Appendix to *Non-bifurcating phylogenetic tree inference via the adaptive LASSO*

7.1. **Lemmas.** Here we perform further theoretical development to establish the main theorems. We remind the reader that we will continue to assume Assumptions 2.1 and 2.2. The following lemma allows gives a lower bound on the fraction of sites with state assignments in a given set. It will prove useful to obtain an upper bound on the likelihood.

Lemma 7.1. *For any non-empty set A of single-site state assignments to the leaves, we define*

$$k_A = |\{i : \mathbf{Y}^i \in A\}|$$

There exist $c_3 > 0, c_4(\delta, n) > 0$ such that for all k, we have

$$\frac{k_A}{k} \geq c_3 - \frac{c_4}{\sqrt{k}} \qquad \forall A \neq \emptyset$$

with probability at least $1 - \delta$.

Proof of Lemma 7.1. Since the tree distance between any pairs of leaves of the true tree is strictly positive, there exists $c_3 > 0$ such that $P_{q^*}(\psi) \geq c_3$ for all state assignments ψ.

Using Hoeffding's inequality, for any state assignment ψ, we have

$$\mathbb{P}\left[\left|\frac{k_{\{\psi\}}}{k} - P_{q^*}(\psi)\right| \geq t\right] \leq 2e^{-2kt^2}.$$

We deduce that

$$\mathbb{P}\left[\exists \psi \text{ such that } \left|\frac{k_{\{\psi\}}}{k} - P_{q^*}(\psi)\right| \geq t\right] \leq 2e^{-2kt^2} \cdot 4^N.$$

For any given $\delta > 0$, by choosing

$$c_4(\delta, N) = \sqrt{\frac{\log(1/\delta) + (2N+1)\log 2}{2}}$$

and $t = c_4(\delta, N)/\sqrt{k}$ we have

$$\left|\frac{k_{\{\psi\}}}{k} - P_{q^*}(\psi)\right| \leq \frac{c_4(\delta, N)}{\sqrt{k}} \qquad \forall \psi$$

with probability at least $1 - \delta$. This proves the Lemma. □

Lemma 7.2 (Generalization bound)**.** *There exists a constant $C(\delta, n, Q, \eta, g_0, \mu) > 0$ such that for any $k \geq 3$, $\delta > 0$, we have:*

$$\left|\frac{1}{k}\ell_k(q) - \phi(q)\right| \leq C\left(\frac{\log k}{k}\right)^{1/2} \qquad \forall q \in \mathcal{T}(\mu)$$

with probability greater than $1 - \delta$.

Proof. Note that for $q \in \mathcal{T}(\mu)$, $0 \geq \log P_q(\psi) \geq -\mu$ for all state assignments ψ. By Hoeffding's inequality,

$$\mathbb{P}\left[\left|\frac{1}{k}\ell_k(q) - \phi(q)\right| \geq y/2\right] \leq 2\exp\left(-\frac{y^2 k}{2\mu^2}\right).$$

For each $q \in \mathcal{T}(\mu)$, $k > 0$, and $y > 0$, define the events

$$A(q, k, y) = \left\{\left|\frac{1}{k}\ell_k(q) - \phi(q)\right| > y/2\right\}$$

and

$$B(q,k,y)=\left\{\exists q'\in\mathcal{T}(\mu)\text{ such that }\|q'-q\|_2\le\frac{y}{4c_2}\text{ and }\left|\frac{1}{k}\ell_k(q)-\phi(q)\right|>y\right\}$$

then $B(q,k,y)\subset A(q,k,y)$ by the triangle inequality, (3.3), and (3.4). Let

$$y=\sqrt{\frac{C\log k}{k}}$$

Since $\mathcal{T}(\mu)$ is a subset of $\mathbb{R}^{2N-3}$, there exist $C_{2N-3}\ge 1$ and a finite set $\mathcal{H}\subset\mathcal{T}(\mu)$ such that

$$\mathcal{T}(\mu)\subset\bigcup_{q\in\mathcal{H}}V(q,\epsilon)\qquad\text{and}\qquad|\mathcal{H}|\le C_{2N-3}/\epsilon^{2N-3}$$

where $\epsilon=y/(4c_2)$, $V(q,\epsilon)$ denotes the open ball centered at q with radius ϵ, and $|\mathcal{H}|$ denotes the cardinality of $\mathcal{H}$. By a simple union bound, we have

$$\mathbb{P}\left[\exists q\in\mathcal{H}:\left|\frac{1}{k}\ell_k(q)-\phi(q)\right|>y/2\right]\le 2\exp\left(-\frac{y^2k}{2\mu^2}\right)C_{2N-3}\Big/\epsilon^{2N-3}.$$

