A Evolutionary model for class switch recombination

A dependence tree $\bar{T}$ with n nodes, sometimes referred to as a state tree, is a tree that defines the conditional independence structure of the random variables associated with the nodes of the tree [24, 25]. Simply put, it is a type of Bayesian network, where the underlying directed acyclic graph is a tree. For each node i in dependence tree $\bar{T}$, we associate a random variable $Y_i \in S$, where S is a discrete state space. Additionally, Y_1 is the random variable associated with the root node. The joint probability $\mathbf{Y} = (Y_1, \ldots, Y_n)$ of these random variables given the underlying structure of dependence tree $\bar{T}$ is defined as follows

$$P(\mathbf{Y} \mid \bar{T}) = P(Y_1) \prod_{i=2}^{n} P(Y_i \mid Y_{\phi(i)}), \tag{5}$$

where $[n] = \{1, \ldots, n\}$ and $\phi(i)$ is a function that returns the parent of node i specified by dependence tree $\bar{T}$. Like Markov chains, this model is parameterized by a distribution over the starting state, $\pi_s = P(Y_1 = s)$, where $\sum_{i \in [r]} \pi_s = 1$, and transition probabilities $p_{s,t} = P(Y_i = t \mid Y_\phi(i) = s)$ from state s to t. Transition probabilities $\mathbf{P} = [p_{s,t}] \in [0,1]^{|S| \times |S|}$ have the property that $\sum_{t \in S} p_{s,t} = 1$ for every state s.

We model class switch recombination with a dependence tree $\bar{T}$ for each lineage tree T with nodes $V(\bar{T}) = V(T)$ and edges $E(\bar{T}) = E(T)$ and isotype labels $\beta(v) = Y_v$ for every node v. In words, the corresponding dependence tree $\bar{T}$ for B cell lineage tree T has the same topology but the dependence tree $\bar{T}$ has a random variable for isotype associated with each node.

For the leaf nodes $v_i \in L(T)$, i.e., the sequenced B cells, we observe isotype b_i directly from scRNA-seq data, i.e., $P(Y_{v_i} = b_i) = 1$. Additionally, root node v_0 in T, represents the naive B cell post V(D)J recombination and therefore $P(Y_{v_0} = 1) = 1$, meaning that root of the dependence tree has isotype state IgM and $\pi_1 = 1$. This means that any dependence tree without $Y_{v_0} = 1$ has zero probability and we omit the initial state probability term. For any other node u in T, we have $\beta(u) = Y_u$. Thus, to compute (5) for isotype labels $\beta(v)$, it suffices to know the isotype transition probabilities $\mathbf{P}$, which we formally define below.

(Main Text) Definition 2. An $r \times r$ matrix $\mathbf{P} = [p_{s,t}]$ is an *isotype transition probability matrix* provided for all isotypes $s, t \in [r]$ it holds that (i) $p_{s,t} = 0$ if $t > s$, (ii) $p_{s,t} \geq 0$, and (iii) $\sum_{t=1}^{r} p_{s,t} = 1$ for all isotypes $s \in [r]$.

The additional conditions on the transition probabilities beyond row stochasticity on the transition probabilities are to properly model the irreversibility of class switch recombination. Using the above model for class switch recombination, we compute the likelihood $\mathrm{CSR}(T, \beta, \mathbf{P})$ for observed isotypes $\mathbf{b}$ given a lineage tree with isotypes β and isotype transition probabilities as follows.

$$\begin{aligned} \mathrm{CSR}(T, \beta, \mathbf{P}) &= \Pr(\mathbf{b} \mid T, \beta, \mathbf{P}) \\ &= \prod_{(u,v) \in E(T)} p_{\beta(u), \beta(v)} \\ &= \prod_{v \in V(T) \setminus \{v_0\}} \prod_{(s,t) \in [r] \times [r]} p_{s,t}^{\mathbf{1}(\beta(v)=t, \beta(\phi(v))=s)} \\ &= \prod_{(s,t) \in [r] \times [r]} p_{s,t}^{N_{s,t}} \end{aligned} \tag{6}$$

where $N_{s,t}$ is the count of occurrences in lineage tree T such that $\beta(v) = t$ and $\beta(\phi(v)) = s$. This is easily extended for a forest of k lineage trees $T_1, \ldots, T_k$ with corresponding isotypes $\beta_1, \ldots, \beta_k$. Given isotype transition probabilities $\mathbf{P}$, the joint probabilities $\mathrm{CSR}(T_1, \beta_1, \mathbf{P}), \ldots, \mathrm{CSR}(T_k, \beta_k, \mathbf{P})$ are conditionally independent, resulting in the joint likelihood

$$\begin{aligned}\prod_{j=1}^{k} \mathrm{CSR}(T_j, \beta_j, \mathbf{P}) &= \prod_{j=1}^{k} \Pr(\mathbf{b}_j \mid T_j, \beta_j, \mathbf{P}) \\ &= \prod_{j=1}^{k} \prod_{(u,v) \in E(T_j)} p_{\beta_j(u), \beta_j(v)} \\ &= \prod_{j=1}^{k} \prod_{v \in V(T_j) \setminus \{v_0\}} \prod_{(s,t) \in [r] \times [r]} p_{s,t}^{\mathbf{1}(\beta_j(v)=t, \beta_j(\phi(v))=s)} \\ &= \prod_{(s,t) \in [r] \times [r]} p_{s,t}^{\sum_{j=1}^{k} N_{j,s,t}}\end{aligned} \tag{7}$$

where $N_{j,s,t}$ is the count of occurrences in lineage tree T_j such that $\beta(v) = t$ and $\beta(\phi(v)) = s$.

B Combinatorial characterization and complexity results

B.1 B cell lineage forest inference

Recall the B CELL LINEAGE FOREST INFERENCE PROBLEM (BLFI) from Sec. 2, restated below for convenience.

(Main Text) Problem 1 (B CELL LINEAGE FOREST INFERENCE (BLFI))**.** Given MSAs $\mathbf{A}_1, \ldots, \mathbf{A}_k$ and isotypes $\mathbf{b}_1, \ldots, \mathbf{b}_k$ for k clonotypes, find isotype transition probabilities $\mathbf{P}^*$ for r isotypes and lineage trees $T_1^*, \ldots, T_k^*$ for $(\mathbf{A}_1, \mathbf{b}_1), \ldots, (\mathbf{A}_k, \mathbf{b}_k)$ whose nodes are labeled by sequences $\alpha_1^*, \ldots, \alpha_k^*$ and isotypes $\beta_1^*, \ldots, \beta_k^*$, respectively, such that $\sum_{j=1}^{k} \mathrm{SHM}(T_j^*, \alpha_j^*)$ is minimum and then $\prod_{j=1}^{k} \mathrm{CSR}(T_j^*, \beta_j^*, \mathbf{P}^*)$ is maximum.

Theorem 1. The BLFI problem is NP-hard even if $k = 1$ and $r = 1$.

We prove that the BLFI problem is NP-hard via a simple reduction from the LARGE PARSIMONY problem [41]. Although this problem is well known, we restate it here for completeness.

Problem 3 (LARGE PARSIMONY (LP))**.** Given a matrix $\mathbf{A} \in \{0,1\}^{n \times m}$, find a rooted tree T whose nodes are labeled by sequences $\alpha : V(T) \to \{0,1\}^m$ such that the n leaves are labeled by the rows of $\mathbf{A}$ and $\sum_{(u,v) \in E(T)} D(\alpha(u), \alpha(v))$ is minimium.

The reduction to BLFI proceeds by using the same MSA $\mathbf{A}$ directly for a single clonotype, i.e., $k = 1$. Additionally, we restrict the number r of isotypes to 1, and set isotypes $\mathbf{b} = [1]^n$.

Lemma 1. Tree T and node labeling α form an optimal solution to LP instance $\mathbf{A}$ if and only if tree T, sequences α and isotypes β, the isotype transition probabilities $\mathbf{P}$ form an optimal solution to BLFI instance $(\mathbf{A}, \mathbf{b})$.

Proof. ($\Rightarrow$) Let tree T and sequence labeling α be an optimal solution to the LP problem. We will show that T and α can be augmented to form an optimal solution to the corresponding BLFI problem. We set $\mathbf{P} = [1]$. We also set $\beta(v) = 1$ for all nodes $v \in T$. We claim that $(T, \alpha, \beta, \mathbf{P})$ form an optimal solution to BLFI. Assume for a contradiction there exists a solution $(T', \alpha', \beta', \mathbf{P}')$ such that $\mathrm{SHM}(T', \alpha') < \mathrm{SHM}(T, \alpha)$, or $\mathrm{SHM}(T', \alpha') = \mathrm{SHM}(T, \alpha)$ and $\mathrm{CSR}(T', \beta', \mathbf{P}') > \mathrm{CSR}(T, \beta, \mathbf{P})$. Clearly, any feasible solution to BLFI must use $\beta(v) = 1$ for all nodes v and $\mathbf{P} = [1]$ as $r = 1$. This means that any feasible solution to BLFI will have a CSR objective value of 1. Therefore, $\mathrm{CSR}(T', \beta', \mathbf{P}') = \mathrm{CSR}(T, \beta, \mathbf{P}) = 1$. Hence, $\mathrm{SHM}(T', \alpha') < \mathrm{SHM}(T, \alpha)$. As can be seen in (1), the SHM objective equals the objective of the LP problem. Therefore, T' and α' have a lower parsimony score than T and α, a contradiction.

($\Leftarrow$) Let $(T, \alpha, \beta, \mathbf{P})$ be an optimal solution to BLFI. Again, as the SHM objective equals the objective of the LP problem, it directly follows that (T, α) form an optimal solution to the LP problem instance. □

B.2 Most parsimonious tree refinement

B.2.1 Combinatorial characterization

Recall from the main text the definition of isotype transition probabilities $\mathbf{P}$, the CSR log-likelihood for isotypes $\mathbf{b}$ of a tree T with nodes labeled by isotypes β, and the MOST PARSIMONIOUS TREE REFINEMENT problem, provided below for convenience.

(Main Text) Definition 2. An $r \times r$ matrix $\mathbf{P} = [p_{s,t}]$ is an *isotype transition probability matrix* provided for all isotypes $s, t \in [r]$ it holds that (i) $p_{s,t} \geq 0$, (ii) $p_{s,t} = 0$ if $s > t$, and (iii) $\sum_{t=1}^{r} p_{s,t} = 1$ for all isotypes $s \in [r]$.

$$\log \mathrm{CSR}(T, \beta, \mathbf{P}) = \log \prod_{(u,v) \in E(T)} p_{\beta(u),\beta(v)} = \sum_{(u,v) \in E(T)} \log p_{\beta(u),\beta(v)}.$$

(Main Text) Problem 2 (MOST PARSIMONIOUS TREE REFINEMENT (MPTR))**.** Given a tree T on n leaves, isotypes $\mathbf{b} = [b_0, \ldots, b_n]$ and isotype transition probabilities $\mathbf{P}$, find a tree T' with root v'_0 and isotype labels $\beta' : V(T') \to [r]$ such that (i) T' is a refinement of T, (ii) $\beta'(v'_0) = b_0 = 1$, (iii) $\beta'(v'_i) = b_i$ for each leaf $v'_i \in \{v'_1, \ldots, v'_n\}$ and (iv) $\log \mathrm{CSR}(T', \beta', \mathbf{P})$ is maximum.

