Assessing the inference of single-cell phylogenies and population dynamics from genetic lineage tracing data

Julia Pilarski[1,2*], Tanja Stadler[1,2*†] and Sophie Seidel[1,2*†]

[1]Department of Biosystems Science and Engineering, ETH Zurich, 4056 Basel, Switzerland.
[2]Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.

*Corresponding author(s). E-mail(s): julia.pilarski@bsse.ethz.ch; tanja.stadler@bsse.ethz.ch; sophie.seidel@bsse.ethz.ch;
†These authors contributed equally to this work.

Abstract

Background: Multicellular organisms develop from a single cell by repeated rounds of cell division, differentiation and apoptosis, which can be displayed in a single-cell phylogenetic tree. Genetic lineage tracing allows us to investigate this development by tracking the ancestry of individual cells as populations grow and change over time. However, accurate reconstruction of the cell phylogeny and quantification of the corresponding phylodynamic parameters - division, differentiation and apoptosis rate - from this tracking data remains challenging and needs to be systematically evaluated.

Results: We perform simulations and assess, using the Bayesian framework, the joint inference of time-scaled cell phylogenies and phylodynamic parameters from CRISPR lineage recordings with random or sequential edits. Principally, we characterize the inference improvement as the recorder capacity increases. Further, we characterize the increase of accuracy of phylogenetic reconstruction when using sequential compared to random recordings, i.e. when using the additional information contained in the order of edits. Moreover, we find that CRISPR lineage recordings carry a strong signal on the rates of cell division. When the phylodynamic parameters are inferred under models that match the true dynamics

and when sufficiently many cells of each type are sampled, also cell death and differentiation rates can be estimated from the data.
Conclusion: In this study, we evaluate how much information on cellular development can be extracted from genetic lineage tracing data using phylogenetic and phylodynamic methodology. We identify important experimental, conceptual, and computational limitations for the inference which can guide future advancements in the field.

Keywords: CRISPR lineage tracing, single-cell phylodynamics, Bayesian inference

1 Background

Uncovering how cells proliferate and specialize is crucial for understanding fundamental biological processes such as the development of multicellular organisms from a single cell, tissue regeneration, and disease progression. In recent years, genetic lineage tracing has emerged as a powerful technology for recording the lineages of individual cells over time. Increasingly, CRISPR-Cas systems are leveraged to induce heritable and irreversible mutations ('edits') in short DNA sequences ('targets' or 'barcodes') which are introduced into the cells' genomes [1–5]. Targets accumulate edits over time, as they are passed on from cells to their descendants. At the end of this process, the diverse barcodes can be retrieved by single-cell sequencing and used to reconstruct cell lineage trees ('phylogenetic trees' or 'phylogenies'). Such trees, in turn, contain information on how cell populations grow and differentiate, and can be analyzed within the statistical framework of phylodynamics [6].

Several CRISPR lineage recorders have been developed in the past decade. Many of them consist of multiple targets, each of which is a CRISPR-Cas9 target site that independently acquires indel mutations upon activation of the editing reagents [7–10] ('non-sequential' recorders). Novel tools employ prime editors that introduce short template-based insertions in tandem arrays ('tapes') of target sites, enabling an ordered recording of edits [11, 12] ('sequential' recorders).

Despite technological advancements, it remains computationally challenging to accurately infer the cell relationships from mutations observed in a sample of cells at a single point in time [1]. Previous studies [13–15] have demonstrated limits of phylogeny reconstruction from CRISPR lineage recordings. In particular, they have shown that constraints on recorder capacity, the editing rate, and diversity of editing outcomes can result in not completely resolved phylogenies. Nevertheless, the studies have identified, using simulation and theory, experimental conditions over which exact tree reconstruction is possible.

Importantly, non-sequential and sequential recorders have been investigated separately until now, and a comparison of their information content is lacking. Moreover, little attention has been paid so far to the quantification of population dynamics, i.e., the dynamics of cell division, differentiation and apoptosis, based on CRISPR lineage recordings. We aim to fill the gap by evaluating the inference of phylodynamic parameters, alongside cell phylogenies, from the data. Given the absence of ground truth information for most linage tracing experiments, we perform simulations. We mimic lineage tracing experiments under a wide range of realistic scenarios and simulate data for both non-sequential and sequential recorders. Then, we apply phylodynamic models and assess how much signal the data contains to accurately infer the cell phylogeny and the population dynamics. In particular, we employ the Bayesian framework, allowing us to incorporate prior knowledge into the analysis and capture the uncertainty associated with the inference results.

Our main goal is to compare the phylogenetic and phylodynamic signal carried by non-sequential and sequential CRISPR lineage recorders. By systematically altering the number of editable target sites, the editing rate, and the editing window in our simulations, we show that recorder capacity significantly affects the inference performance. Furthermore, we investigate in what scenarios the rates of cell division, differentiation and apoptosis can be estimated reliably from the data, and identify important

limitations of the currently available phylodynamic models to capture realistic cell population dynamics.

2 Results

2.1 The workflow

To evaluate and compare the information contained in genetic lineage tracing data, we established the following workflow Fig. 1 (for details, see Methods): First, we simulated time-scaled cell phylogenies which represent the development of cell populations from a single ancestor. Note that we use here the terms and 'phylogeny' and 'phylogenetic tree' interchangeably to refer to the ancestry of a sample of cells from a population. Others might call that object also 'lineage tree', as discussed in [6]. We considered multiple tree generating processes (population dynamics or phylodynamic models) and simulated 20 trees per model. Each process was motivated by realistic cell population dynamics, from completely synchronous divisions in a homogeneous cell population resembling early embryonic development, to stochastic divisions in a heterogeneous cell population resembling differentiating cells.

Second, we simulated sequential and non-sequential CRISPR lineage recordings along the phylogenies. We assumed that all target sites are unedited in the ancestor cell and evolve over time. Thus, for each tree and recorder, we obtained an alignment of barcodes. This simulation was done on fixed trees, implicitly assuming that the cell division, apoptosis and differentiation process generating the trees is independent of the barcode evolution process. Indeed, experimentalists aim to insert barcodes which do not alter the cellular dynamics.

Third, from each simulated alignment, we jointly reconstructed the time-scaled cell phylogeny and estimated the parameters of population dynamics in a Bayesian Markov chain Monte Carlo (MCMC) framework in BEAST 2 [16]. Specifically, we applied the editing models *TiDeTree* [17] and *SciPhy* [18] to our simulated non-sequential and

sequential CRISPR lineage recordings, respectively, and fitted birth-death sampling models [19–21] to the data.

Fourth, we evaluated phylogeny reconstruction by computing the weighted Robinson-Foulds (wRF) distance [22, 23] between each inferred tree and true tree. Importantly, this metric considers branch lengths in addition to the tree topology and thus evaluates time-scaled cell phylogenies. Additionally, we inspected the posterior distributions of the tree and phylodynamic parameters. We assessed the coverage by computing the fraction of 95% highest posterior density (HPD) intervals which contained the true parameters. Also, we assessed the information gain (when using sequence data in addition to the prior assumptions) by computing the 'HPD proportion', i.e., the width of the estimated HPD intervals with respect to prior distributions. Further, we determined the relative bias of the inferred parameters by comparing posterior medians to the true values, and quantified the certainty in the estimates by computing relative HPD widths.

Overall, our workflow served to analyze systematically the phylogenetic and phylodynamic signal in CRISPR lineage tracing data.