Using the fact that $B(q,k,y)\subset A(q,k,y)$ for all $q\in\mathcal{H}$, we deduce

$$\mathbb{P}\left[\exists q\in\mathcal{T}(\mu):\left|\frac{1}{k}\ell_k(q)-\phi(q)\right|>y\right]\le 2\exp\left(-\frac{y^2k}{2\mu^2}\right)C_{2N-3}\Big/\epsilon^{2N-3}.$$

To complete the proof, we need to chose C in such a way that

$$C_{2N-3}\left(\frac{4\sqrt{k}g_0c_2}{\sqrt{C\log k}}\right)^{2N-3}\times 2\exp\left(-\frac{C\log k}{2\mu^2}\right)\ \le\ \delta.$$

Since $k\ge 3$ and $C\ge 1$, the inequality is valid if

$$C_{2N-3}\left(4g_0c_2\right)^{2N-3}\times 2k^{\frac{2N-3}{2}-\frac{C}{2\mu^2}}\ \le\ \delta$$

and can be obtained if

$$\frac{2N-3}{2}-\frac{C}{2\mu^2}<0,\qquad\text{and}\qquad C_{2N-3}\left(4g_0c_2\right)^{2N-3}\times 2\cdot 3^{\frac{2N-3}{2}-\frac{C}{2\mu^2}}\ \le\ \delta.$$

In other words, we need to choose C such that

$$C\ge 2\mu^2\left(\log(1/\delta)+\log C_{2N-3}+(2N-3)\log(4\sqrt{3}g_0c_2)\right).$$

This completes the proof. □

7.2. **Proofs of main theorems.**

Proof of Theorem 3.10. By definition of the estimator, we have

$$-\frac{1}{k}\ell_k(q^{k,R_k})+\lambda_kR_k(q^{k,R_k})\le-\frac{1}{k}\ell_k(q^*)+\lambda_kR_k(q^*)$$

which is equivalent to $U_k(q^{k,R_k})\le\lambda_kR_k(q^*)-\lambda_kR_k(q^{k,R_k})$.

We have $q^{k,R_k}\in\mathcal{T}(\mu)$ with probability at least $1-2\delta$ from Lemma 3.9 for k sufficiently large. Therefore by Lemma 3.6,

$$\mathbb{E}[U_k(q^{k,R_k})]\le\frac{1}{k}\qquad\text{or}\qquad\frac{1}{2}\mathbb{E}[U_k(q^{k,R_k})]\le U_k(q^{k,R_k})+\frac{C\log k}{k^{2/\beta}},$$

with probability at least $1 - 3\delta$. The second case implies that

$$\begin{aligned}\frac{c_1^\beta}{2}\|q^{k,R_k} - q^*\|_2^\beta &\le \frac{1}{2}\mathbb{E}[U_k(q^{k,R_k})] \\ &\le \lambda_k R_k(q^*) - \lambda_k R_k(q^{k,R_k}) + \frac{C\log k}{k^{2/\beta}} \le \frac{C\log k}{k^{2/\beta}} + \lambda_k R_k(q^*)\end{aligned}$$

while for the first case, we have

$$\frac{c_1^\beta}{2}\|q^{k,R_k} - q^*\|_2^\beta \le \mathbb{E}[U_k(q^{k,R_k})] \le \frac{1}{k} \le \frac{C\log k}{k^{2/\beta}} + \lambda_k R_k(q^*)$$

since $\beta \ge 2$ and $C \ge 1$. This demonstrates (3.7).