Let σ be a mapping from $V(T')$ to $V(T)$ that reverses all EXPAND operations of each node u' in refinement T' in order to obtain back the node $\sigma(u') = u$ from which it was derived in the original tree T. We say that an isotype labeling $\beta' : V(T') \to [r]$ of T' is transitory if along each directed edge (u', v') of T' either the isotype changes or u' and v' correspond to two distinct nodes of T. More formally, we have the following definition.

Definition 3. Let T' be a refinement of a tree T whose leaves are labeled by isotypes $\mathbf{b}$. Then, an isotype labeling β' of T' is *transitory* provided (i) $\beta'(v'_0) = 1$ where v'_0 is the root of T', (ii) $\beta'(v') = b_{\sigma(v')}$ for each leaf $v' \in L(T')$, (iii) $\beta'(u') \leq \beta(v')$ for each edge (u', v') of T', and (iv) $\beta'(u') = \beta'(v')$ only if $\sigma(u') \neq \sigma(v')$ for each edge (u', v') of T'.

Importantly, among the set of optimal solutions (T', β') to each MPTR problem instance $(T, \mathbf{b}, \mathbf{P})$ there exist solutions where β' is transitory.

Lemma 2. Let $(T, \mathbf{b}, \mathbf{P})$ be an MPTR problem instance. There exist an optimal solution (T', β') where β' is transitory.

Proof. We prove this by contradiction. Let (T', β') be an optimal solution where β' is not transitory. First, observe that it holds that $\beta'(u') \leq \beta(v')$ for each edge (u', v') of T'. To see why, if there were an edge (u', v') such that $\beta'(u') > \beta(v')$ then $\mathrm{CSR}(T', \beta', \mathbf{P}) = -\infty$ as $\log p_{s,t} = -\infty$ if $s > t$. However, setting $\beta'(u') = 1$ for nodes u' would result in log-likelihood greater than $-\infty$. Since (T', β') is a feasible solution to MPTR respecting irreversibility of isotype transitions, it means that condition (iv) of Definition 3 is violated. Let (u', v') be an edge such that $\beta'(u') = \beta'(v')$ and $\sigma(u') = \sigma(v')$. We can contract this edge, retaining the isotype labeling β' for the remaining nodes, such that the resulting tree remains a refinement of T and the objective value remains unchanged as $\log p_{s,s} = 0$. Repeating this procedure for all edges (u', v') such that $\beta'(u') = \beta'(v')$ and $\sigma(u') = \sigma(v')$ results in (T'', β''), where T'' is a refinement of T labeled by β'', with the same optimal score as (T', β'). Clearly, (T'', β'') is transitory, proving the lemma. ☐

B.2.2 Complexity

Note that maximizing the CSR log-likelihood is equivalent to maximizing the CSR likelihood, which is the objective function we will use in this subjection. That is,

$$\mathrm{CSR}(T, \beta, \mathbf{P}) = \prod_{(u,v) \in E(T)} p_{\beta(u),\beta(v)}.$$

We now prove the following theorem.

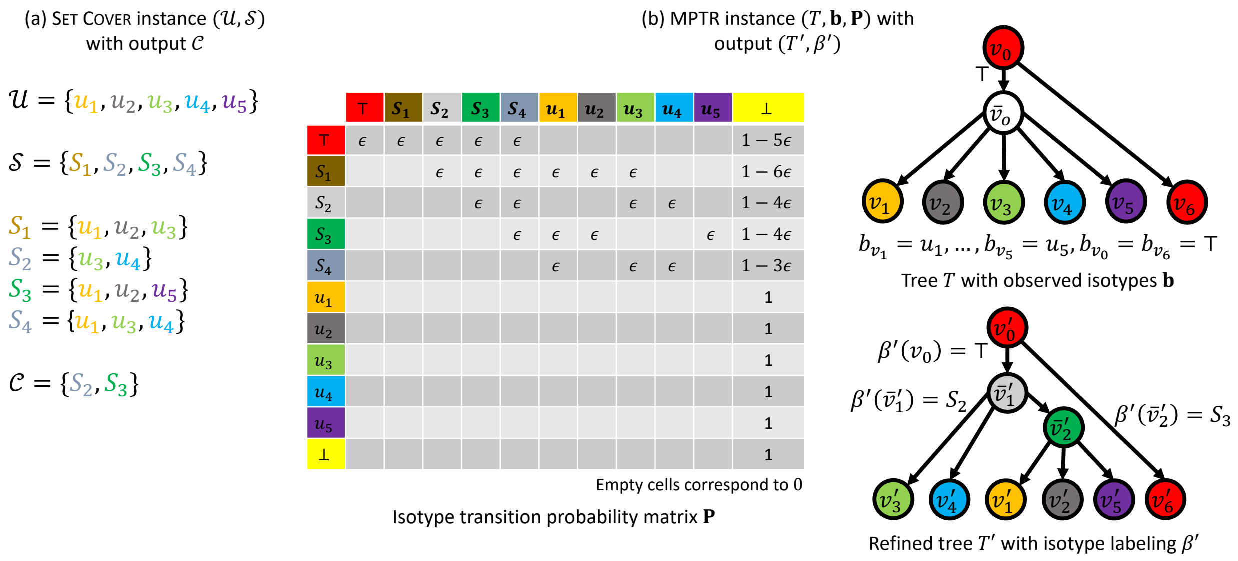
Theorem 2. The MPTR problem is NP-hard.

We show that MPTR is NP-hard by reduction from SET COVER.

Problem 4 (SET COVER)**.** Given a universe $\mathcal{U}$ of elements $\{u_1, \dots u_{|\mathcal{U}|}\}$ and a collection $\mathcal{S}$ of subsets $\{S_1, \dots, S_{|\mathcal{S}|}\}$ such that $\bigcup_{i=1}^{|\mathcal{S}|} S_i = \mathcal{U}$, find a cover $\mathcal{C} \subseteq \mathcal{S}$ such that $\bigcup_{S \in \mathcal{C}} S = \mathcal{U}$ and the size $|\mathcal{C}|$ of the cover is minimum.

Note that while the order of the subsets in collection $\mathcal{S}$ does not matter for SET COVER, our reduction will assume the subsets to be in an arbitrary but fixed order. Similarly, we will assume $\mathcal{U}$ to be ordered arbitrarily. SET COVER has been proven to be NP-hard in Karp's 21 NP-complete problems [42]. We describe a polynomial time reduction from SET COVER to MPTR. To that end, given the set $\mathcal{U}$ of elements and the collection $\mathcal{S}$ of subsets, we construct a tree T with $|\mathcal{U}| + 1$ leaves, $r = |\mathcal{U}| + |\mathcal{S}| + 2$ isotypes, observed isotypes $\mathbf{b} \in [r]^{|\mathcal{U}|+1}$, and $r \times r$ transition probabilities $\mathbf{P}$. The steps are as follows.

1. To construct tree T, we begin by adding the root node v_0. Following that, we attach two children, denoted as $\bar{v}_0$ and $v_{|\mathcal{U}|+1}$, to the root node v_0. Finally, for each element $u_q \in \mathcal{U}$, we add an edge $(\bar{v}_0, v_q)$ in tree T. The constructed tree T has $|\mathcal{U}| + 3$ nodes and $|\mathcal{U}| + 2$ edges.
2. We consider a total of $r = |\mathcal{S}| + |\mathcal{U}| + 2$ isotypes, each corresponding to either a subset $S_i \in \mathcal{S}$, an element $u_q \in \mathcal{U}$, or one of the special symbols $\top$ or $\bot$. Specifically, the first isotype stands for the special symbol $\top$, followed by $|\mathcal{S}|$ isotypes representing each subset $S_i \in \mathcal{S}$, succeeded by $|\mathcal{U}|$ isotypes representing each element $u_q \in \mathcal{U}$, and concluding with the last isotype signifying the special symbol $\bot$. For convenience, we define a function $R : \mathcal{S} \cup \mathcal{U} \cup \{\top, \bot\} \to [r]$ to map the subsets $S_i \in \mathcal{S}$, the elements $u_q \in \mathcal{U}$, and the special symbols $\top$ and $\bot$ to their representative isotype indices as follows.
$$R(X) = \begin{cases} 1, & \text{if } X = \top, \\ i+1, & \text{if } X = S_i, \\ |\mathcal{S}| + q + 1, & \text{if } X = u_q, \\ |\mathcal{S}| + |\mathcal{U}| + 2, & \text{if } X = \bot. \end{cases}$$
3. For the observed isotypes, we set $b_0 = b_{|\mathcal{U}|+1} = R(\top) = 1$, and $b_q = R(u_q)$ for $1 \leq q \leq |\mathcal{U}|$.
4. We define ϵ to be a constant such that $0 < \epsilon \leq 1/(|\mathcal{S}| + |\mathcal{U}| + 1)$. Next, we construct the isotype transition probabilities $\mathbf{P}$ parameterized by ϵ as follows.
   (a) We set the transition probability from $R(\top)$ to $R(\top)$ or $R(S_i)$ for any set $S_i \in \mathcal{S}$ to be ϵ and to $R(u_q)$ for any $u_q \in \mathcal{U}$ to be 0.
$$\begin{aligned} p_{R(\top),R(\top)} &= \epsilon, & \\ p_{R(\top),R(S_i)} &= \epsilon & \forall 1 \leq i \leq |\mathcal{S}|, \\ p_{R(\top),R(u_q)} &= 0 & \forall 1 \leq q \leq |\mathcal{U}|. \end{aligned}$$
   (b) We set the transition probability from $R(\top)$ to $R(\bot)$ to be $1 - (1 + |\mathcal{S}|)\epsilon$.
$$p_{R(\top),R(\bot)} = 1 - (1 + |\mathcal{S}|)\epsilon.$$
   (c) We set the transition probability $p_{R(S_i),R(S_j)}$ for any $S_i, S_j \in \mathcal{S}$ to be ϵ if $i < j$, and 0 otherwise.
$$p_{R(S_i),R(S_j)} = \begin{cases} \epsilon, & \text{if } i < j, \\ 0, & \text{if } i \geq j, \end{cases} \qquad \forall 1 \leq i, j \leq |\mathcal{S}|.$$
   (d) We set the transition probability from $R(S_i)$ to $R(u_q)$ for any set $S_i \in \mathcal{S}$ and any element $u_q \in \mathcal{U}$ to be ϵ if $u_q \in S_i$, and 0 otherwise.
$$p_{R(S_i),R(u_q)} = \begin{cases} \epsilon, & \text{if } u_q \in S_i, \\ 0, & \text{if } u_q \notin S_i, \end{cases} \qquad \forall 1 \leq i \leq |\mathcal{S}|, 1 \leq q \leq |\mathcal{U}|.$$