2.2 Assessing the phylogenetic and phylodynamic signal in CRISPR lineage recordings - the baseline

Initially, we generated cell phylogenies where all cells undergo the same dynamics, and simulated CRISPR lineage recordings under a representative experimental design (baseline).

Specifically, we simulated cell populations growing either by synchronous and regular cell divisions, or in a stochastic manner. In the former, we assumed that the time to division is equal for each cell and defined four models as follows: (1) cells divide and are sampled completely, (2) cells divide and are sampled incompletely, (3) cells divide, die with a fixed probability and are sampled completely, and (4) cells divide,

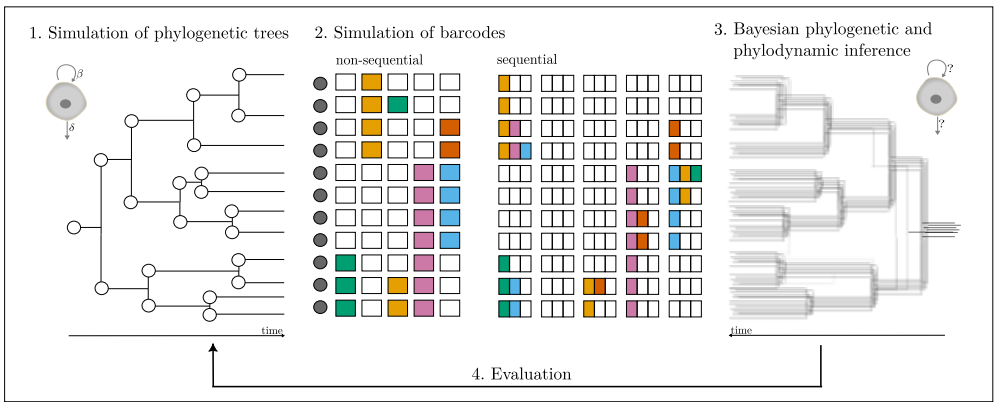

Fig. 1: Schematic of the workflow. First, we generated phylogenetic trees according to different phylodynamic models, starting with a single ancestor cell. Branches represent cells, internal nodes represent cell divisions, tips represent the sampled cells, branch lengths correspond to time. Second, we simulated CRISPR lineage recordings along the trees. Non-sequential recordings consist of independent targets, where each target acquires an edit at random. In contrast, sequential recordings comprise arrays of target sites, where each array acquires edits in sequential order. We collected barcodes (accumulated edits on target sites) for all sampled cells. Colors indicate different editing outcomes. Third, on each set of barcodes, we applied Bayesian phylogenetic and phylodynamic inference to jointly reconstruct the time-scaled phylogenetic tree and estimate the parameters of cell population dynamics. Fourth, we evaluated the inference by comparing the inferred tree and parameters to the ground truth.

die with a fixed probability and are sampled incompletely (see details in section 5.1). For stochastic dynamics, we used the constant rate birth-death process [19, 20] where the time to birth and death is exponentially distributed. In total, we obtained 100 phylogenies, i.e. 20 per model.

Along them, we simulated non-sequential recordings on 20 targets and sequential recordings on 20 tapes of length 5. We used an editing rate of 0.05 so that most, but not all, target sites per cell mutated. In total, we performed 100 simulations (one per tree) per lineage recorder.

In the inference, we have to specify a barcode evolution model and a tree generation model. For the barcode evolution model, we employed the frameworks TiDeTree and SciPhy to the non-sequential and sequential barcode evolution, respectively; we used these frameworks for both simulation and inference. For the cell population growth,

i.e. the tree generation, we assumed the commonly employed stochastic birth-death process.

Note that this inference model matches the simulation model perfectly when analyzing birth-death trees, and we expect good estimates of the model parameters given enough data and appropriate priors. However, when analyzing the synchronous cell division trees, there is a model misspecification which may lead to a reduced quality of the estimates. As there is currently no inference model corresponding to the simulation model of synchronous cell divisions, we explored to what extend the birth and death rates inferred under the birth-death model can be interpreted. In this section, we characterize how well the model parameters can be inferred across dynamics, and in later sections we explore how these estimates change when changing aspects of the simulation.

As shown in Fig. 2, the true editing rate and phylogenetic tree parameters (height and length) were recovered for more than 80% of the simulations with a small relative bias. The true birth and death rates were recovered for all simulations in which the birth-date process generated the phylogenetic tree. However, their recovery was substantially lower for simulations on phylogenetic trees which grew by synchronous cell divisions, as expected due to the model misspecificaton. In particular, death rates were consistently underestimated for simulations on synchronous trees with cell death, and overestimated for simulations on synchronous trees without cell death. The latter could be explained by the nature of the MCMC chain and the choice of prior distribution on the death rate. We used an exponential distribution, defined on $[0, \infty)$ and continuous, thus, the probability for the chain to take the single value 0 was zero, and the samples were taken from small but positive real numbers. Importantly, all 95% HPD intervals of the posterior estimates approached the true value 0.

On average, the posterior distributions of the birth and editing rates shrank by more than 90% relative to prior distributions, suggesting high confidence in the estimates and strong signal in the data. Compared to these parameters, the uncertainty about the inferred death rates was high (cf. HPD proportion in Fig. 2, and Fig. B2).

Further, the wRF distance between the inferred trees and true trees oscillated around 0.2 for non-sequential recordings. The distance was halved (median 0.1) for sequential recordings, which could be due to the overall larger number of target sites in the sequential recorder or due to the sequential edits themselves.

Overall, these results indicate that our simulated CRISPR lineage recordings contained information on both cell relationships and the timing of cell divisions. However, they bore a weaker signal on cell death rates. The estimation of death rates was also more sensitive to model misspecification.

2.3 Varying experimental parameters

To evaluate how different experimental parameters affect the inference from CRISPR lineage recordings, we varied the editing rate, the number of editable target sites, and the editing window in our simulations. To ensure that any differences to the baseline are due to the varied parameters, we re-used the 100 phylogenetic trees from the baseline. Thus, we simulated 100 alignments per setting. Fig. 3 illustrates the general trends, whereas Fig. B3 and B4 show differences between the tree generating processes.

Of particular interest in CRISPR lineage tracing experiments is calibrating the rate at which target sites accumulate edits. The recorders are characterized by a limited number of target sites, so tuning the editing rate is necessary to obtain sufficiently diverse barcodes at the end of an experiment. A too low editing rate may provide only scarce information, while a too high editing rate leads to recorder saturation before the end of the experiment and missing information on later cell development.