If the additional assumption (3.6) is satisfied, we also have

$$\|q^{k,R_k} - q^*\|_2^\beta \le \frac{C'\log k}{k^{2/\beta}} + C_3\lambda_k\|q^{k,R_k} - q^*\|_2.$$

Using Lemma 3.7 with

$$\nu = 1/\beta, \qquad x = \|q^{k,R_k} - q^*\|_2^\beta, \qquad a = C_3\lambda_k \qquad \text{and} \qquad b = \frac{C'\log k}{k^{2/\beta}},$$

we obtain

$$x \le C_1 a^{1/(1-\nu)} + C_2 b,$$

which implies

$$\|q^{k,R_k} - q^*\|_2^\beta \le C'(\delta, C_3)\left(\frac{\log k}{k^{2/\beta}} + \lambda_k^{\beta/(\beta-1)}\right).$$

This completes the proof. □

Proof of Theorem 3.11. We first note that by Theorem 3.10, the estimator q^{k,R_k} is consistent, which guarantees $\lim_{k\to\infty} q^{k,R_k} = q^*$ almost surely. Thus

$$\lim_{k\to\infty} S_k(q^*) = \lim_{k\to\infty} \sum_{q_i^* \neq 0} (q_i^*)^{1-\gamma} < \infty.$$

The hypotheses of this theorem imply that $\lambda_k \to 0$ and thus by Theorem 3.10, we also deduce that q^{k,S_k} is also a consistent estimator. This validates (i).

To establish topological consistency under (ii), we divide the proof into two steps.

As the first step, we prove that $\lim_k \mathbb{P}(\mathcal{A}(q^*) \subset \mathcal{A}(q^{k,S_k})) = 1$. If $q_{i_0}^* = 0$ for some i_0, then from Theorem 3.10, we have

$$q_{i_0}^{k,R_k} \le C'(\delta)\left(\frac{\log k}{k^{2/\beta}} + \lambda_k^{\beta/(\beta-1)}\right)^{1/\beta} \qquad \forall k$$

with probability at least $1 - \delta$. By the definition of w_{k,i_0}, we have

$$\begin{aligned}\lim_{k\to\infty} \alpha_k w_{k,i_0} \quad &\ge \quad \lim_{k\to\infty} \alpha_k (C'(\delta))^{-\gamma}\left(\frac{\log k}{k^{2/\beta}} + \lambda_k^{\beta/(\beta-1)}\right)^{-\gamma/\beta} \\ &= \quad (C'(\delta))^{-\gamma} \lim_{k\to\infty}\left(\frac{\log k}{\alpha_k^{\beta/\gamma} k^{2/\beta}} + \alpha_k^{-\beta/\gamma}\lambda_k^{\beta/(\beta-1)}\right)^{-\gamma/\beta}\end{aligned}$$

which goes to infinity since by the hypotheses of the Theorem

$$\alpha_k^{\beta/\gamma} \succ \frac{\log k}{k^{2/\beta}} \qquad \text{and} \qquad \alpha_k^{\beta/\gamma} \succ \lambda_k^{\beta/(\beta-1)}.$$

Since $\delta > 0$ is arbitrary, we deduce that $\lim_{k\to\infty} \alpha_k w_{k,i_0} = \infty$ with probability one.

Now for any branch length vector q, we define $f(q)$ as the vector obtained from q by setting the i_0 component of q to 0. By definition of the estimator q^{k,S_k}, we have

$$-\frac{1}{k}\ell_k(q^{k,S_k}) + \alpha_k \sum_i w_{k,i}\, q_i^{k,S_k} \leq -\frac{1}{k}\ell_k(f(q^{k,S_k})) + \alpha_k \sum_i w_{k,i}[f(q^{k,S_k})]_i$$

or equivalently

$$\alpha_k w_{k,i_0} q_{i_0}^{k,S_k} \leq \frac{1}{k}\ell_k(q^{k,S_k}) - \frac{1}{k}\ell_k(f(q^{k,S_k})).$$

Lemma 3.8 establishes that there exist, $\mu^* > 0$ and a neighborhood V of q^* in $\mathcal{T}$ such that $V \subset \mathcal{T}(\mu^*)$. Since the estimator q^{k,S_k} is consistent and $q^*_{i_0} = 0$, we can assume that both q^{k,S_k} and $f(q^{k,S_k})$ belong to $\mathcal{T}(\mu^*)$ with k large enough. Thus, from Lemma 3.5, we have

$$\left|\frac{1}{k}\ell_k(q^{k,S_k}) - \frac{1}{k}\ell_k(f(q^{k,S_k}))\right| \leq c_2\|q^{k,S_k} - f(q^{k,S_k})\|_2 = c_2 q_{i_0}^{k,S_k}.$$