$\mathcal{U} = \{u_1, u_2, u_3, u_4, u_5\}$

$\mathcal{S} = \{S_1, S_2, S_3, S_4\}$

$S_1 = \{u_1, u_2, u_3\}$
$S_2 = \{u_3, u_4\}$
$S_3 = \{u_1, u_2, u_5\}$
$S_4 = \{u_1, u_3, u_4\}$

$\mathcal{C} = \{S_2, S_3\}$

	$\top$	S_1	S_2	S_3	S_4	u_1	u_2	u_3	u_4	u_5	$\bot$
$\top$	ϵ	ϵ	ϵ	ϵ	ϵ						$1-5\epsilon$
S_1			ϵ	ϵ	ϵ	ϵ	ϵ	ϵ			$1-6\epsilon$
S_2				ϵ	ϵ			ϵ	ϵ		$1-4\epsilon$
S_3					ϵ	ϵ	ϵ			ϵ	$1-4\epsilon$
S_4						ϵ		ϵ	ϵ		$1-3\epsilon$
u_1											1
u_2											1
u_3											1
u_4											1
u_5											1
$\bot$											1

Empty cells correspond to 0

Isotype transition probability matrix $\mathbf{P}$

Figure S1: **Polynomial time reduction from SET COVER to MPTR.** (a) shows a SET COVER instance $(\mathcal{U}, \mathcal{S})$, with the corresponding minimum set cover $\mathcal{C}$. The constructed MPTR instance $(T, \mathbf{b}, \mathbf{P})$, along with the output (T', β') is shown in (b). Isotypes are indicated through colors. The mapping function R is omitted, with the isotypes directly represented by elements, subsets, $\top$, or $\bot$. The empty boxes in the transition probability matrix $\mathbf{P}$ corresponds to 0.

(e) For each $S_i \in \mathcal{S}$, we set the transition probability from $R(S_i)$ to $R(\top)$ to be 0 and to $R(\bot)$ to be $1 - (|\mathcal{S}| - i + |S_i|)\epsilon$.

$$p_{R(S_i),R(\top)} = 0 \qquad 1 \le i \le |\mathcal{S}|,$$
$$p_{R(S_i),R(\bot)} = 1 - (|\mathcal{S}| - i + |S_i|)\epsilon \qquad 1 \le i \le |\mathcal{S}|.$$

(f) For any $u_q \in \mathcal{U}$, we set the transition probability from $R(u_q)$ to any other isotype except $\bot$ to be 0. We set $p_{R(u_q),R(\bot)}$ for any $u_q \in \mathcal{U}$ to be 1.

$$p_{R(u_q),R(X)} = 0 \qquad \forall 1 \le q \le |\mathcal{U}|, X \in \mathcal{S} \cup \mathcal{U} \cup \{\top\},$$
$$p_{R(u_q),R(\bot)} = 1 \qquad \forall 1 \le q \le |\mathcal{U}|.$$

(g) Last, we set the transition probability $p_{R(\bot),R(\bot)}$ to be 1.

$$p_{R(\bot),R(\bot)} = 1$$

Clearly, by construction matrix $\mathbf{P}$ obtained from a SET COVER instance $(\mathcal{U}, \mathcal{S})$ is an isotype transition probability matrix as $\mathbf{P}$ is upper triangular, each entry is non-negative and each row sums to 1. In addition, this reduction takes polynomial time.

To prove hardness, let (T', β') be an optimal solution to the MPTR instance composed of the input tree T, observed isotypes $\mathbf{b}$, and isotype transition probabilities $\mathbf{P}$ corresponding to SET COVER instance $(\mathcal{U}, \mathcal{S})$.

Lemma 3. $\mathrm{CSR}(T', \beta', \mathbf{P}) > 0$ for the refined tree T' and the isotype labeling β' inferred by MPTR.

Proof. We prove this by showing that for any constructed input tree T, observed isotypes $\mathbf{b}$ and isotype transition probabilities $\mathbf{P}$, there exists a refined tree T' and isotype labeling β' such that $\mathrm{CSR}(T', \beta', \mathbf{P}) > 0$. We provide a proof by constructing a refined tree T' with isotype labeling β'. The tree T' will expand the unique polytomous node $\bar{v}_0$ into a chain $\bar{v}'_1 \to \ldots \to \bar{v}'_{|\mathcal{S}|}$. We leave the remaining nodes $v_0, v_1, \ldots, v_{|\mathcal{U}|+1}$ of T unaltered, letting $v'_0, v'_1, \ldots, v'_{|\mathcal{U}|+1}$ denote their corresponding nodes in T'. Next, for each $1 \le q \le |\mathcal{U}|$, we pick a subset S_i such that $u_q \in S_i$, and add edge $(\bar{v}'_i, v'_q)$ in T' and set $\beta'(v'_q) = R(u_q)$. We add the edges $(v'_0, v'_{|\mathcal{U}|+1})$ and $(v'_0, \bar{v}'_1)$. Finally, we set $\beta'(v'_0) = \beta'(v'_{|\mathcal{U}|+1}) = R(\top)$. Clearly all the edges in T' have nonzero isotype transition probabilities, so $\mathrm{CSR}(T', \beta', \mathbf{P}) > 0$. □

Corollary 2. The root v'_0 of T' is labeled by isotype $\top$.

Proof. Due to the presence of leaf $v_{|\mathcal{U}|+1}$ with isotype $b_{|\mathcal{U}|+1} = R(\top)$, the root v'_0 of T' must be labeled by isotype $\beta'(v'_0) = R(\top)$, otherwise there would be a zero-probability edge. □

Corollary 3. No node v' of T' is labeled by isotype $\bot$.

Corollary 4. Each edge (v', v'') of T' has an isotype transition probability of $p_{\beta'(v'),\beta'(v'')} = \epsilon$.

Observe that $\bar{v}_0$ is the only polytomous node in T. We will now prove that $\bar{v}_0$ is the only node of T that is expanded in the refined tree T'.

Lemma 4. Node $\bar{v}_0$ is the only node of T that is expanded in T'.

Proof. By Lemma 2, we may assume that β' is transitory. Let v'_0 be the root of T'. We prove this lemma by contradiction. Let $v \neq \bar{v}_0$ be a distinct node of T that is expanded in T'. We distinguish the following three cases.

- $v = v_{|\mathcal{U}|+1}$: In this case, v equals the leaf node $v_{|\mathcal{U}|+1}$ whose parent is the root v_0. Consider the corresponding node $v'_{|\mathcal{U}|+1}$ of T' such that $\sigma(v'_{|\mathcal{U}|+1}) = v_{|\mathcal{U}|+1}$ and $v'_{|\mathcal{U}|+1}$ is a leaf of T'. Since β' is transitory, we have that $\beta'(v'_0) = \beta'(v'_{|\mathcal{U}|+1}) = R(\top)$. Since node $v_{|\mathcal{U}|+1}$ was expanded, node $v'_{|\mathcal{U}|+1}$ has a unique parent $v''_{|\mathcal{U}|+1} \neq v'_0$. As β' is transitory and $\beta'(v'_{|\mathcal{U}|+1}) = R(\top)$ and $R(\top) \leq s$ for all $s \in [r]$, we must have that $\beta'(v''_{|\mathcal{U}|+1}) = R(\top)$. This, however, implies that β' is not transitory as $\sigma(v''_{|\mathcal{U}|+1}) = \sigma(v'_{|\mathcal{U}|+1}) = v_{|\mathcal{U}|+1}$ and $\beta'(v''_{|\mathcal{U}|+1}) = \beta'(v'_{|\mathcal{U}|+1}) = R(\top)$, which yields a contradiction.

- $v \in \{v_1, \ldots, v_{|\mathcal{U}|}\}$: Note that v is a leaf of T. Consider the corresponding node v' of T' such that $\sigma(v') = v$ and v' is a leaf of T'. The parent of v in T is node $\bar{v}_0$. Since node v was expanded, node v' has a unique parent v'' such that $\sigma(v'') = v$. Let v''' be the unique parent of v''. By Corollary 4, we have that the two edges (v'', v') and (v''', v'') both have probabilities ϵ, contributing a factor of 2ϵ to the overall probability $\mathrm{CSR}(T', \beta', \mathbf{P})$. However, by contracting the edge (v'', v') and removing the node v'', we obtain another solution with higher probability, leading to a contradiction.

- $v = v_0$: Consider the corresponding node v'_0 such that $\sigma(v'_0) = v_0$ and v'_0 is the root of T'. There are two cases two consider. Let v''_0 be a child of v'_0 such that $\sigma(v''_0) = v_0$. We distinguish two cases.
  - First, $\beta'(v'_0) = \beta'(v''_0)$. By Corollary 2, we have that $\beta'(v'_0) = \beta'(v''_0) = R(\top)$. By Corollary 4, we have that the edge (v'_0, v''_0) contributes a factor of ϵ to the overall probability $\mathrm{CSR}(T', \beta', \mathbf{P})$. We can remove this factor by simply contracting the edge (v'_0, v''_0), resulting in a more optimal solution, which is a contradiction.
  - Second, $\beta'(v'_0) \neq \beta'(v''_0)$. By Corollary 2, we have that $\beta'(v'_0) = R(\top)$. By Lemma 3, we have $\beta'(v''_0) \in \{R(S_1), \ldots, R(S_{|\mathcal{S}|})\}$. Again, by the same lemma, all children of v''_0 will be labeled by isotypes different than v''_0. In particular, each child of v''_0 will either correspond to node v_0 or $\bar{v}_0$ of T, labeled from the set $\{R(S_1), \ldots, R(S_{|\mathcal{S}|})\} \setminus \{\beta'(v''_0)\}$. Thus, we may contract the edge (v'_0, v''_0), with probability ϵ, and remove the node v''_0, reassigning all children of v''_0 to v'_0. The resulting tree and isotype labeling will have a larger probability, a contradiction.

□

Assume that a series of EXPAND operations on $\bar{v}_0$ in T has generated k nodes in T', where k ranges from 1 (no EXPAND operation) to $|\mathcal{U}|$. We denote $\bar{v}'_1, \ldots, \bar{v}'_k$ to be the new nodes in T' originating from $\bar{v}_0$ in T, i.e., $\sigma(\bar{v}'_1) = \ldots = \sigma(\bar{v}'_k) = \bar{v}_0$. Let $\bar{T}'$ be the subtree of T' induced by nodes $\bar{v}'_1, \ldots, \bar{v}'_k$.

Lemma 5. The refined tree T' has $|\mathcal{U}| + k + 2$ nodes, $|\mathcal{U}| + k + 1$ edges, and $\mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$.