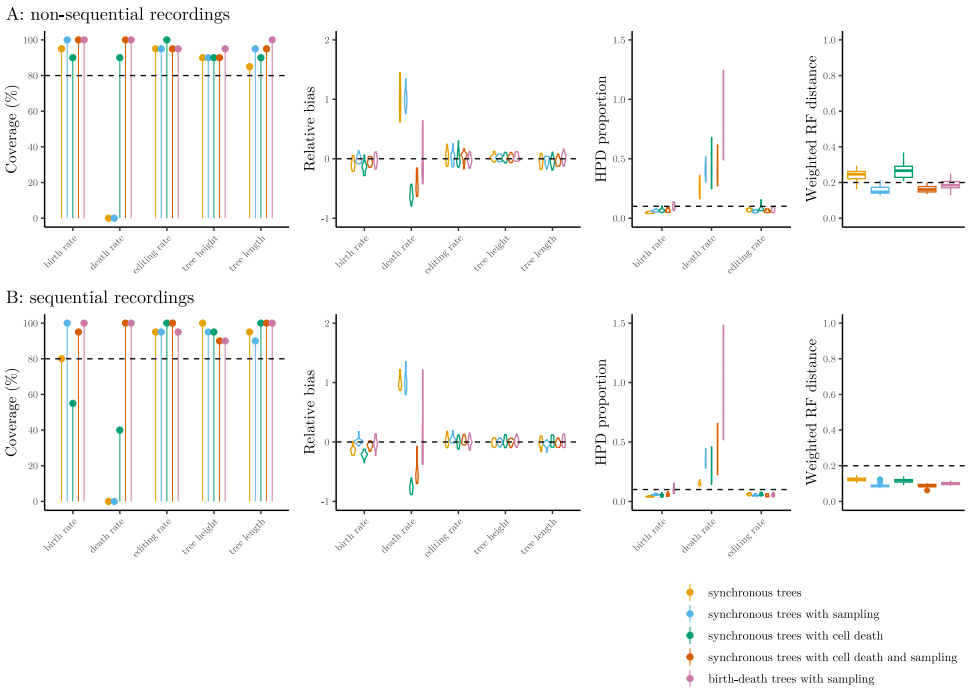

Fig. 2: Inference performance on (A) non-sequential and (B) sequential CRISPR lineage recordings. The graphs show from left to right: the coverage, relative bias and HPD proportion of the numerical parameters, and the weighted Robinson-Foulds (RF) distance between the inferred and true trees. The inference results are summarized across 20 simulations per tree generating process (indicated by colors). Dashed lines display thresholds for each metric to facilitate visual comparability.

Here, we investigated what consequences these scenarios have for phylogenetic and phylodynamic inference.

First, we reduced the editing rate to 0.01 so that less than half of the available targets or tapes per cell mutated. Then, we increased the editing rate to 0.15 so that all targets in the non-sequential recorder and at least the first position in all tapes in the sequential recorder mutated. We observed that phylogeny reconstruction from non-sequential recordings was less accurate for the lower and higher editing rate compared to the baseline. In fact, at the highest editing rate, groups of cells acquired identical barcodes, indicating recorder saturation. Under these conditions, the distance from

the inferred to true phylogenetic trees almost doubled. In contrast, reconstruction of the cell phylogeny from sequential recordings was most accurate for the highest editing rate. As target sites in the tapes were activated in order, the tapes did not saturate too early and accumulated information until the end of the experiment.

Next, we decreased and increased the number of targets in the non-sequential recorder, and the number and length of tapes in the sequential recorder. We observed that the inference improved consistently when more target sites were available, as expected. In particular, the average distance between the inferred and true phylogenetic trees has halved for recorders with 20 compared to only 5 targets or tapes, and decreased further for 40 targets or tapes. This result nicely agrees with general statistical theory where adding four times as many observations in the sample reduces the standard error of the mean by a factor of two.

Finally, many systems of interest develop over long time spans over which current technologies cannot sustain editing entirely. Thus, we examined the effect of a limited editing period and its placement during the experiment on parameter estimates. In particular, we restricted the editing window in the less capacious, non-sequential recorder to half of the experimental period. Editing occurred either in the first half, in an interval around the middle, or in the second half of the experiment. In the last case, cell barcodes evolved to the most diverse set. Surprisingly, phylogeny reconstruction from barcodes edited in the middle or second half of the experiment was on average as accurate as reconstruction from barcodes edited throughout the entire experiment, as measured by the weighted RF distance. This result might be affected by the timing of the cell division events. As more cell divisions occur towards the end of the experiment, the number of splits increases, and later editing appears better under that metric. Two generalized RF metrics, measuring tree topology only, indicate that phylogeny reconstruction was less accurate when editing was restricted to

the first or second half, compared to the baseline, but was similarly accurate when editing occurred in an interval around the middle of the experiment (cf. Fig. B4).

Again, we evaluated the inference of cell division and apoptosis rates alongside cell phylogenies. Notably, its performance differed much more subtly across the experimental parameters. In general, we observed that HPD intervals of the birth and death rates became narrower as the capacity of the recorders increased, indicating higher confidence in the estimates. Also, relative HPD widths mirrored the trends in phylogeny reconstruction for the different editing rates (as illustrated in Fig. 3 right).

In summary, recorders with more editable target sites, an editing rate tuned to match the recorder capacity, and more diverse barcodes carried a stronger phylogenetic and phylodynamic signal. It is also notable that some parameters, such as the cell division rate, can be robustly estimated even when the editing window is halved.

2.4 Evaluating sequential editing

While the non-sequential recorder is limited to one edit per target, the sequential recorder has multiple editable positions per tape. At baseline, we saw that the sequential recorder accumulated more edits throughout the experiment, so more diverse barcodes arose. Therefore, phylogeny reconstruction improved and uncertainty in the parameter estimates decreased. But did the inference improve only due to an increase in recorder capacity, or due to the sequential acquisition of edits per se?

To test this, we simulated a scenario where we held everything essentially equivalent and contrasted 20 non-sequential sites with 20 sequential sites (one tape of length 20). For the non-sequential recorder, we could reuse the baseline. For the sequential recorder, we used the same number and distribution of possible editing outcomes as for non-sequential recordings. We further tuned the editing rate to 0.45, such that the average number of edits per barcode and the diversity of barcodes were comparable to the non-sequential recordings. Note that the editing rate r is defined per active target