If $q_{i_0}^{k,S_k} > 0$, we deduce that $\alpha_k w_{k,i_0}$ is bounded from above by c_2, which is a contradiction. This implies that $q_{i_0}^{k,S_k} = 0$, and we conclude that

$$\lim_k \mathbb{P}(\mathcal{A}(q^*) \subset \mathcal{A}(q^{k,S_k})) = 1.$$

As the second step, we prove that $\lim_k \mathbb{P}(\mathcal{A}(q^{k,S_k}) \subset \mathcal{A}(q^*)) = 1$. Indeed, the consistency of q^{k,S_k} guarantees that

$$\lim_{k\to\infty} q^{k,S_k} = q^*$$

almost surely. Therefore, if $q^*_{i_0} > 0$ for some i_0, then $q_{i_0}^{k,S_k} > 0$ for k large enough. In other words, we have $\lim_k \mathbb{P}(\mathcal{A}(q^{k,S_k}) \subset \mathcal{A}(q^*)) = 1$.

Combing step 1 and step 2, we deduce that the adaptive estimator is topologically consistent. □

Proof of Lemma 3.12. Since q^{k,S_k} is topologically consistent and q^{k,R_k} is consistent, we have

$$\mathcal{A}(q^{k,S_k}) = \mathcal{A}(q^*) \qquad \text{and} \qquad q_i^{k,R_k} \geq q_i^*/2 \qquad \forall i \notin \mathcal{A}(q^*)$$

with probability one for sufficiently large k. Defining $b = \min_{i\notin\mathcal{A}(q^*)} q_i^*$, we have

$$|S_k(q^{k,S_k}) - S_k(q^*)| = \left|\sum_{q_i^*\neq 0} w_{k,i}(q_i^{k,S_k} - q_i^*)\right| \leq \sqrt{2N-3}\,(b/2)^{-\gamma}\,\|q^{k,S_k} - q^*\|_2$$

via Cauchy-Schwarz which completes the proof. □

Proof of Theorem 3.13. We note that for the LASSO estimator, $R_k^{[0]}(q^*) = \sum_i q_i^*$ is uniformly bounded from above. Hence, the LASSO estimator is consistent. We can then use this as the base case to prove, by induction, that adaptive LASSO and the multiple-step LASSO are consistent via Theorem 3.11 (part (i)). Moreover, $R_k^{[0]}$ is uniformly Lipschitz and satisfies (3.6), so using part (ii) of Theorem 3.11, we deduce that adaptive LASSO (i.e., the estimator with penalty function $R_k^{[1]}$) is topologically consistent.

We will prove that the multiple-step LASSOs are topologically consistent by induction. Assume that $q^{k,R_k^{[m]}}$ is topologically consistent, and that $q^{k,R_k^{[m-1]}}$ is consistent. From Lemma 3.12, we deduce that there exists $C > 0$ independent of k such that

$$\left| R_k^{[m]}\left(q^{k,R_k^{[m]}}\right) - R_k^{[m]}(q^*)\right| \le C \left\| q^{k,R_k^{[m]}} - q^*\right\|_2 \qquad \forall k. \tag{7.1}$$

This enables us to use part (ii) of Theorem 3.11 to conclude that $q^{k,R_k^{[m+1]}}$ is topologically consistent. This inductive argument proves part (i) of the Theorem. We can now use (7.1) and Theorem 3.10 to derive the convergence rate of the estimators. □

7.3. **Technical proofs.**

Lemma 2.3. *If the penalty R_k is continuous on $\mathcal{T}$, then for $\lambda > 0$ and observed sequences $\mathbf{Y}^k$, there exists a $q \in \mathcal{T}$ minimizing*

$$Z_{\lambda,\mathbf{Y}^k}(q) = -\frac{1}{k}\ell_k(q) + \lambda R_k(q).$$

Proof of Lemma 2.3. Let $\{q^n\}$ be a sequence such that

$$Z_{\lambda,\mathbf{Y}^k}(q^n) \to \nu := \inf_q Z_{\lambda,\mathbf{Y}^k}(q).$$

We note that since $\ell_k(q^*) \neq -\infty$ and R_k is continuous on the compact set $\mathcal{T}$, ν is finite. Since $\mathcal{T}$ is compact, we deduce that a subsequence $\{q^m\}$ converges to some $q^0 \in \mathcal{T}$. Since the log likelihood (defined on $\mathcal{T}$ with values in the extended real line $[-\infty, 0]$) and the penalty R_k are continuous, we deduce that q^0 is a minimizer of $Z_{\lambda,\mathbf{Y}^k}$. □