Proof. Since T has $|\mathcal{U}| + 3$ nodes, and, by Lemma 4, the only node $\bar{v}_0$ of T that is expanded, expands to k nodes $\bar{v}'_1, \ldots, \bar{v}'_k \in V(T')$, the total number of nodes in T' is $|\mathcal{U}| + 2 - 1 + k = |\mathcal{U}| + k + 2$. Similarly, the number of edges in T is $|\mathcal{U}| + 2$, and since $\bar{T}'$ is a tree containing k nodes, it has $k - 1$ edges. So the total number of edges in T' is $|\mathcal{U}| + 2 + k - 1 = |\mathcal{U}| + k + 1$. It follows from Corollary 4 that $\mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$. □

Lemma 6. Nodes $\bar{v}'_1, \ldots, \bar{v}'_k$ of T' are labeled by k distinct isotypes from the set $\{R(S_1), \ldots, R(S_{|\mathcal{S}|})\}$.

Proof. By construction of $\mathbf{P}$, $R(u_q)$ can only be transitioned into from $R(S_i)$ with nonzero probability where $u_q \in S_i$. So if there is an edge $(\bar{v}'_j, v'_q)$ in T' connecting expanded node $\bar{v}'_j$ with leaf v'_q labeled with $R(u'_q)$ then $\beta'(\bar{v}'_j) = S_i$ for some $S_i \in \mathcal{S}$. Using the observation, we begin by showing that each expanded node $\bar{v}'_i$ has at least one child $v'_q \in L(T')$. We do so by contradiction. Suppose the refined tree T' has an expanded node $\bar{v}'_i$ that does not have any leaf $v'_q \in L(T')$ as a child. Without loss of generality, assume that $\bar{v}'_i$ has a child $\bar{v}''_i$, which, in turn, is the parent of a leaf $v'_q \in L(T')$. This means that $\bar{v}''_i$ is labeled with $\beta'(\bar{v}''_i) = R(S_i)$ for some $S_i \in \mathcal{S}$. Since $R(S_i)$ can only be transitioned into from $R(S_j)$, where $j < i$, or $R(\top)$ with nonzero probability, it holds that $\beta(\bar{v}'_i)$ is either $R(S_j)$ where $j < i$ or $R(\top)$. Similarly, the parent of $\bar{v}'_i$ should also be labeled either with $R(S_{j'})$ where $j' < j$ or $R(\top)$. Now we create a new tree T'' by (i) adding the children of $\bar{v}'_i$ as the children of the parent of $\bar{v}'_i$, and (ii) deleting the edge between $\bar{v}'_i$ and its parent. Clearly T'' has nonzero transition probabilities on all the edges, but has one fewer edge than T'. So $\mathrm{CSR}(T'', \beta', \mathbf{P}) < \mathrm{CSR}(T', \beta', \mathbf{P})$, which contradicts with the premise that T' minimizes $\mathrm{CSR}(T', \beta', \mathbf{P})$. So each expanded node $\bar{v}'_j$ is labeled with $R(S_i)$ for some $S_i \in \mathcal{S}$.

It remains to show that the k nodes $\bar{v}'_1, \ldots, \bar{v}'_k$ are labeled by k distinct isotypes from the set $\{R(S_1), \ldots, R(S_{|\mathcal{S}|})\}$. To see why, observe that, by construction of $\mathbf{P}$, the incident nodes of each edge among nodes $\bar{v}'_1, \ldots, \bar{v}'_k$ must be labeled by distinct isotypes from the set $\{R(S_1), \ldots, R(S_{|\mathcal{S}|})\}$, as $p_{R(S_i),R(S_i)} = 0$ for all $S_i \in \mathcal{S}$. □

Lemma 7. There exists an minimum set cover of size k if and only if there is an optimal solution (T', β') such that $\mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$.

Proof. ($\Rightarrow$) Let $\mathcal{C} = \{S^*_1, \ldots, S^*_k\}$ be a set cover of minimum size k. Without loss of generality, we further assume that $R(S^*_i) < R(S^*_{i+1})$ for any $1 \le i \le k-1$. Next, we build a refined tree T' with isotype labeling β' by expanding the node $\bar{v}_0 \in V(T)$ to k nodes $\bar{v}'_1, \ldots, \bar{v}'_k \in V(T')$. More specifically, we replace $\bar{v}_0$ with $\bar{v}'_1, \ldots, \bar{v}'_k \in V(T')$ such that (i) v_0 is connected to $\bar{v}'_1$ by an edge, (ii) there is an edge $(\bar{v}'_i, \bar{v}'_{i+1})$ in T' for each $1 \le i \le k-1$, (iii) $\bar{v}'_i$ is labeled with $R(S^*_i)$, i.e. $\beta'(\bar{v}'_i) = R(S^*_i)$, and (iv) for each child v_q of $\bar{v}_0$ in T, there exists exactly one edge $(\bar{v}'_i, v_q)$ in T' where $u_q \in S^*_i$. Clearly T' is a refinement of tree T, and all the newly added edges have nonzero transition probabilities ϵ. Hence, $\mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$.

All that remains to show is that (T', β') is optimal. We show this by contradiction. Let (T'', β'') be an optimal solution such that $\mathrm{CSR}(T'', \beta'', \mathbf{P}) < \mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$. By Lemma 4, we have that only the node $\bar{v}_0$ of T is expanded in T'' corresponding $\bar{v}''_1, \ldots, \bar{v}''_{k'}$ nodes in T''. Since $\mathrm{CSR}(T'', \beta'', \mathbf{P}) < \mathrm{CSR}(T', \beta', \mathbf{P})$, it must hold that $k'' < k$. By Lemma 6 we have that the k' labels of nodes $\bar{v}''_1, \ldots, \bar{v}''_{k'}$ correspond to k' distinct subsets of $\mathcal{S}$. By Lemma 3, we have that these k' subsets of $\mathcal{S}$ form a cover of the universe $\mathcal{U}$, leading to a contradiction. Hence, (T', β') is optimal.

($\Leftarrow$) Now assume that there exists an optimal solution (T', β') such that $\mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$. Note that the restriction that $\mathrm{CSR}(T', \beta', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k+1}$ is without loss of generality due to Lemma 5. Now according to Lemma 6, there are k expanded nodes in T' labeled with $R(S^*_1), \ldots, R(S^*_k)$. We define $\mathcal{C} = \{S^*_1, \ldots, S^*_k\}$. Now each leaf $v'_q \in L(T')$ labeled with $R(u_q)$ is the child of an expanded node $\bar{v}'_i \in V(T')$ labeled with $R(S^*_i)$. Since $\mathrm{CSR}(T', \beta', \mathbf{P}) > 0$ by Lemma 3, the transition probability from $R(S^*_i)$ to $R(u_q)$ is strictly greater than 0, which means $u_q \in S^*_i$. So every element in $\mathcal{U}$ is covered by one of the subsets from $\mathcal{C}$. So $\mathcal{C}$ is a set cover of size k.

It remains to show that $\mathcal{C}$ is a minimum-size set cover. Assume for a contradiction that there exists a cover $\mathcal{C}' =\subseteq \mathcal{S}$ such that $|\mathcal{C}'| = k' < k = |\mathcal{C}|$. Let $\mathcal{C}' = \{C'_1, \ldots, C'_{k'}\}$ where the subsets follow the same order as in the original reduction to MPTR. We construct a refined tree T'' with isotype labeling β'' corresponding to $\mathcal{C}'$ by expanding the unique polytomous node $\bar{v}_0$ of T into a chain $\bar{v}''_1 \to \ldots \to \bar{v}''_{k'}$, with one node $\bar{v}''_i$ for each subset $C_i \in \mathcal{C}'$ labeled by $\beta''(\bar{v}''_i) = R(C_i)$, and connecting each leaf $v_q \in \{v_1, \ldots, v_{|\mathcal{U}|}\}$ to a single expanded node $\bar{v}''_i$ such that $u_q \in C'_i$. Since $\mathcal{C}'$ is a cover of $\mathcal{U}$, each leaf $v_q \in \{v_1, \ldots, v_{|\mathcal{U}|}\}$ will be connected. Moreover, tree T'' with isotype labeling β'' form a solution to MPTR. Clearly, T'' has $|\mathcal{U}| + k' + 2$ nodes and $|\mathcal{U}| + k' + 1$ edges. Moreover, each edge of T'' has a nonzero isotype transition probability equal to ϵ, so $\mathrm{CSR}(T'', \beta'', \mathbf{P}) = \epsilon^{|\mathcal{U}|+k'+1} < \epsilon^{|\mathcal{U}|+k+1} = \mathrm{CSR}(T', \beta', \mathbf{P})$, a contradiction. □

C Supplementary Methods

C.1 Tree refinement

We solve an instance $(T, \mathbf{b}, \mathbf{P})$ of the MPTR problem (Fig. S2a) by reducing it to a graph problem. Given an instance $(T, \mathbf{b}, \mathbf{P})$ of the MPTR, we construct a directed graph $G_{T,\mathbf{b}}$, called the expansion graph, with nodes $V(G_{T,\mathbf{b}}) \subseteq$

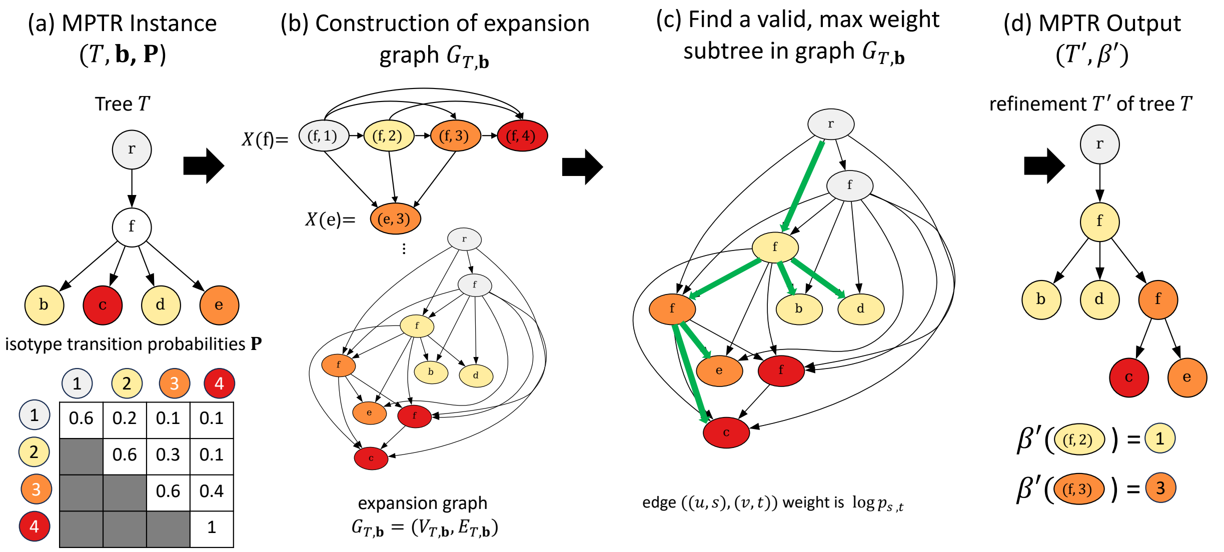


Figure S2: **Algorithm for solving the MPTR problem** (a) An instance $(T, \mathbf{b}, \mathbf{P})$ of the MPTR problem. (b) To construct the expansion graph $G_{T,\mathbf{b}}$ for tree T whose leaves have isotypes $\mathbf{b}$, each original node u in $V(T)$ corresponds to a set $X(u)$ of nodes in $G_{T,\mathbf{b}}$. Edges are added to capture all transitory refinements of tree T. (c) We use the expansion graph $G_{T,\mathbf{b}}$ with weighted edges to find a valid, maximum weight subtree in $G_{T,\mathbf{b}}$, depicted in green. (d) This selected subtree is an optimal solution (T', β') to the MPTR problem instance $(T, \mathbf{b}, \mathbf{P})$.