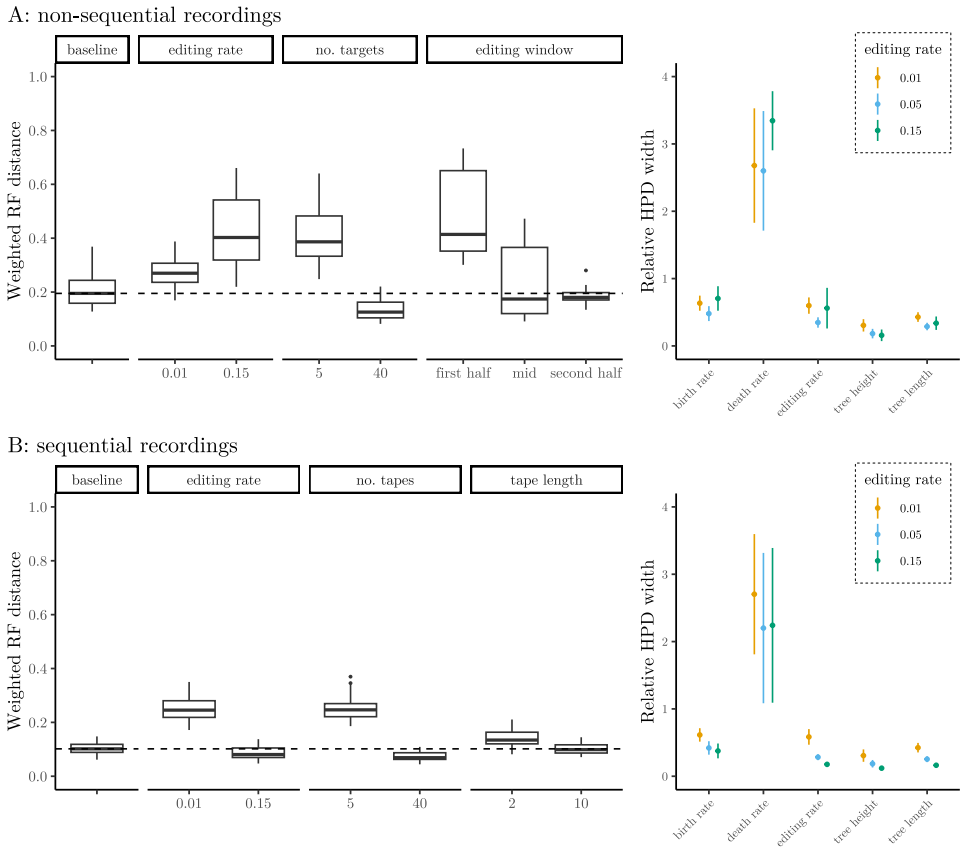

Fig. 3: Inference performance on (A) non-sequential and (B) sequential CRISPR lineage recordings with varying experimental parameters. At baseline, the editing rate was 0.05, the recorders carried 20 targets or 20 tapes of length 5, respectively, and editing occurred throughout the entire experiment. Left graphs show the weighted RF distance between the inferred and true trees. Facets indicate the varied experimental parameters, dashed lines display the baseline medians. Right graphs show the relative HPD width (mean ± standard deviation) for numerical parameters, colored by the editing rate used for simulation. The inference results are summarized across 100 simulations per setting (including both simulated synchronous and birth-death trees).

site. In the non-sequential model, all m target sites are active from the start on, while in the sequential model, target sites in a tape are activated one by one. Hence, the overall editing rate in the non-sequential model is initially $r \cdot m$ and decreases over time until all targets are edited. In contrast, in the sequential model, it is initially r,

constant over time, and changes to 0 only when all target sites are edited. Therefore, tuning the editing rate was necessary to reach a similar amount of edits per cell in the two types of recordings.

Importantly, phylogeny reconstruction, as measured by weighted RF distance, improved significantly when edits accumulated in order (Welch's t-test, $t(193.28) = 5.81, p < 0.001$, see Fig. 4A). However, the inference of phylodynamic parameters did not change substantially. In particular, sequential editing did not increase the accuracy nor reduce the uncertainty in the estimated birth and death rates (Fig. 4B and Fig. B5).

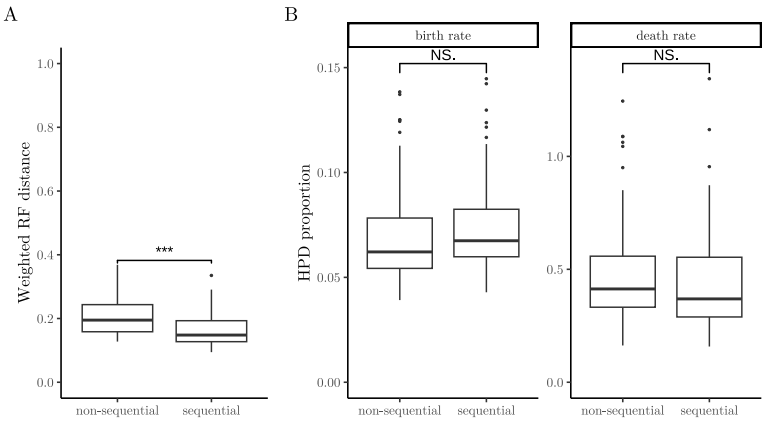

Fig. 4: Comparison of the sequential and non-sequential accumulation of edits in CRISPR recorders. Panel (A) shows the weighted RF distance between the inferred and true trees. Panel (B) shows the HPD proportion of the inferred birth and death rates. The inference results are summarized across 100 simulations per setting (including both simulated synchronous and birth-death trees). Statistical significance was assessed by Welch's t-test, ***: p-value < 0.001, NS.: not significant.

2.5 Inferring cellular differentiation dynamics

A key interest of developmental biologists are the differentiation dynamics of different cell types. To investigate the inference of these dynamics, we simulated non-sequential and sequential CRISPR lineage recordings with baseline experimental parameters

along multi-type birth-death trees [21]. We considered two prototypical dynamics of cell type transitions, terminal and chain-like, of three different cell types (1, 2 and 3) from a single progenitor type 0 (see details in 5.1). From the recordings, we inferred time-scaled and typed cell phylogenies, cell type-specific division (birth) and apoptosis (death) rates, as well as cell type transition (migration) rates.

Importantly, the coverage of all parameters was above 80% for both recorders and both cell type transition dynamics (Fig. B6). In particular, the inferred migration rates varied around the true values for most simulations, and their HPD intervals were much narrower than the prior distribution (median HPD proportion 0.35). For trees with terminal transitions, uncertainty was lowest for migration rate to type 1 and highest for migration rate to type 3. Given that we simulated the trees with a transition rate to type 1 two times higher than the transition rate to type 3, we hypothesized that the higher the true transition rate, the more cells of the corresponding terminal type were sampled, and the higher was the signal for inference.

For both recorders, the migration rate from type 1 to type 2 was overestimated almost tenfold for some trees with chain-like transitions (see Fig. 5I). A closer look at these trees revealed that none of them contained a single tip of type 1. This implies that transitions from type 0 to type 1, or transitions from type 1 to type 2, or both, have occurred very fast such that no cells of type 1 have been sampled at the end of the experiment. It remains unclear, however, why the migration rate to type 2 and not to type 1 was overestimated. It remains to be explored in future work on more simulations whether this result is specific to the selected trees or a general behavior of the inference framework.

Notably, the topology and branch lengths of multi-type phylogenies were inferred comparably to those of single-type phylogenies. Additionally, cell types were correctly inferred for more than 80% of ancestral cells across settings (Fig. 5II). They were consistently inferred better in trees with terminal transitions (median 97.8%) than

those with chain-like transitions (median 87.4%) . This could easily be explained by the observation that each cell could be either of the same type as its descendant or of type 0, whereas more possibilities exist in the case of chain-like transitions.

Taken together, the rates of cellular differentiation were inferred robustly from CRISPR lineage recordings alongside cell relationships when enough cells of the different types were sampled.