Lemma 3.5. *For any $\mu > 0$, there exists a constant $c_2(N, Q, \eta, g_0, \mu) > 0$ such that*

$$\left|\frac{1}{k}\ell_k(q) - \frac{1}{k}\ell_k(q')\right| \le c_2\|q - q'\|_2 \tag{3.3}$$

and

$$|\phi(q) - \phi(q')| \le c_2\|q - q'\|_2 \tag{3.4}$$

for all $q, q' \in \mathcal{T}(\mu)$.

Proof of Lemma 3.5. Using the same arguments as in the proof of Lemma 4.2 of Dinh et al. (2018), we have

$$\left|\frac{\partial P_q(\psi)}{\partial q_i}\right| \le \varsigma 4^n$$

for any state assignment ψ where ς is the element of largest magnitude in the rate matrix Q. By the Mean Value Theorem, we have

$$|\log P_q(\psi) - \log P_{q'}(\psi)| \le c_2\sqrt{2N-3}\|q - q'\|_2 \qquad \forall q, q', \psi$$

where $c_2 := \varsigma 4^n/e^{-\mu}$, and $\|\cdot\|_2$ is the ℓ_2-distance in $\mathbb{R}^{2N-3}$. This implies both (3.3) and (3.4). □

Lemma 3.6. *Let G_k be the set of all branch length vectors $q \in \mathcal{T}(\mu)$ such that $\mathbb{E}[U_k(q)] \geq 1/k$. Let $\beta \geq 2$ be the constant in Lemma 3.3. For any $\delta > 0$ and previously specified variables there exists $C(\delta, N, Q, \eta, g_0, \mu, \beta) \geq 1$ (independent of k) such that for any $k \geq 3$, we have:*

$$U_k(q) \geq \frac{1}{2}\mathbb{E}[U_k(q)] - \frac{C \log k}{k^{2/\beta}} \qquad \forall q \in G_k$$

with probability greater than $1 - \delta$.

Proof of Lemma 3.6. The difference of average likelihoods $U_k(q)$ is bounded by Lemma 3.5 and the boundedness assumption on $\mathcal{T}$, thus by Hoeffding's inequality

$$\mathbb{P}\left[U_k(q) - \mathbb{E}\left[U_k(q)\right] \leq -y\right] \leq \exp\left(-\frac{2y^2 k}{c_2^2 \|q - q^*\|^2}\right).$$

By choosing $y = \frac{1}{2}\mathbb{E}\left[U_k(q)\right] + t/2$, we have $y^2 \geq t\mathbb{E}\left[U_k(q)\right]$. For any $q \in G_k$, we deduce using (3.5) (and the fact that $\beta \geq 2$) that

$$\mathbb{P}\left[U_k(q) \leq \frac{1}{2}\mathbb{E}\left[U_k(q)\right] - t/2\right] \leq \exp\left(-\frac{2c_1^2 t k \mathbb{E}[U_k(q)]}{c_2^2 \mathbb{E}[U_k(q)]^{2/\beta}}\right) \leq \exp\left(-\frac{2c_1^2 t k^{2/\beta}}{c_2^2}\right).$$

For each $q \in G_k$, define the events

$$A(q, k, t) = \left\{U_k(q) - \frac{1}{2}\mathbb{E}\left[U_k(q)\right] \leq -t/2\right\}$$

and

$$B(q, k, t) = \left\{\exists q' \in G_k \text{ such that } \|q' - q\|_2 \leq \frac{t}{4c_2} \text{ and } U_k(q') - \frac{1}{2}\mathbb{E}\left[U_k(q')\right] \leq -t\right\}$$

then $B(q, k, t) \subset A(q, k, t)$ by the triangle inequality, (3.3), and (3.4). Let

$$t = \frac{C \log k}{k^{2/\beta}}.$$

To obtain a union bound and complete the proof, we need to chose C in such a way that

$$C_{2N-3}\left(\frac{4k^{2/\beta} g_0 c_2}{C \log k}\right)^{2N-3} \times 2\exp\left(-\frac{2c_1^2 C \log k}{c_2^2}\right) \;\leq\; \delta$$

where C_{2N-3} is defined as in the proof of Lemma 7.2. This can be done by choosing

$$C \geq \frac{4\beta c_2^2}{9c_1^2}\Big(\log(1/\delta) + \log C_{2N-3} + (2N-3)\log(4 \cdot 3^{2/\beta} g_0 c_2)\Big).$$