$V(T) \times [r]$ and edges $E(G_{T,\mathbf{b}})$. At a high level, nodes of $V(G_{T,\mathbf{b}})$ are of the form (u, s) where $u \in V(T)$ is a node of the input tree T and $s \in [r]$ is an isotype state. Formally, we have the following definition.

Definition 4. A directed graph $G_{T,\mathbf{b}}$ is an *expansion graph* of a rooted tree T whose leaves are labeled by isotypes $\mathbf{b}$ provided $V(G_{T,\mathbf{b}}) = \bigcup_{u \in V(T)} X(u)$ where

$$X(u) = \begin{cases} \{(u, b_u)\}, & \text{if } u \in L(T), \\ \{(u, s) \mid s \in \{1, \ldots, \max\{b_v \mid v \in L(T_u)\}\}\}, & \text{if } u \in V(T) \setminus L(T), \end{cases} \tag{8}$$

and $E(G_{T,\mathbf{b}}) = \{((u, s), (v, t)) \mid (u, v) \in E(T), s \leq t\} \cup \{((u, s), (u, t)) \mid u \in V(T), s < t\}$.

In the above definition $X(u)$ is the set of nodes of $G_{T,\mathbf{b}}$ corresponding to node u of T, accounting for the fact that leaves u of T retain their isotype state in any refinement T' of T. On the other hand, internal nodes u of T may be subject to EXPAND operations such that the corresponding nodes of T' are assigned isotypes s ranging from state 1 to the maximum isotype state among all descendant leaves of u in T. The edges of $G_{T,\mathbf{b}}$ respect the irreversibility property of isotypes as well as the parental relationships of nodes of T. See Fig. S2c for an example expansion graph $G_{T,\mathbf{b}}$.

We now define constrained subtrees, termed valid, of the expansion graph $G_{T,\mathbf{b}}$.

Definition 5. A subtree T' of $G_{T,\mathbf{b}}$ is *valid* provided (i) T' is rooted at $(v_0, 1)$ where v_0 is the root of T and (ii) there is a unique edge $((u, s), (v, t))$ in $E(T')$ for each edge (u, v) of T.

We now show that the set of valid subtrees of $G_{T,\mathbf{b}}$ corresponding to trees T' with isotype labelings β' is equivalent to the set composed of pairs (T', β') where T' is a refinement of T and β' is a transitory isotype labeling of T'.

Lemma 8. Let T' be a refinement of T whose leaves are labeled by isotypes $\mathbf{b}$ and let β' be an isotype labeling of T'. Then, β' is transitory if and only if (T', β') induces a valid subtree of $G_{T,\mathbf{b}}$.

Proof. ($\Rightarrow$) Let β' be a transitory isotype labeling of T'. We start by showing that (T', β') induce a connected subtree of $G_{T,\mathbf{b}}$. First, let u' be a node of T' labeled by isotype $\beta(u')$. We claim that $(u', \beta(u')) \in X(u)$. We distinguish the two cases. First, $u' \in L(T')$. Let $u = \sigma(u')$ be the original leaf node u of T. Since β' is transitory, we have $\beta(u') = b_{\sigma(u')} = b_u$. Hence, $(u', \beta(u')) \in X(u)$ for each leaf node $u' \in L(T')$. Second, $u' \in V(T') \setminus L(T')$. Let

$u = \sigma(u')$ be the original internal node u of T. Suppose for a contradiction $(u', \beta'(u')) \notin X(u))$. This means that $\beta'(u') > \max\{b_v \in L(T_u)\}$. As such, there would be an edge (u'', v'') such that $\beta'(u'') > \beta'(v'')$ where u'' is a node in the subtree $T'_{u'}$ rooted at node u'. However, this would mean that β' would violate condition (iii) of Definition 3, a contradiction. Thus, $(u', \beta(u')) \in X(u)$ for each internal node $u' \in V(T' \setminus L(T')$. Hence, $(u', \beta(u')) \in V(G_{T,\mathbf{b}})$.

We now prove that each edge (u', v') of T' whose incident nodes are labeled by $(\beta'(u'), \beta'(v'))$ corresponds to an edge $((u', \beta'(u')), (v', \beta'(v')))$ of $G_{T,\mathbf{b}}$. This follows directly from conditions (iii) and (iv) of Definition 3 and the definition of $E(G_{T,\mathbf{b}})$ in Definition 4. This implies that the subgraph of $G_{T,\mathbf{b}}$ induced by (T', β') is a (connected) subtree of $G_{T,\mathbf{b}}$.

We now must show that this induced subtree of $G_{T,\mathbf{b}}$ is valid. By condition (i) of Definition 3, we have that $\beta'(v'_0) = 1$ for the root v'_0 of T'. As such, the induced subtree of $G_{T,\mathbf{b}}$ is rooted at $(v'_0, 1)$. Finally, we must show there is a unique edge $((u, s), (v, t))$ in the induced subtree of $G_{T,\mathbf{b}}$ for each original edge (u, v) of T. This follows from the fact that T' is a refinement of T. Thus the subgraph of $G_{T,\mathbf{b}}$ induced by (T', β') is a valid subtree of $G_{T,\mathbf{b}}$.

$(\Leftarrow)$ Consider a valid subtree of $G_{T,\mathbf{b}}$, resulting in a tree T' and isotype labeling β'. To see why T' is a refinement of T, observe that edges $((u, s), (u, t))$ correspond to an EXPAND operation on node u of T. It remains to show that β' is transitory. By condition (i) of Definition 5, we have that the root of T' is labeled by state 1, satisfying condition (i) of Definition 3. Conditions (ii) and (iii) of Definition 3 are met by construction of $G_{T,\mathbf{b}}$. Finally, condition (iv) of Definition 3 follows from condition (ii) of Definition 5. Hence, the isotype labeling β' of T' is transitory. □

The following key proposition follows from the previous two lemmas.

Proposition 2. Let $G_{T,\mathbf{b}}$ be an expansion graph of a rooted tree T whose leaves are labeled by isotypes $\mathbf{b}$. Then, given isotype transition probablities $\mathbf{P}$, a valid subtree (T', β') of $G_{T,\mathbf{b}}$ maximizing $\sum_{(u',v') \in E(T')} \log p_{\beta'(u'),\beta'(v')}$ is an optimal solution to MPTR instance $(T, \mathbf{b}, \mathbf{P})$.

To find such a valid subtree with maximum log-likelihood, we formulate the following MILP based on a multi-commodity flow formulation for modeling connectivity. We make use of two sets of decision variables. The first is $f^{t}_{(u,s),(v,t)} \in \mathbb{R}_{\geq 0}$, which represents the amount of flow on edge (u, v) designated for sink $q \in L(T)$. The second is $x_{(u,s),(v,t)} \in \{0, 1\}$, which indicates if edge (u, v) has non-zero flow.

$$\min \sum_{((u,s),(v,t)) \in E(G_{T,\mathbf{b}})} x_{(u,s),(v,t)} \log p_{s,t} \tag{9}$$

s.t.

$$\sum_{(v,t) \in \eta^{+}((u,s))} f^{q}_{(u,s),(v,t)} = \sum_{(v,t) \in \eta^{-}((u,s))} f^{q}_{(v,t),(u,s)}, \qquad \forall (v, s) \in V(G_{T,\mathbf{b}}) \setminus \{(v_0, 1)\}, v \notin L(T), q \in L(T), \tag{10}$$

$$\sum_{(u,s) \in \eta^{-}((q,b_q))} f^{q}_{(u,s),(q,b_q)} = 1, \qquad \forall q \in L(T), \tag{11}$$

$$\sum_{(v,t) \in \eta^{+}((v_0,1))} f^{q}_{(v_0,1),(v,t)} = 1, \qquad \forall q \in L(T), \tag{12}$$

$$f^{q}_{(u,s),(v,t)} \leq x_{(u,s),(v,t)}, \qquad \forall q \in L(T), ((u, s), (v, t)) \in E(G_{T,\mathbf{b}}), \tag{13}$$

$$\sum_{(u,s) \in X(u)} \sum_{(v,t) \in X(v)} x_{(u,s),(v,t)} = 1, \qquad \forall (u, v) \in E(T), \tag{14}$$

$$0 \leq f^{q}_{(u,s),(v,t)} \leq 1, \qquad \forall q \in L(T), ((u, s), (v, t)) \in E(G_{T,\mathbf{b}}), \tag{15}$$

$$x_{(u,s),(v,t)} \in \{0, 1\}, \qquad \forall ((u, s), (v, t)) \in E(G_{T,\mathbf{b}}), \tag{16}$$

where $\eta^{+}((u, s))$ is the set of direct successors of node (u, s) in graph $E(G_{T,\mathbf{b}})$ and $\eta^{-}((u, s))$ is the set of direct predecessors of node (u, s).

Constraints (10), (11), (12) enforce flow conversation and ensure that each terminal receives one unit of flow. Below is a description of each of the above constraints. Constraint (13) links the flow variables to the choice of edges in the resulting refinement. Finally, constraint (14) ensures that refined tree T' can be obtained from tree T via a series of EXPAND operations.