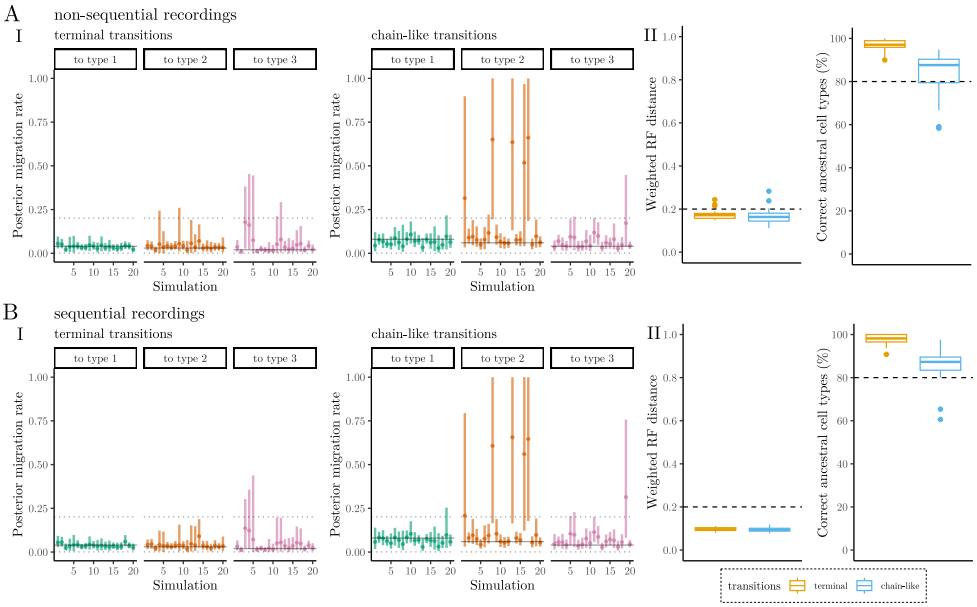

Fig. 5: Inference from non-sequential (A) and sequential (B) recordings on multi-type birth-death trees with terminal (left) and chain-like (right) cell type transitions. Panels (I) show the medians (dots) and 95% HPD intervals (lines) of the inferred migration rates per simulation. Solid horizontal lines indicate the true values, dashed horizontal lines indicate the 95% credible interval of the prior distribution. Panels (II) show a boxplot of the weighted RF distances between the inferred and true tree, and the percentage of correctly estimated ancestral cell types. The metrics are summarized across 20 simulations per recorder and cell type transition dynamics. Dashed lines display thresholds for comparability.

3 Discussion

We have carried out a simulation study to assess the inference of single-cell phylogenies and population dynamics from CRISPR lineage tracing data. In line with Salvador-Martínez et al. [13] and recent theoretical analyses [14, 15], we find that phylogeny reconstruction largely depends on the capacity of CRISPR lineage recorders. Principally, recorders with more target sites can accumulate more edits throughout the experiment and produce more informative barcodes. Crucially, when the editing rate is too low or too high for a given number of target sites, not all cell divisions can be recorded which leads to only partially resolved cell phylogenies. Our results indicate that less accurate phylogeny reconstruction goes along with less certain estimation of phylodynamic parameters.

The study was designed to compare non-sequential and sequential CRISPR lineage recorders. As expected, replacing targets with tandem arrays of target sites results in an increase in recorder capacity, and thus improves the inference. Importantly, we find that the order of edits in sequential recordings carries additional information that significantly improves phylogeny reconstruction. Surprisingly, sequential editing does not significantly improve the inference of phylodynamic parameters.

Further, we have observed that non-sequential recorders saturate faster than sequential recorders at the same editing rate per active target site. Due to constraints on the order in which target sites become active, a sequential recorder can thus track lineages over longer periods of time. One strategy to prevent saturation of a non-sequential recorder during the experiment, is restricting the editing window to the most interesting period of development. Previously, it has been reported [4] that early editing facilitates the recognition of specific clones in a cell population, but inferring relationships among cells within the clones is difficult. Consistent with this, we find that editing in only the first half of the experiment leads to less well resolved phylogenetic trees. Our results indicate, however, that phylogeny reconstruction can be

almost as accurate when editing lasts half the time in an interval around the middle or the second half of the experiment as when it spans throughout the entire experiment. However, this observation likely depends on our simulation setup, where few cell divisions occur early on, and may not generalize to scenarios where most divisions happen at the beginning of the experiment.

Most importantly, our study has revealed how well the parameters of population dynamics can be estimated from CRISPR lineage recordings using the Bayesian inference framework. Notably, we used different cell population models (either models based on synchronous cell divisions or a birth-death process) for simulation and inferred accurate cell division rates across dynamics despite model misspecification. Thus, we find that the recordings carry a strong signal on the rates of cell division, as evidenced by high accuracy and certainty of the estimates across scenarios.

However, our results indicate that signal on the rates of cell death is weak and their estimation is sensitive to model misspecification. Theoretical considerations on the inference of birth and death rates from molecular phylogenies can explain this result. It has been shown that death events leave a characteristic signature in the shape of phylogenetic trees, and can be derived from the increase in the number of lineages through time [24, 25]. However, in small populations, the number of lineages through time is very noisy, and often, not sufficiently many death events occur to reliably estimate the death rate [26]. Furthermore, the quantification of death rate becomes more challenging when only a sample of cells is analysed [24]. Thus, while the signal for death rates is weaker than for cell division rates, this is in line with the use of phylodynamic methods in adjacent fields and not a characteristic of lineage tracing data per se.

Another interesting result is that cell type-specific division rates and differentiation rates can be estimated from CRISPR lineage recordings, when enough cells of each type are sampled (and given no model misspecification). In our simulations, at least

one cell of each type had to be present in the dataset to inform the type-specific phylodynamic parameters, and we required cells to be sampled with the same probability (irrespectively of the type) at the end of the experiment. This finding has implications for developing appropriate procedures to sample cells in real-life experiments. Sampling schemes on experimental replicates should capture sufficiently many cells of all types. It should be explored in further simulations how non-uniform sampling and sampling biases of certain cell types affect the inference. This might be particularly relevant when the real cell type transitions are much more complex than assumed in our simulations, and when more different cell types arise during development, some of which abundantly and others rarely.

In future research, more realistic models of cell population dynamics should be developed and evaluated. Such models would overcome biases in estimating the parameters of cell population dynamics due to model misspecification. Our analysis has shown, for example, that estimates of the apoptosis rate were biased when we simulated synchronous and regular cell divisions, but inferred the rates of cell division and apoptosis under the stochastic birth-death model. More realistic models might further consider, for instance, that cell divisions are initially regular and synchronous, and later become asynchronous and stochastic. They should also account for more complex dynamics of cellular differentiation, and varying rates of cell division and apoptosis across types, lineages, and over time. Inference models that faithfully describe cell population dynamics would improve phylodynamic inference from empirical lineage tracing data.

In our simulation study, we did not consider several factors that might affect the phylogenetic and phylodynamic inference from CRISPR lineage recordings. For example, in real-life experiments, the rate of editing can change over time [8], as it is challenging to maintain the desired level of activity of the editing enzyme over long periods of time [4]. Furthermore, technical issues such as target dropout, barcode

loss, barcode homoplasy, and noise in barcode readout commonly occur and result in missing and erroneous data [1, 13]. Future studies should account for these factors in simulations and evaluate their impact on inference.

Our workflow has faced the computational limitation of high runtime complexity. Each MCMC chain required at least a few hours to collect sufficiently many samples and converge to the posterior distribution. Inference with the multi-type phylodynamic model lasted up to three weeks. Scaling Bayesian inference methods to analyze lineage recordings on populations consisting of thousands or even millions of cells is a core challenge. On the other hand, it remains an open question how many cells are required to get accurate estimates of cell division, differentiation and apoptosis rates, and thus, to what number of cells the methods need to scale.

4 Conclusions

Altogether, phylogenetic and phylodynamic inference from CRISPR lineage recordings is promising, but faces various challenges, as we have shown in our simulations. In terms of experiment design, CRISPR lineage recorders have a limited capacity and thus require careful tuning of the editing rate or restricting the editing window to collect sufficient information for the inference. Conceptually, inference models which accurately represent various cell population dynamics, such as synchronous cell divisions, are missing and model misspecification leads to biases. Finally, high runtime and applicability to small samples are major computational limitations of the inference. These issues should be addressed as genetic lineage tracing technologies and analysis methods advance. Going forward, it will be exciting to integrate genetic lineage tracing data with other molecular single-cell measurements, with the goal of establishing more comprehensive models of cellular development.