□

Lemma 3.8. *There exist $\mu^* > 0$ and an open neighborhood V of q^* in $\mathcal{T}$ such that $V \subset \mathcal{T}(\mu^*)$.*

Proof of Lemma 3.8. Let

$$\mu^* = -2 \min_{\psi} \log P_{q^*}(\psi)$$

then we have $\log P_{q^*}(\psi) > -\mu^*$ for all state assignments ψ.

For a fixed value of ψ, $\log P_q(\psi)$ is a continuous function of q around q^*. Hence, there exists an neighborhood V_ψ of q^* such that V_ψ is open in $\mathcal{T}$ and $\log P_q(\psi) > -\mu^*$. Let $V = \cap_\psi V_\psi$. Because the set of all possible labels ψ of the leaves is finite, V is open in $\mathcal{T}$ and

$$\log P_q(\psi) > -\mu^* \qquad \forall \psi, \forall q \in V.$$

In other words, we have $V \subset \mathcal{T}(\mu^*)$. □

Lemma 3.9. *If the sequence $\{\lambda_k R_k(q^*)\}$ is bounded, then for any $\delta > 0$, there exist $\mu(\delta) > 0$ and $K(\delta) > 0$ such that for all $k \geq K$, $q^{k,R_k} \in \mathcal{T}(\mu)$ with probability at least $1 - 2\delta$.*

Proof of Lemma 3.9. We first assume that $\mu > \mu^*$, where μ^* is defined in Lemma 3.8. Thus, we have $q^* \in \mathcal{T}(\mu^*) \subset \mathcal{T}(\mu)$. By definition, we have

$$-\frac{1}{k}\ell_k(q^{k,R_k}) + \lambda_k R_k(q^{k,R_k}) \leq -\frac{1}{k}\ell_k(q^*) + \lambda_k R_k(q^*)$$

which implies via Lemma 7.2 that

$$\phi(q^*) - C(\delta)\frac{\log k}{\sqrt{k}} + \lambda_k R_k(q^{k,R_k}) - \lambda_k R_k(q^*) \leq \frac{1}{k}\ell_k(q^{k,R_k}) \tag{7.2}$$

with probability at least $1 - \delta$.

Let c_3 and $c_4(\delta, N)$ be as in Lemma 7.1, and assume that k is large enough such that

$$c_3 - c_4(\delta, N)\frac{\log k}{\sqrt{k}} > 0. \tag{7.3}$$

Denoting the upper bound of $\{\lambda_k R_k(q^*)\}$ by U, we define

$$\mu = \max\left\{-2\left(c_3 - c_4(\delta, N)\frac{\log k}{\sqrt{k}}\right)^{-1}\left(\phi(q^*) - C(\delta)\frac{\log k}{\sqrt{k}} - U\right), \mu^*\right\}.$$

If we assume that $q^{k,R_k} \notin \mathcal{T}(\mu)$, then the set $I = \{\psi : \log P_{q^{k,R_k}}(\psi) \leq -\mu\}$ is non-empty. Using Lemma 7.1, we have

$$\frac{1}{k}\ell_k(q^{k,R_k}) \leq \frac{1}{k}\sum_{Y_i \in I} \log P_{q^{k,R_k}}(Y_i) \leq -\mu \cdot \frac{k_I}{k} \leq -\mu \cdot \left(c_3 - c_4(\delta)\frac{\log k}{\sqrt{k}}\right) \tag{7.4}$$

with probability at least $1 - \delta$.

Combining equations (7.2) and (7.4), and using the fact that $\{\lambda_k R_k(q^*)\}$ is bounded by U, we obtain

$$\phi(q^*) - C(\delta)\frac{\log k}{\sqrt{k}} - U \leq -\mu \cdot \left(c_3 - c_4(\delta, N)\frac{\log k}{\sqrt{k}}\right).$$

This contradicts the choice of μ for k large enough such that (7.3) holds.