C.2 Maximum likelihood estimate of isotype transition probabilities

Given a forest $T_1, \dots, T_k$ of lineage trees correspondingly labeled by isotypes $\beta_1, \dots, \beta_k$, we seek the maximum likelihood estimate of isotype transition probabilities $\mathbf{P}^*$.

$$\mathbf{P}^* = \arg\max_{\mathbf{P}} \prod_{(s,t)\in[r]\times[r]} p_{s,t}^{\sum_{j=1}^{k} N_{j,s,t}} \tag{17}$$

subject to

$$\sum_{t\in[r]} p_{s,t} = 1, \qquad \forall s \in [r]. \tag{18}$$

We solve this constrained optimization problem using Lagrange multipliers λ_s for each state s. We first take the log of likelihood $\prod_{j=1}^{k} \mathrm{CSR}(T_j, \beta_j, \mathbf{P})$ with respect to isotype transition probabilities $\mathbf{P}$.

$$\begin{aligned} \log \prod_{j=1}^{k} \mathrm{CSR}(T_j, \beta_j, \mathbf{P}) &= \log \prod_{(s,t)\in[r]\times[r]} p_{s,t}^{\sum_{j=1}^{k} N_{j,s,t}} \\ &= \sum_{(s,t)\in[r]\times[r]} \left(\sum_{j=1}^{k} N_{j,s,t}\right) \log p_{s,t}. \end{aligned} \tag{19}$$

To our log-likelihood, we add the term $\lambda_s \left(\sum_{s\in[r]} p_{s,t} - 1\right)$ for each isotype s, resulting in new objective

$$\mathcal{L}(\mathbf{P}, \lambda_1, \dots, \lambda_r) = \left[\sum_{(s,t)\in[r]\times[r]} \left(\sum_{j=1}^{k} N_{j,s,t}\right) \log p_{s,t} + \sum_{s\in[r]} \lambda_s \left(\sum_{t\in[r]} p_{s,t} - 1\right)\right] \tag{20}$$

Then, we set the partial derivative of $\mathcal{L}(\mathbf{P}, \lambda_1, \dots, \lambda_r)$ with respect to each parameter $p_{s,t}$ and λ_s and solve the resulting system of equations. For each λ_s, we obtain our constraint,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \lambda_s} &= 0 = \left(\sum_{t\in[r]} p_{s,t} - 1\right) \\ \sum_{t\in[r]} p_{s,t} &= 1 \end{aligned} \tag{21}$$

For each parameter $p_{s,t}$, we set the partial derivative to 0 and solve for $p_{s,t}$ as a function of λ_s.

$$\begin{aligned} \frac{\partial \mathcal{L}}{p_{s,t}} &= 0 = \frac{\sum_{j=1}^{k} N_{j,s,t}}{p_{s,t}} - \lambda_s \\ \lambda_s &= \frac{\sum_{j=1}^{k} N_{j,s,t}}{p_{s,t}} \\ p_{s,t} &= \frac{\sum_{j=1}^{k} N_{j,s,t}}{\lambda_s} \end{aligned}$$

Given the constraint (21), we have that

$$\sum_{t\in[r]} p_{s,t} = \frac{\sum_{t\in[r]} \sum_{j=1}^{k} N_{j,s,t}}{\lambda_s} = 1, \tag{22}$$

and

$$\lambda_s = \sum_{t \in [r]} \sum_{j=1}^{k} N_{j,s,t}.$$

This yields the following maximum likelihood estimate $p^*_{s,t}$,

$$p^*_{s,t} = \frac{\sum_{j=1}^{k} N_{j,s,t}}{\sum_{t \in [r]} \sum_{j=1}^{k} N_{j,s,t}}$$

Lastly, to account for unobserved isotype transitions where isotype $s \leq t$, we add a pseudocount of 1, resulting in updated isotype transition probabilities

$$p^*_{s,t} = \frac{\sum_{j=1}^{k} N_{j,s,t} + 1}{\sum_{t \in [r]} \left(\sum_{j=1}^{k} N_{j,s,t} + 1 \right)}. \tag{23}$$

In the main text, we additionally use the shorthand $N_{s,t} = \sum_{j=1}^{k} N_{j,s,t}$.

D Simulation details

We designed *in silico* experiments to evaluate TRIBAL with known ground-truth isotype transition probabilities $\mathbf{P}$ and lineage trees T labeled by sequences α and isotypes β. Specifically, we used an existing BCR phylogenetic simulator [13] that models SHM)(Appendix D.1 but not CSR. We generated isotype transition probabilities $\mathbf{P}$ with $r = 7$ isotypes (as in mice) under two different models of CSR (Appendix D.2). Briefly, both CSR models assume the probability of not transitioning is higher than the probability of transitioning, but in the *sequential model* there is clear preference for transitions to the next contiguous isotype, while in the *direct model* the probabilities of contiguous and non-contiguous class are similar (Fig. S3) Given $\mathbf{P}$, we evolved isotype characters down each ground truth lineage tree T.

We generated 5 replications of each CSR model for $k = 75$ clonotypes and $n \in \{35, 65\}$ cells per clonotype, resulting in 20 *in silico* experiments, yielding a total of 1500 ground truth lineage trees. We generated our *in silico* experiments to evaluate all aspects of TRIBAL while benchmarking against existing methods including dnapars [7], dnaml [7] and IgPhyML [10].

D.1 SHM simulation and benchmarking

The Davidsen and Matsen SHM simulator models the generation of B cell lineage trees via a Poisson branching process with selection towards BCRs with increased affinity [13]. We used the provided Docker Hub image container [1] to generate our ground truth B cell lineage trees T and sequence labels α. In addition, we used the provided benchmarking pipeline to run dnapars [7], dnaml [7] and IgPhyML [10]. Below is the command to generate our *in silico* experiments for $n \in \{35, 65\}$ cells and $k = 75$ clonotypes and run comparison methods.

```
simulate
    --igphyml
    --dnapars
    --dnaml
    --selection
    --target_dist=5
    --target_count=100
    --carry_cap=1000
    --T=35
    --lambda=2.0
    --lambda0=0.365
    --n={n}
    --nsim={k}
    --random_naive=sequence_data/AbPair_naive_seqs.fa
```

D.2 CSR simulation

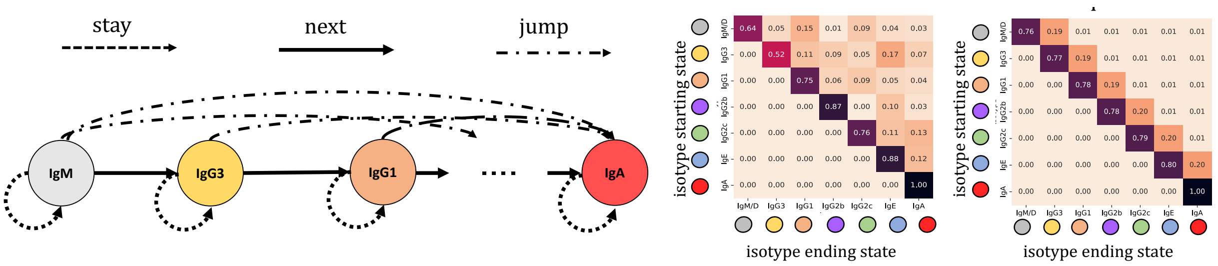
After generating each ground truth B cell lineage tree T as described above, we then evolved isotype characters down each tree T using two different models for class switch recombination to obtain ground truth isotypes β. First, we describe the two different CSR models that we used to generate ground truth isotype transition probabilities $\mathbf{P}$. Then, we describe the generation of these isotype transition probability matrices under these two models.

We grouped each isotype transition probability $p_{s,t}$ where $s <= t$ into one of three categories: (i) *stay*, (ii) *next*, and (iii) *jump*(Fig S3a). In *stay*, the B cell does not undergo any class switching and the isotype does not change. In *next*, a B cell class switches to the next contiguous heavy chain locus. In *jump*, the B cell class switches by jumping to an isotype heavy chain constant locus that is not contiguous.

Next, we describe how we generated ground truth isotype transition probabilities $\mathbf{P}$ under both direct and sequential CSR models. To simulate isotype transition probabilities with direct switching, we randomly sampled a probability of transitioning $1 - \theta \in \{0.1, 0.15, \ldots, 0.35\}$. We then set the initial isotype transition probabilities as

---

[1] krdav/bcr-phylo-benchmark



(a) class switch category (b) direct model (c) sequential model

Figure S3: **Class switch recombination models for *in silico* experiments**. a) Examples of different isotype transition probability parameter groups. (b) Examples of simulated isotype transition probabilities **P** for the direct model of CSR. In the direct model, when a B cell class switches is no systematic preference for transition to the *next* sequential state or *jumping* to a non-contiguous isotype. (c) In the sequential model, a B cell undergoing CSR has a strong affinity for the *next* contiguous heavy chain locus.

$$p'_{s,t} = \begin{cases} 0, & \text{if } s > t, \\ \min(\theta + \epsilon, \tau), & \text{if } s = t, \\ \min(\frac{1-\theta}{r-s} + \epsilon, \tau), & \text{if } s < t, \end{cases}$$

where we add Gaussian noise $\epsilon \sim \mathcal{N}(\mu, \sigma)$ with mean $\mu = 0.05$ and standard deviation $\sigma = 0.025$ to each parameter. To avoid negative transition probabilities we set $\tau = 0.01$. Fig. S3b shows an example of a simulated isotype transition probability matrix under the direct CSR model.

$$p'_{s,t} = \begin{cases} 0, & \text{if } s > t, \\ \min(\theta + \epsilon, \tau), & \text{if } s = t, \\ \min(1 - \theta + \epsilon, \tau), & \text{if } t = s + 1, \\ \tau, & \text{otherwise.} \end{cases}$$

We then set parameter $p'_{s,t} := p'_{s,t} / \sum_{s \in [r]} p'_{s,t}$ to ensure each row in the isotype transition probability matrix **P** sums to 1. Fig. S3b shows an example of a simulated isotype transition probability matrix under the direct model. Fig. S3c shows an example of a simulated isotype transition probability matrix under the sequential CSR model.

D.3 Inference using TRIBAL

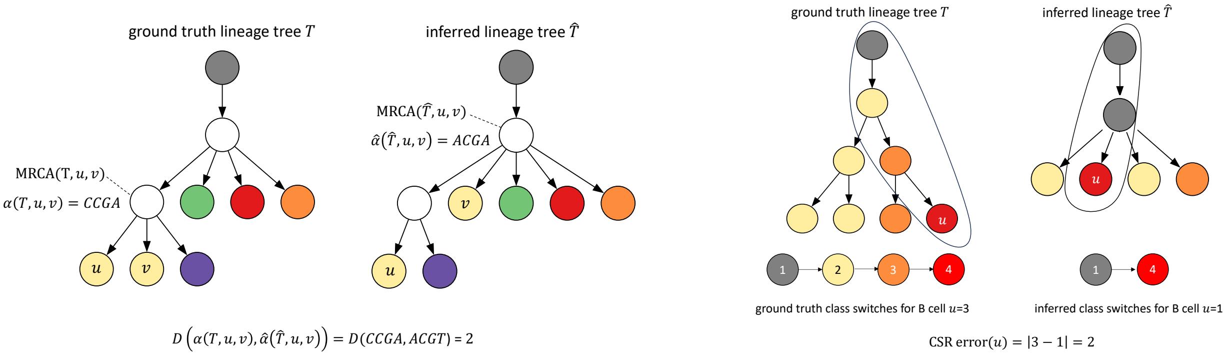
We ran TRIBAL in two ways, referred to as TRIBAL and TRIBAL-NO REFINEMENT, in order to assess the importance of the tree refinement stage of our algorithm. As the naming convention implies, the main difference between TRIBAL and TRIBAL-NO REFINEMENT, is that in TRIBAL-NO REFINEMENT the input trees are not refined and the isotypes $\hat{\beta}$ are inferred using the Sankoff [29] algorithm with weights $w_{s,t}^{(\ell)} = -\log p_{s,t}^{(\ell)}$. All other steps of TRIBAL algorithm remain the same.

