Therefore, by Theorem 1 and Proposition 1 we have the following corollary, which states that the lower bound on the reconstruction accuracy holds for any two terminal taxa. In particular, considering two taxa rather than one cannot improve the lower bound given by Theorem 1.

**Corollary 2** For any rooted binary phylogenetic ultrametric tree and the \( N_4 \)-model, the Fitch algorithm using all terminal taxa is more accurate, or at least as accurate, for ancestral state reconstruction than using any two terminal taxa, that is,

\[
RA(X) \geq 1 - 3p.
\]

This statement completes Sect. 3, and we now have a look on similar results obtained for the \( N_3 \)-model.

### 4 The Accuracy of Ancestral Sequence Reconstruction With Three-state Characters

Under the same assumptions as for the four-state model, similar results can be obtained for the three-state alphabet \( A = \{\alpha, \beta, \gamma\} \). In this case, the reconstruction accuracy is given by

\[
RA(X) = P_\alpha(X) + \frac{1}{2} \cdot (P_{\alpha\beta}(X) + P_{\alpha\gamma}(X)) + \frac{1}{3} \cdot P_{\alpha\beta\gamma}(X)
\]

\[
= P_\alpha(X) + P_{\alpha\beta}(X) + \frac{1}{3} \cdot P_{\alpha\beta\gamma}(X).
\]

Then, Theorem 2 can be formulated similarly to the statement before. The proof is left out, since it can be done analogously. However, we want to emphasize that the conjecture stated by Li et al. also holds for the \( N_3 \)-model.

**Theorem 2** For any rooted binary phylogenetic ultrametric tree and the \( N_3 \)-model, the Fitch algorithm using all terminal taxa is more accurate, or at least as accurate, for ancestral state reconstruction than using any particular terminal taxon, that is,

\[
RA(X) \geq 1 - 2p.
\]
By Theorem 2, a lower bound on $RA(X)$ for rooted binary ultrametric phylogenetic trees is also given for $A = \{\alpha, \beta, \gamma\}$. Note that the analogs of Lemma 1, Corollary 1, Proposition 1 and Corollary 2 also hold under the $N_3$-model. In particular, the reconstruction accuracy for ultrametric trees is then at least $1 - 2p$. The exact statements and the proof of Proposition 2 are given subsequently. The proof of Lemma 2 can be found in the online supplementary material (See footnote 1).

**Lemma 2** For any rooted binary phylogenetic tree and the $N_3$-model, we have that

$$P_\alpha(X) \geq P_\beta(X),$$

$$P_{\alpha\beta}(X) \geq P_{\beta\gamma}(X).$$

Note that Lemma 2 does also not require the underlying tree to be ultrametric.

**Corollary 3** Let $T$ be a rooted binary ultrametric phylogenetic tree on taxon set $X$ with $|X| = 2$. Let $p$ denote the probability of change from the root to any leaf under the $N_3$-model. Then, the reconstruction accuracy for ancestral state reconstruction using the Fitch algorithm is given by

$$RA(X) = 1 - 2p.$$

**Proposition 2** For any rooted binary phylogenetic ultrametric tree and the $N_3$-model, the reconstruction accuracy for the Fitch algorithm using any two terminal taxa $x_1, x_2 \in X$ for ancestral state reconstruction is given by

$$RA(\{x_1, x_2\}) = 1 - 2p.$$  

**Proof** Let $x_1, x_2 \in X$ be two terminal taxa of any rooted binary ultrametric phylogenetic tree $T$. Moreover, we consider the standard decomposition of $T$ into its two maximal pending subtrees $T_1$ and $T_2$ as depicted in Fig. 2. Thus, the proof is divided into two cases. In the first case, we have without loss of generality $x_1 \in Y_1$ and $x_2 \in Y_2$. By Corollary 3, the reconstruction accuracy using $x_1$ and $x_2$ is then $RA(\{x_1, x_2\}) = 1 - 2p$.