5 Methods

5.1 Simulation of phylogenetic trees

We started each tree generating process with a single cell at time $t = 0$ and let the cell populations evolve until time $T = 40$, measured in arbitrary time units. We generated time-scaled phylogenetic trees in which branches represent cells, internal nodes represent cell divisions, tips represent contemporaneous cells and branch lengths correspond to time.

Homogeneous cell populations. In the first set of simulations, we assumed all cells share the same population rates of cell division, apoptosis and sampling, and generated phylogenetic trees in R using the packages *ape* [27] and *TreeSim* [28]. We considered five population dynamics: First, we assumed that cells divide synchronously in regular time intervals of length $\frac{T}{n+1}$, where n is the number of cell division time points, and all their descendants survive (*synchronous trees*). Next, we added the possibility that cells die with a fixed probability p_d before the synchronous cell division time points (*synchronous trees with cell death*). For both processes, we then sampled living cells from the population at the end of the process with sampling fraction $\rho < 1$. The resulting *synchronous trees with sampling* or *synchronous trees with cell death and sampling*, respectively, contained only lineages of the sampled cells. Finally, we generated trees using the constant rate birth-death sampling model [19, 20], parameterized by a birth (cell division) rate β, a death (apoptosis) rate δ, and sampling probability ρ with which cells at time $T = 40$ are sampled. In these *birth-death trees with sampling*, cells divided or died in a stochastic manner.

Heterogeneous cell populations. In the second set of simulations, we generated cell phylogenies under the multi-type birth-death branching model [21] using the

package *BDMM-Prime* [29–31] in *BEAST v2.7.4* [16]. Here, we assumed that cell populations consist of multiple types, each having its own birth rate β_i and death rate δ_i. Migration rates γ_{ij} describe the rate of cells changing from one type i to another j. We considered four cell types such that $i, j = 0, 1, 2, 3$. Again, we sampled cells at the end of the process, at time $T = 40$, with sampling probability ρ, i.e. each cell was equally likely to be sampled.

Based on experimental observations, we considered two prototypical dynamics of cell type transitions. The first one was motivated by a study on *C. elegans* which reported that cell type changes can be abrupt and many distinct terminal cell types arise from progenitor cells [32]. Hence, in trees with *terminal transitions*, cells of type 0 could transition to type 1, 2 or 3, but all descendants of cells of type $1, 2, 3$ were of the same type.

The second dynamics was inspired by research on mouse embryonic stem cells which revealed a stochastic, chain-like network of cell-state transitions [33]. Thus, in trees with *chain-like transitions*, cells of type 0 could only transition to type 1, cells of type 1 to type 2, and cells of type 2 to type 3. In this study, we considered only irreversible cell type transitions.

We chose the parameters of all population dynamics described above such that the expected number of sampled cells, corresponding to the number of tree tips, was roughly 100. Our choice of parameters was also guided by simulations in previous studies [17] and an empirical observation of cell development from mouse embryonic stem cells [34]. For comparability across all simulations, we selected trees with 20 to 200 tips. In the multi-type case, we assigned slightly different birth, death and migration rates to the different types. From phylogenies with terminal transitions, we selected trees in which the sampled population consisted of at least one cell of type

1, 2 and 3. From phylogenies with chain-like transitions, we selected trees in which at least one sampled cell reached type 3.

The parameters of the trees are summarized in Tables 1 and 2 below.

Table 1: Parameters for simulating homogeneous phylogenetic trees

Trees	n	p_d	β	δ	ρ
synchronous	**7**	**0**	0.12	0	**1**
synchronous with sampling	**10**	**0**	0.17	0	**0.1**
synchronous with cell death	**10**	**0.2**	0.14	0.035	**1**
synchronous with cell death and sampling	**12**	**0.1**	0.19	0.02	**0.1**
birth-death with sampling	-	0.11	**0.2**	**0.025**	**0.1**

We specified the values in bold, and approximated the remaining parameters (on population level, see 5.5).

Table 2: Parameters for simulating heterogeneous phylogenetic trees

Parameter	Value
birth rate	$\beta = (\beta_i)_{i=0,1,2,3} = (0.2, 0.18, 0.18, 0.18)$
death rate	$\delta = (\delta_i)_{i=0,1,2,3} = (0.01, 0.03, 0.03, 0.03)$
migration rate	$\gamma_{\text{terminal}} = (\gamma_{ij})_{i,j=0,1,2,3} = \begin{pmatrix} 0 & 0.04 & 0.03 & 0.02 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
	$\gamma_{\text{chain-like}} = (\gamma_{ij})_{i,j=0,1,2,3} = \begin{pmatrix} 0 & 0.08 & 0 & 0 \\ 0 & 0 & 0.06 & 0 \\ 0 & 0 & 0 & 0.04 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
sampling proportion	$\rho = 0.1$

5.2 Simulation of barcodes

We simulated the cumulative acquisition of edits in CRISPR lineage recorders in BEAST 2 using the packages *TiDeTree* [17] and *SciPhy* [18].

The frameworks TiDeTree and SciPhy model the evolution of each target as a continuous time Markov chain on the state space of all possible editing outcomes.

The unedited state is encoded as 0, the possible edits in TiDeTree (scars) are indexed by $1, ..., S$ and the possible edits in SciPhy (insertions) are indexed by $1, ..., I$.

Non-sequential recordings. TiDeTree models the following CRISPR-Cas9 editing process: At time $t = 0$, a single cell exists with m unedited targets at independent genomic loci. At time $0 \leq t_e < T$, where t_e denotes the editing height and T the experiment duration, the editing window begins by injecting the editing reagents or inducing their expression. The editing reagents are guided to target sites which, consequently, acquire irreversible indels at a constant editing rate r. The scarring probabilities $s_1, ..., s_S$ indicate the relative frequency of each edit, also called scar, appearing on a target. In subsequent rounds of cell division, the accumulated edits are passed on to descendant cells. Editing may be suspended after a time interval Δt_e, the editing duration, for example, when the editing reagents degrade. At the end of the experiment, at time T, a subset of cells is selected for single-cell sequencing and the accumulated edits at target sites (barcodes) are read out.

Sequential recordings. In contrast, SciPhy models ordered editing in a CRISPR lineage recorder with a prime editor. The ancestor cell at time $t = 0$ contains k tagged tandem arrays (tapes) of l target sites. At start, all but the first targets in each tape are inactive. Once the editing reagents are guided to a target site and an editing event takes place, the position of the active target is shifted by one unit along the array. Editing outcomes are, in this case, irreversible short template-based insertions that occur with insert probabilities $i_1, ..., i_I$ at editing rate r. Dividing cells pass their tapes with accumulated edits to their descendants. Editing continues until the experiment ends at time T. Then, cells are sampled and their barcodes are read out.

We simulated CRISPR lineage recordings along time-scaled cell phylogenies. Thus, for

all tips in each tree, we obtained one vector of length m for non-sequential recordings and k vectors of length l for sequential recordings.

For our simulations, we fixed the experiment duration to $T = 40$. Empirical observations indicate that few editing outcomes are much more common than others [11, 13], hence, we sampled their frequencies from an exponential distribution and scaled them to obtain scarring and insert probabilities (Table A1). We varied the remaining experimental parameters as summarized below. We have used *Snakemake* [35] to automate the simulations.