Due to large input sets $\mathcal{T}_j$ for some simulated clonotypes j, we sample 50 trees from $\mathcal{T}_j$ for consideration of candidate lineage tree $T_j^{(\ell)}$ within each iteration ℓ. We additionally include the previous optimal lineage tree $T^{(\ell-1)}$ of iteration $\ell - 1$ in the sampled trees for each clonotype j to ensure convergence.

We used a convergence threshold of 0.5 and a maximum of 10 iterations per restart. A total of 5 restarts were performed by iterating through $\theta \in \{0.55, 0.65, 0.75, 0.85, 0.95\}$ for each restart.

D.4 Performance metrics

To evaluate performance of isotype transition probability inference, $\hat{\mathbf{P}}$, we utilized *Kullback-Leibler (KL) divergence*. To assess accuracy of lineage tree inference, we used normalized Robinson-Foulds (RF) distance to assess accuracy of the topology of the inferred tree $\hat{T}$, most recent common ancestor (MRCA) distance to assess accuracy of the inferred sequences $\hat{\alpha}$, and Class Switch Recombination (CSR) error to assess accuracy of the inferred isotypes $\hat{\beta}$.



(a) MRCA distance (b) CSR error

Figure S4: **Performance metrics for B cell lineage tree inference.** (a) An example calculation for MRCA distance leaves u and v. (b) An example calculation of CSR error for lineage u.

Kullback-Leibler (KL) divergence. To evaluate accuracy of isotype transition probability inference, we used *Kullback–Leibler (KL) divergence* [27] to compare the inferred transition probability distribution $\hat{\mathbf{p}}_s$ of each isotype s to the simulated ground truth distribution $\mathbf{p}_s$. KL divergence D_{KL} is defined as

$$D_{\mathrm{KL}}(\hat{\mathbf{p}}_s || \mathbf{p}_s) = \sum_{q \in [r]} \hat{p}_{s,t} \log(\hat{p}_{s,t} / p_{s,t}) \tag{24}$$

The lower the KL divergence, the more similar the two distributions.

Normalized Robinson-Foulds (RF) distance. To assess the accuracy of topology of the inferred B cell lineage tree $\hat{T}$ with respect to simulated ground truth tree T, we used *normalized Robinson-Foulds (RF) distance*. For this metric, we treat both trees as unrooted. For an unrooted tree, if you remove an edge (but not its endpoints), it defines a bipartition of the leaf set [43]. Doing this for every edge in tree T yields a set $B(T)$ of bipartitions. RF distance is defined as the size of the symmetric difference between bipartitions $B(T)$ and $B(\hat{T})$ [28]. We then normalize this by the total number of bipartitions in each tree. Thus, normalized RF is computed as follows

$$\text{normalized RF}(T, \hat{T}) = \frac{|B(T) \triangle B(\hat{T})|}{|B(T)| + |B(\hat{T})|}. \tag{25}$$

Most Recent Common Ancestor (MRCA) distance. To assess the accuracy of the inferred ancestral sequence reconstruction $\hat{\alpha}$ with respect to simulated ground truth α, we used a metric called *Most Recent Common Ancestor (MRCA) distanc*e introduced by Davidsen and Matsen [13]. For any two simulated B cells (leaves), the MRCA distance is the Hamming distance between the MRCA sequences of these two B cells in both the ground truth and inferred lineage trees. This distance is then averaged over all pairs of simulated B cells. A graphical depiction of this metric is show in Fig. S4a.

More formally,

$$\text{MRCA distance}(\alpha, \hat{\alpha}) = \frac{2}{n(n-1)m} \sum_{u,v \in L(T)} D(\hat{\alpha}(\hat{T}, u, v), \alpha(T, u, v)), \tag{26}$$

where in a slight abuse of notation $\alpha(T, u, v)$ is the sequence of the most recent common ancestor (MRCA) of nodes u and v in lineage tree T and m is the length of MSA.

Class switch recombination (CSR) error. We assessed the accuracy of isotype inference by a new metric called *CSR error*, which is computed for each B cell i and clonotype j and is the absolute difference between the number of ground-truth class switches and inferred number of class switches that occurred along its evolutionary path from the root (Fig. S4b). Since dnaml, dnapars and IgPhyML do not infer isotypes for internal nodes, we pair these methods with the Sankoff algorithm [29] using $w_{s,t}$ equals 1 if $s = t$, 0 if $s < t$ and ∞ otherwise.

E Supplementary Results

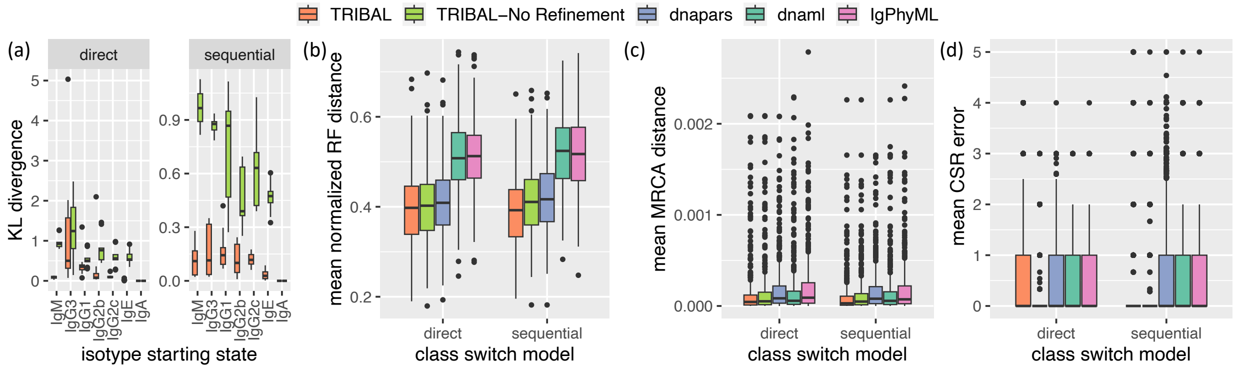
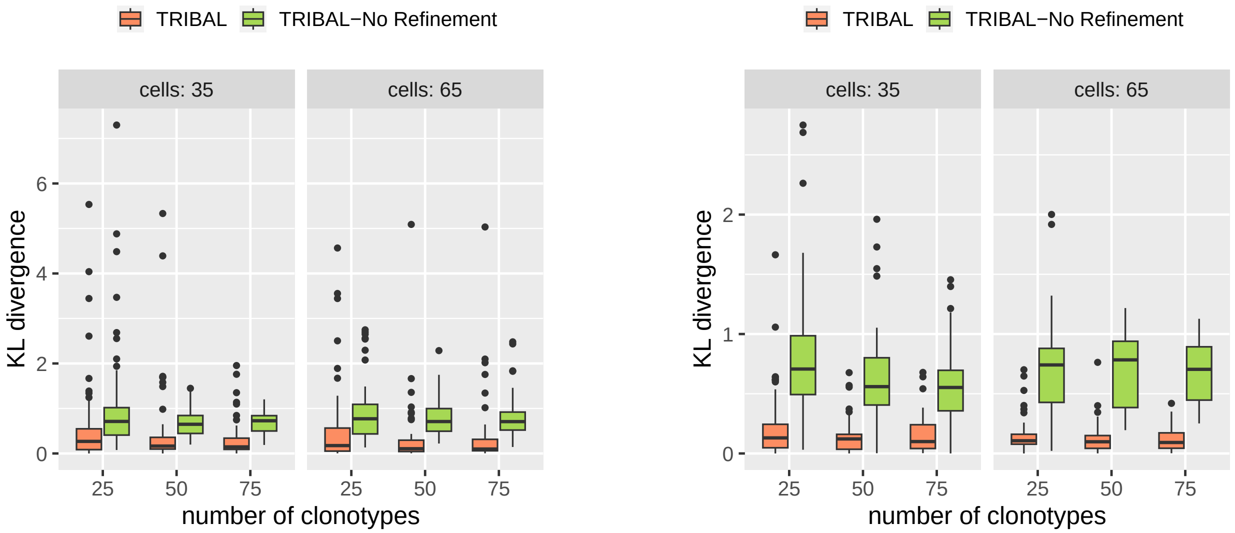


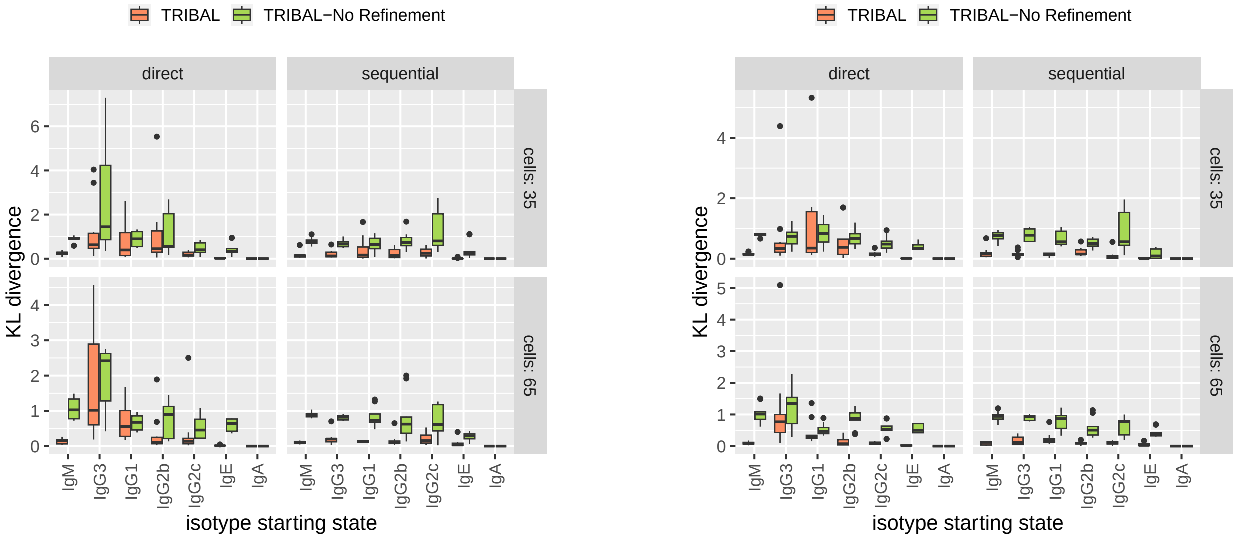
Figure S5: **Simulations results for $k = 75$ clonotypes and $n = 65$ cells per clonotype.**