We deduce that $q^{k,R_k} \in \mathcal{T}(\mu)$ with probability at least $1 - 2\delta$. □

7.4. **More experimental results.** Here we present additional experimental results for the case of $\gamma > 1$.

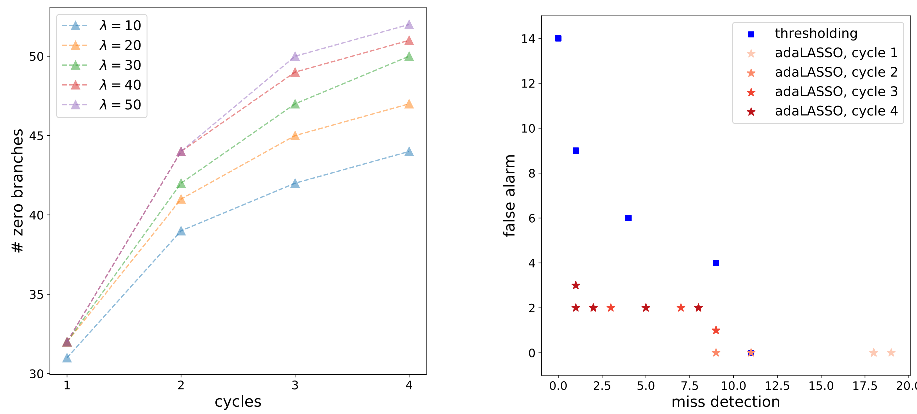

FIGURE S1. Topological consistency comparison of different phylogenetic LASSO procedures on simulation 2. $\gamma = 1.01$.

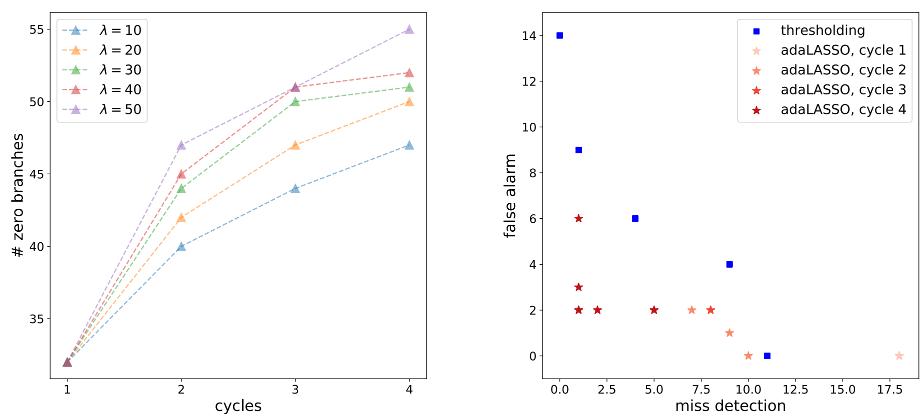

FIGURE S2. Topological consistency comparison of different phylogenetic LASSO procedures on simulation 2. $\gamma = 1.1$.

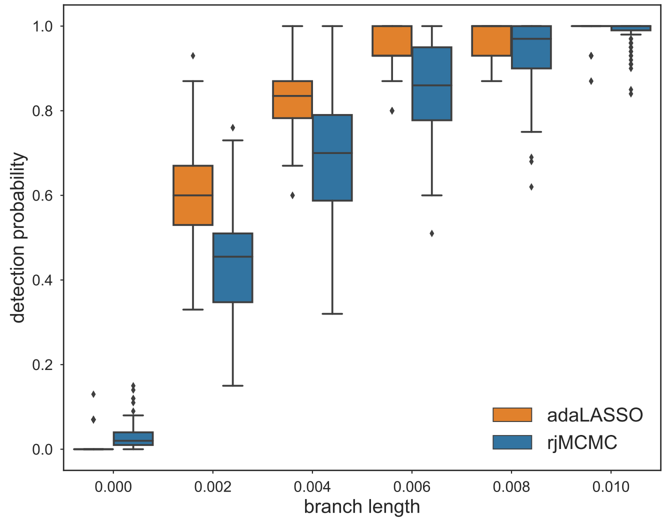

FIGURE S3. Box plot showing performance of multistep adaptive phylogenetic LASSO and rjMCMC at detecting short branches. $\gamma = 1.1$