In the second case, we have either $x_1, x_2 \in Y_1$ or $x_1, x_2 \in Y_2$. Thus, without loss of generality we consider $x_1, x_2 \in Y_1$ as depicted in Fig. 3. Let $y$ be the last common ancestor of $x_1$ and $x_2$, i.e., the first node that occurs both on the path from $x_1$ to $\rho$ as well as on the path from $x_2$ to $\rho$. Let $\hat{T}$ be the subtree of $T_1$ consisting of the paths from $y$ to $x_1$ and $x_2$, respectively. $\hat{T}$ is depicted with dotted lines in Fig. 3. Thus, the root of $\hat{T}$ is $y$. In addition, let $\overline{p}$ be the probability for one specific change from $\rho$ to $y$, and let $\hat{p}$ be the probability for one specific change from $y$ to $x_1$ or $x_2$. By (23), we have

$$RA(\{x_1, x_2\}) = P_\alpha(\{x_1, x_2\}) + P_{\alpha\beta}(\{x_1, x_2\}).$$  

(24)

Note that $P_{\alpha\beta\gamma}(\{x_1, x_2\}) = 0$ since we cannot obtain sets with more than two elements with the Fitch algorithm when only $x_1$ and $x_2$ are used for the reconstruction.
In the following, we use the notation \( f|_{[x_1, x_2]} \) for the restriction of character \( f \) on taxa \( x_1 \) and \( x_2 \).

Furthermore, we have

\[
P_\alpha([x_1, x_2]) = \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha\}|\rho = \alpha) \\
= (1 - 2\overline{p}) \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha\}|y = \alpha, \rho = \alpha) \\
+ \overline{p} \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha\}|y = \beta, \rho = \alpha) \\
+ \overline{p} \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha\}|y = \gamma, \rho = \alpha) \\
= (1 - 2\overline{p}) \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha\}|y = \alpha, \rho = \alpha) \\
+ 2\overline{p} \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\beta\}|y = \alpha, \rho = \alpha)
\]

by the symmetry of the \( N_3 \)-model 

\[
= (1 - 2\overline{p})(1 - 2\hat{p})^2 + 2\overline{p}\hat{p}^2. \quad (25)
\]

Moreover,

\[
P_{\alpha\beta}([x_1, x_2]) = \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha, \beta\}|\rho = \alpha) \\
= (1 - 2\overline{p}) \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha, \beta\}|y = \alpha, \rho = \alpha) \\
+ \overline{p} \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha, \beta\}|y = \beta, \rho = \alpha) \\
+ \overline{p} \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha, \beta\}|y = \gamma, \rho = \alpha) \\
= (1 - \overline{p}) \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\alpha, \beta\}|y = \alpha, \rho = \alpha) \\
+ \overline{p} \mathbb{P}(\text{MP}(f|_{[x_1, x_2]}, \hat{T}) = \{\beta, \gamma\}|y = \alpha, \rho = \alpha)
\]

by the symmetry of the \( N_3 \)-model 

\[
= (1 - \overline{p}) 2\overline{p} (1 - 2\hat{p}) \hat{p} + \overline{p} 2\hat{p}^2. \quad (26)
\]

Thus, by (25) and (26), (24) becomes

\[
RA([x_1, x_2]) = (1 - 2\overline{p})(1 - 2\hat{p})^2 + 2\overline{p}\hat{p}^2 + (1 - \overline{p}) 2\overline{p} (1 - 2\hat{p}) \hat{p} + \overline{p} 2\hat{p}^2 \\
= 1 - 2\overline{p} - 2\hat{p} + 6\overline{p}\hat{p} \\
= 1 - 2\overline{p} \quad \text{since} \quad \overline{p} = \overline{p} + \hat{p} - 3\overline{p}\hat{p}.
\]

Therefore, in both cases \( RA([x_1, x_2]) = 1 - 2\overline{p} \) which completes the proof. \( \square \)