(a) direct CSR model

(b) sequential CSR model

Figure S6: **KL divergence from ground truth isotype transition probabilities aggregated over all isotype starting states, except IgA, by isotype starting state with varying the number k clonotypes, the number n of cells and CSR model.**



(a) $k = 25$ clonotypes

(b) $k = 50$ clonotypes

Figure S7: **KL divergence from ground truth isotype transition probabilities by isotype starting state for $k \in \{25, 50\}$ clonotypes**

dataset	clonotypes k	total cells n	median cells per clonotype	max cells per clonotype	median distinct isotypes per clonotype
NP-KLH-1	167	1776	7	89	3
NP-KLH-2a	70	537	6	32	2
NP-KLH-2b	58	357	5	21	2

Table S1: **Summary of NP-KLH mouse scRNA-seq datasets**

E.1 Average clade entropy for a leaf labeling

We describe a metric used to assess the average entropy contained within a leaf-labeling of the clades of a tree. First, we introduce some notation. Let Σ be an alphabet. Let clade u of tree T be the subtree T_u rooted at node u. Let $\delta(u) \subseteq L(T)$ be the subset of leaves that are descendants of node u. Let $\ell : L(T) \to \Sigma$ be a leaf labeling. Given a clade u and leaf-labeling ℓ, the entropy of a clade with respect to its leaf labels is defined as

$$H(u, \ell) = -\sum_{s \in [r]} p(s) \log p(s), \tag{27}$$

where $p(s) = \sum_{v \in \delta(u)} \mathbf{1}(\ell(v) = s)/|\delta(u)|$. The average clade entropy $\bar{H}$ is computed over all clades except the leaves $L(T)$ and the root r as follows

$$\bar{H}(T, \ell) = \frac{\sum_{u \in \bar{V}} H(u, \ell)}{|\bar{V}|}, \tag{28}$$

where $\bar{V} = V(T) \setminus (\{r\} \cup L(T))$ is the set of non-trivial clades.

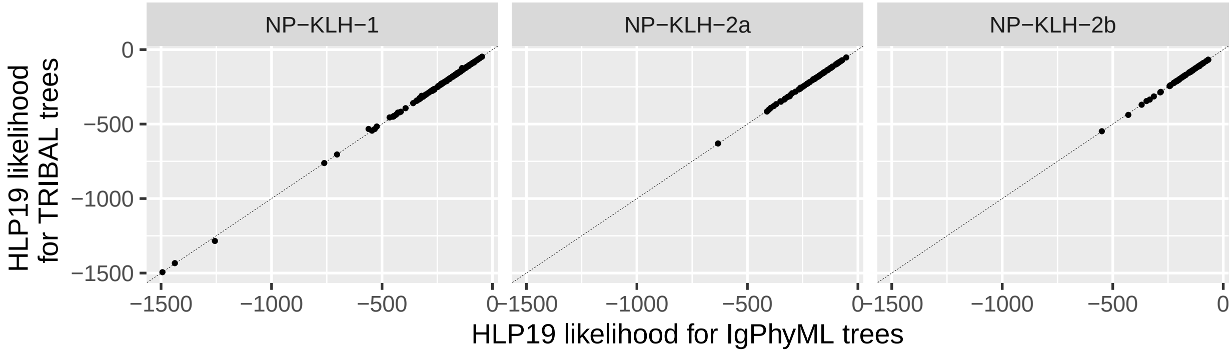


Figure S8: **Comparison of HLP19 likelihood computed for IgPhyML and TRIBAL inferred B cell lineage trees for NP-KLH datasets.**

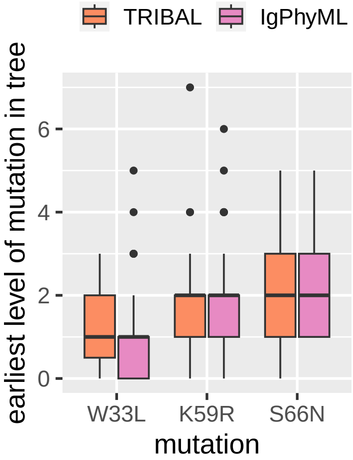
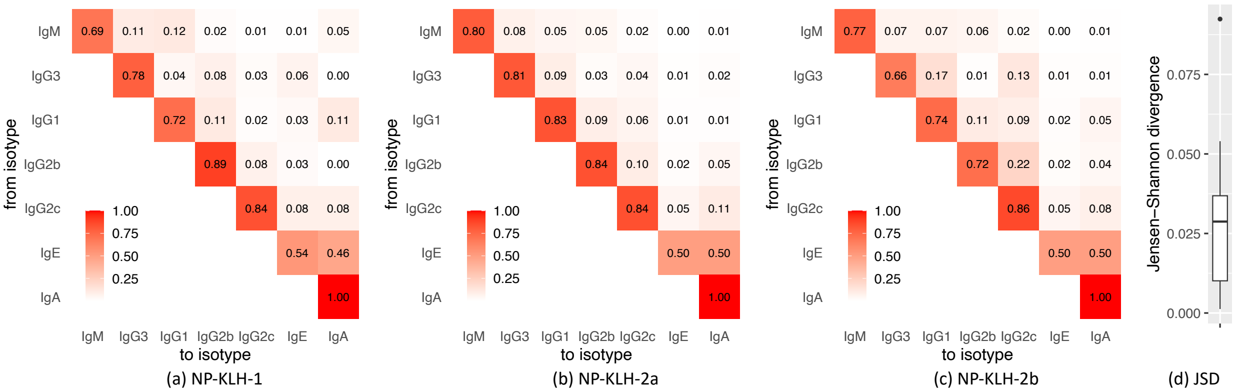


Figure S9: **Earliest observed level of mutation in a B cell lineage tree.** Level 0 represents the MRCA.



(a) NP-KLH-1 (b) NP-KLH-2a (c) NP-KLH-2b (d) JSD

Figure S10: **TRIBAL inferred isotype transition probabilities for NP-KLH.** (a) Isotype transition probabilities for NP-KLH-1. (b) Isotype transition probabilities for NP-KLH-2a. (c) Isotype transition probabilities for NP-KLH-2b. (d) The distribution of Jensen-Shannon divergence (JSD) for pairwise comparisons of rows of the inferred isotype transition probabilty matrices for IgM through Ig2c. IgE was excluded from comparison due to a lack of observed B cells within each dataset to yield informative estimates. IgA is excluded as the inference of this row is trivial.

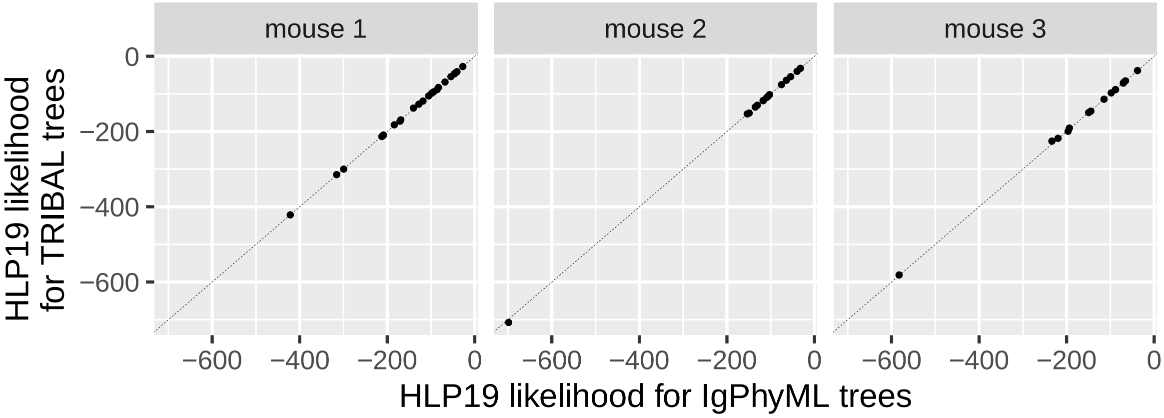
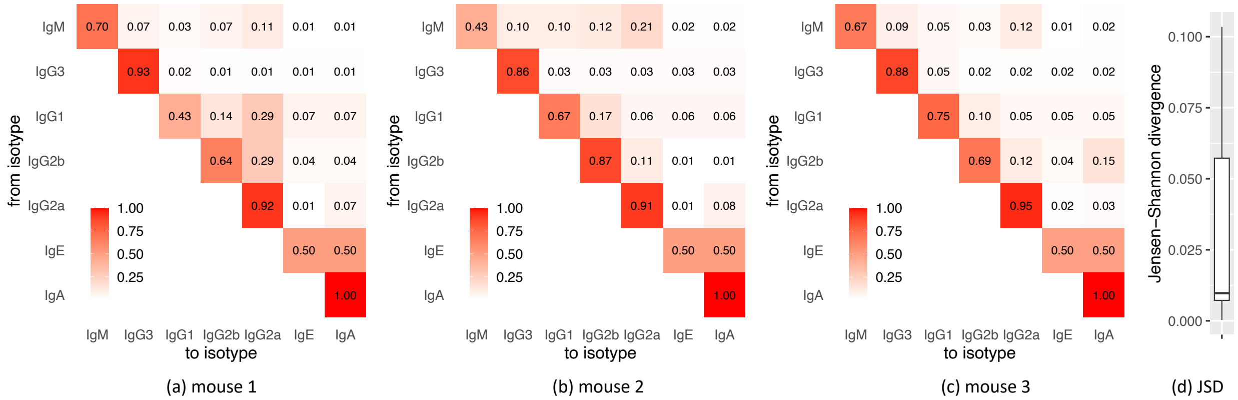


Figure S11: **Scatterplot comparing HLP19 likelihood for IgPhyML trees to the HLP19 likelihood computed for TRIBAL trees for ABC datasets.**

dataset	clonotypes k	total cells n	median cells per clonotype	max cells per clonotype	median distinct isotypes per clonotype
mouse 1	24	224	7.5	31	2
mouse 2	15	218	7	81	3
mouse 3	15	157	7	39	3

Table S2: **Summary of ABC mouse scRNA-seq datasets**



(a) mouse 1 (b) mouse 2 (c) mouse 3 (d) JSD

Figure S12: **TRIBAL inferred isotype transition probabilities for ABC datasets.** (a) Isotype transition probabilities for Mouse 1. (b) Isotype transition probabilities for Mouse 2. (c) Isotype transition probabilities for NP-Mouse 3. (d) The distribution of Jensen-Shannon divergence (JSD) for pairwise comparisons of rows of the inferred isotype transition probability matrices for IgM through Ig2c. IgE was excluded from comparison due to a lack of observed B cells within each dataset to yield informative estimates. IgA is excluded as the inference of this row is trivial.