Table 3: Parameters for simulating CRISPR lineage recordings

TiDeTree			SciPhy		
Parameter	Baseline	Variations	Parameter	Baseline	Variations
editing rate r	0.05	0.01, 0.15	editing rate r	0.05	0.01, 0.15, 0.45
number of targets m	20	5, 40	number of tapes k	20	1, 5, 40
editing height t_e	40	20, 30	tape length l	5	2, 10, 20
editing duration Δt_e	40	20			

5.3 Bayesian inference

In general, Bayesian phylogenetic and phylodynamic inference [36] starts with a collection of n sequences, denoted alignment A. It assumes that the sequences evolved from a common ancestor according to a substitution model with rate matrix $\mathcal{Q}$ along the branches of a tree $\mathcal{T}$. The tree itself resulted from a population dynamic process specified by a set of parameters η. The goal is to infer the unknown aspects of the past process, that is, the tree and parameters of the substitution and phylodynamic models from the sequences. Prior information on the parameters of the assumed model M have to be specified. Applying the Bayes rule and assumptions of independence between the components results in the expression

$$P(\mathcal{T}, \mathcal{Q}, \eta | A, M) = \frac{P(A|\mathcal{T}, \mathcal{Q}, M) P(\mathcal{T}|\eta, M) P(\mathcal{Q}|M) P(\eta|M)}{P(A|M)}$$

for the posterior distribution of the variables of interest. The term $P(A|\mathcal{T}, \mathcal{Q}, M)$ is called the phylogenetic likelihood and $P(\mathcal{T}|\eta, M)$ the phylodynamic likelihood.

Due to the large dimensionality of the state space and the impractical evaluation of the marginal likelihood $P(A|M)$, Monte Carlo algorithms, such as Markov chain Monte Carlo (MCMC), are commonly used to estimate the posterior distribution. The MCMC algorithm explores the state space by collecting samples from the posterior distribution. Each sample consists of the numerical parameters of the substitution and phylodynamic models, as well as the tree topology and the associated branch lengths.

We employed the Bayesian MCMC framework in BEAST 2 to jointly infer time-scaled cell phylogenies and parameters of the cell population processes as well as editing process from genetic lineage tracing data. Our simulated barcodes constituted the input alignments.

We fit the editing (substitution) models from TiDeTree and SciPhy to the non-sequential and sequential barcodes, respectively. We inferred the editing rate and fixed the scarring and insert probabilities to true values, because we reasoned that the occurrence and relative frequency of each editing outcome can be quantified in real experiments using sequencing data.

To infer population dynamics, we used the birth-death skyline contemporary model implemented in *BDSKY* [37] for homogeneous cell populations, and the multi-type birth-death model implemented in BDMM-Prime [29–31] for heterogeneous cell populations. Due to identifiability reasons [20], we fixed the sampling proportion to the truth and inferred the birth, death and non-zero migration rates. Fixing the sampling proportion can also be done for empirical data as the sampling proportion at the end of the experiment can be determined.

We fixed the origin of the phylogenetic trees at 40, reflecting the experiment duration, and specified the editing window. We inferred the tree topology and branch lengths, from which we derived the tree length (the sum of all branch lengths)

and tree height. For heterogeneous populations, we first reconstructed tip-typed trees [31] (meaning types ancestral to the tips were not inferred) and, subsequently, stochastically mapped ancestral type changes on the trees.

We used weakly informative distributions as priors in our analysis (Table 4). We let the analysis run for 10^8 steps or until the effective sample size (ESS) was above 200 and discarded 10% of the analysis to account for burn-in.

Table 4: Prior distributions for inference

Parameter	Prior	Percentiles $[5^{\text{th}}, 95^{\text{th}}]$
editing rate r	LogNormal(-3, 1)[1]	$[0.01, 0.26]$
birth rate β	LogNormal(-1.5, 1)	$[0.04, 1.16]$
death rate δ	Exponential(25)	$[0.002, 0.12]$
migration rate γ	LogNormal(-3.25, 1)	$[0.007, 0.20]$

In the multi-type model, we used the same prior distribution for each non-zero element of the vectors β, δ and the matrices γ.

[1]For $r = 0.45$, we used LogNormal(-1, 1) with percentiles $[0.07, 1.91]$.

5.4 Evaluation

We characterized different proxies for information content in our simulated alignments by calculating the number of different editing outcomes, the number of edits per barcode, and the pairwise Hamming distance between barcodes. We defined alignment diversity as the number of unique barcodes relative to their total number.

For each simulation and the respective inference run, and for each inferred parameter, we computed the median and the 95% highest posterior density (HPD) interval of the posterior probability distribution. To evaluate the inference performance for parameters of particular interest (editing, birth, death and migration rates, tree height and tree length) across simulations, we calculated the following metrics:

- Coverage: fraction of times the true parameter was contained in HPD interval.

- Relative bias: difference between the posterior median and the true parameter, divided by the truth.
- Relative absolute error: absolute difference between the posterior median and the true parameter, divided by the truth.
- Relative HPD width: difference between the upper and lower bound of the posterior 95% HPD interval divided by the true value of the parameter.
- HPD proportion (only for parameters with a explicitly specified prior distribution; but not for tree height and tree length): difference between the upper and lower bound of the posterior 95% HPD interval divided by the difference between the upper and lower bound of the 95% credible interval of the prior distribution.

For synchronous trees with birth only, the true death rate was 0, so we normalized the metrics by the mean of the median estimates instead of the truth. To be able to compare the inference of the death rate across population models, we also calculated not normalized metrics for this parameter.

To evaluate the reconstruction of cell phylogenies, we derived for each inference the 95% credible set of tree topologies, that is the smallest set of all tree topologies that accounts for 95% of the posterior probability. We measured its relative size by dividing the number of trees in the credible set by the number of sampled trees. Then, we calculated tree coverage as the fraction of times the true tree was in the credible set across simulations. Furthermore, we summarized the posterior trees to the maximum clade credibility (MCC) tree with the node heights rescaled to the posterior median node heights for the clades contained in the tree. We quantified the topological distance between each MCC ('inferred') tree and the corresponding true phylogenetic tree in three ways. We computed the weighted Robinson-Foulds (RF) distance [22, 23] which considers branch lengths alongside tree topology. Additionally, we computed two generalized RF distance metrics, Nye similarity [38] and Shared Phylogenetic Information [39], which account for differences between similar but not identical pairs

of tree splits. These calculations were done using the R packages *TreeDist* [40] and *phangorn* [41].

For multi-type phylogenies, we evaluated the inference of ancestral cell types by identifying clades that a given MCC tree and the corresponding true tree have in common, and calculating the proportion of internal nodes in these clades with correctly assigned types.

Additionally, we used the R packages *treeio* [42, 43], *tracerer* [44] and *tidyverse* [45] for evaluation.

5.5 Approximation

As described in 5.3, we fitted the birth-death model, parametrized by a birth and death rate, to all homogeneous cell phylogenies. However, the population model underlying synchronous trees is parametrized differently, by the number of time points at which all living cells simultaneously divide, and optionally, the probability of cell death. To be able to evaluate the phylodynamic inference for phylogenetic trees with synchronous cell divisions, we approximated their 'generative' birth and death rates in two ways. We compared these approximate 'generative' birth and death rates to the estimated birth and death rates.