**Corollary 4** For any rooted binary phylogenetic ultrametric tree and the \( N_3 \)-model, the Fitch algorithm using all terminal taxa is more accurate, or at least as accurate, for ancestral state reconstruction than using any two terminal taxa, that is,

\[
RA(X) \geq 1 - 2\overline{p}.
\]
5 Conclusion and Discussion

In this paper, we considered the reconstruction accuracy of the Fitch algorithm for ancestral state reconstruction. In particular, we analyzed rooted binary ultrametric phylogenetic trees under the $N_4$-model. For an ultrametric tree, the probability of a change from the root to any leaf is the same. For such trees, we investigated a lower bound on the reconstruction accuracy by answering affirmatively the conjecture by Li, Steel and Zhang, which stated that for rooted binary ultrametric phylogenetic trees under the symmetric $N_2$-model the reconstruction accuracy using all terminal taxa is at least as high as the conservation probability of any leaf. In 2009, Fischer and Thatte had already shown that this conjecture holds for two-state characters, but it remained unknown whether this result could be extended to three or more character states. In particular, the biologically relevant case of $r = 4$, which corresponds to the DNA or RNA alphabet, remained unclear.

The main result of this manuscript is the proof of the conjecture for $r = 4$, which provides a lower bound on the reconstruction accuracy. As mentioned before, the conjecture also holds for the $N_3$-model. In the past, several studies showed that in some cases, the Fitch algorithm provides better results when some data are disregarded, e.g., for the symmetric $N_2$-model (Li et al. 2008; Fischer and Thatte 2009). This led to a critical view on maximum parsimony as a method for ancestral state reconstruction. But as we have shown here, at least for ultrametric trees, the extreme case of disregarding all data except for one or two leaves can never improve the reconstruction accuracy of the Fitch algorithm. In this sense, our results are good news for maximum parsimony as a method for ancestral state reconstruction. Moreover, it is known from the literature that the maximum parsimony root state estimate, which can be calculated using the Fitch algorithm, and the maximum likelihood root state estimate coincide under the symmetric $r$-state model $N_r$ (Tuffley and Steel 1997, Theorem 6). Thus, our results also hold for maximum likelihood.

To conclude, the generalization to the $N_r$-model for $r > 4$ (e.g., $r = 20$ for the protein alphabet) is still open, but we conjecture that it also holds. Moreover, an explicit analysis of a similar setting for other models, e.g., the Kimura 2-parameter model (Kimura 1980), is an interesting topic for future research.

References

Fischer M, Thatte BD (2009) Maximum parsimony on subsets of taxa. J Theor Biol 260(2):290–293. https://doi.org/10.1016/j.jtbi.2009.06.010
Fitch WM (1971) Toward defining the course of evolution: Minimum change for a specific tree topology. Syst Zool 2(4):406–416. https://doi.org/10.2307/2412116
Jukes T, Cantor C (1969) Evolution of protein molecules. Mamm Protein Metab 3:21–132. https://doi.org/10.1016/B978-1-4832-3211-9.50009-7
Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16(2):111–120. https://doi.org/10.1007/BF01731581
Li G, Steel M, Zhang L (2008) More taxa are not necessarily better for the reconstruction of ancestral character states. Syst Biol 57(4):647–653. https://doi.org/10.1080/10635150802203898
Liberles (ed) DA (2007) Ancestral Sequence Reconstruction. Oxford University Press, NY. https://doi.org/10.1093/acprof:oso/9780199299188.001.0001
Regner P (2012) Phylogenetic trees-selected combinatorial problems. Diplomarbeit, Technical University of Vienna, Vienna
Salisbury BA, Kim J (2001) Ancestral state estimation and taxon sampling density. Syst Biol 50(4):557–564. https://doi.org/10.1080/106351501750435103
Semple C, Steel M (2003) Phylogenetics. Oxford Lecture Series in Mathematics and its Application
Tuffley C, Steel M (1997) Links between maximum likelihood and maximum parsimony under a simple model of site substitution. Bull Math Biol 59(3):581–607. https://doi.org/10.1007/BF02459467
Yang J, Li J, Dong L, Grünewald S (2011) Analysis on the reconstruction accuracy of the fitch method for inferring ancestral states. BMC Bioinform 12(18). https://doi.org/10.1186/1471-2105-12-18
Zhang L, Shen J, Yang J, Li G (2010) Analyzing the fitch method for reconstructing ancestral states on ultrametric phylogenetic trees. Bull Math Biol 72:1760–1782. https://doi.org/10.1007/s11538-010-9505-8