Approximation per lineage. First, we consider the time intervals between subsequent cell divisions and deaths. In the birth-death model, both events follow a Poisson process, so the expected waiting time for the next birth event is exponentially distributed with parameter β, and the expected waiting time for the next death event is exponentially distributed with parameter δ. The expected waiting time for the next event of any kind is exponential with parameter $\beta + \delta$ and mean $\frac{1}{\beta+\delta}$. Then, a cell divides with probability $\frac{\beta}{\beta+\delta}$ or dies with probability $\frac{\delta}{\beta+\delta}$.

In our synchronous tree simulations, cell divisions and deaths occur after fixed time intervals. We now approximate this process with 'generative' birth and death rates. If n denotes the number of cell division time points within $[0, T]$, the time interval between subsequent divisions is $\frac{T}{n+1}$. Taking the inverse results in a per-lineage event rate of $\beta + \delta \approx \frac{n+1}{T}$. Given a death probability p_d and noting that $\delta = p_d \cdot (\beta + \delta)$, the birth and death rates are approximated by the 'generative' rates

$$\beta \approx (1 - p_d) \cdot \frac{n+1}{T} \quad \text{and} \quad \delta \approx p_d \cdot \frac{n+1}{T} \tag{1}$$

In birth-only synchronous trees, $p_d = 0$, and thus the rates are simply $\beta \approx \frac{n+1}{T}$ and $\delta = 0$.

Approximation on population level. In our second approximation, we derive the birth and death rates from the expected number of cells at the end of the tree generating process, at time T. Under the birth-death model, the expected number of living individuals (cells) after arbitrary time t is

$$N(t) = e^{(\beta - \delta) \cdot t}$$

assuming that the population consists initially of one cell [19]. Hence, the number of cells grows exponentially through time as long as $\beta > \delta$. Given a sampling probability ρ, on average $\rho \cdot e^{(\beta - \delta) \cdot T}$ cells are sampled at time T.

Under synchronous cell divisions and sampling fraction ρ, the expected number of cells after n division time points is $\rho \cdot 2^n$. Hence, for birth-only trees, in which $\delta = 0$, we can approximate

$$\rho \cdot e^{\beta \cdot T} \approx \rho \cdot 2^n$$

meaning the 'generative' birth rate is $\beta \approx \frac{n}{T} \ln 2$.

In synchronous trees with cell death, instead of considering the expected number of surviving cell lineages, we used the event rate $\beta+\delta$ to compare the expected number of cells $\rho \cdot 2^n$ to the birth-death model, treating deaths as birth events in the calculation. We arrive at the equation

$$\rho \cdot e^{(\beta+\delta)\cdot T} \approx \rho \cdot 2^n \quad \Longleftrightarrow \quad \beta + \delta \approx \frac{n}{T} \ln 2$$

For a given death probability p_d, approximated birth and death rates on the population level are thus

$$\beta \approx (1 - p_d) \cdot \frac{n}{T} \ln 2 \quad \text{and} \quad \delta \approx p_d \cdot \frac{n}{T} \ln 2 \tag{2}$$

Again, for synchronous trees with birth only, $\beta \approx \frac{n}{T} \ln 2$ and $\delta = 0$. We use these approximations as the 'generative' birth and death rates.

Altogether, for $n \to \infty$, the population level birth and death rates differ by a factor of $\ln 2$ from the per lineage rates. We suggest that the difference stems from the assumed base in the population growth functions. The per lineage rates would match with the population level rates, if the population growth over time was exponential with base e, as in the birth-death model. However, when cell divisions are synchronous, the population growth over time is slower - exponential with base 2. Multiplying the per lineage rates with factor $\ln 2$ changes the base in the growth function from e to 2.

Applying the inference framework to our simulated data, we have observed that the posterior medians consistently approached the birth and death rates approximated on the population level. Therefore, we treated those as the 'true' rates in the evaluation. Our results further imply that birth rates inferred under the birth-death model can be divided by $\ln 2$ to estimate the doubling time of cell populations growing by synchronous and regular cell divisions in empirical analyses.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Code to rerun the simulations and analysis and generate the figures is available at https://github.com/pilarskj/celldev.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors thank ETH Zürich for funding. This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement no. 101001077 (PhyCogy).

Authors' contributions

JP, SS and TS designed the study. JP performed the simulations, analyzed the data, and was a major contributor in writing the manuscript. SS and TS provided conceptual advice and analysis tools. All authors read and approved the final manuscript.

Acknowledgments

The authors thank Antoine Zwaans, Nicola Mulberry and other members of the cEvo group at ETH Zurich for helpful discussions.

References

[1] McKenna A, Gagnon JA. Recording development with single cell dynamic lineage tracing. Development. 2019;146(12):dev169730. https://doi.org/10.1242/dev.169730.

[2] Wu SHS, Lee JH, Koo BK. Lineage tracing: computational reconstruction goes beyond the limit of imaging. Molecules and Cells. 2019;42(2):104. https://doi.org/10.14348/molcells.2019.0006.

[3] Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nature Reviews Genetics. 2020;21(7):410–427. https://doi.org/10.1038/s41576-020-0223-2.

[4] Chen C, Liao Y, Peng G. Connecting past and present: single-cell lineage tracing. Protein & Cell. 2022;13(11):790–807. https://doi.org/10.1007/s13238-022-00913-7.

[5] Askary A, Chen W, Choi J, Du LY, Elowitz MB, Gagnon JA, et al. The lives of cells, recorded. Nature Reviews Genetics. 2024;p. 1–20. https://doi.org/10.1038/s41576-024-00788-w.

[6] Stadler T, Pybus OG, Stumpf MP. Phylodynamics for cell biologists. Science. 2021;371(6526):eaah6266. https://doi.org/10.1126/science.aah6266.

[7] Alemany A, Florescu M, Baron CS, Peterson-Maduro J, Van Oudenaarden A. Whole-organism clone tracing using single-cell sequencing. Nature. 2018;556(7699):108–112. https://doi.org/10.1038/nature25969.

[8] Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nature Biotechnology. 2018;36(5):469–473. https://doi.org/10.1038/nbt.4124.

[9] Kalhor R, Mali P, Church GM. Rapidly evolving homing CRISPR barcodes. Nature Methods. 2017;14(2):195–200. https://doi.org/10.1038/nmeth.4108.

[10] He Z, Maynard A, Jain A, Gerber T, Petri R, Lin HC, et al. Lineage recording in human cerebral organoids. Nature Methods. 2022;19(1):90–99. https://doi.org/10.1038/s41592-021-01344-8.

[11] Choi J, Chen W, Minkina A, Chardon FM, Suiter CC, Regalado SG, et al. A time-resolved, multi-symbol molecular recorder via sequential genome editing. Nature. 2022;608(7921):98–107. https://doi.org/10.1038/s41586-022-04922-8.

[12] Loveless TB, Carlson CK, Dentzel Helmy CA, Hu VJ, Ross SK, Demelo MC, et al. Open-ended molecular recording of sequential cellular events into DNA. Nature Chemical Biology. 2024;p. 1–10. https://doi.org/10.1038/s41589-024-01764-5.

[13] Salvador-Martínez I, Grillo M, Averof M, Telford MJ. Is it possible to reconstruct an accurate cell lineage using CRISPR recorders? eLife. 2019;8:e40292. https://doi.org/10.7554/eLife.40292.

[14] Wang R, Zhang R, Khodaverdian A, Yosef N. Theoretical guarantees for phylogeny inference from single-cell lineage tracing. Proceedings of the National Academy of Sciences. 2023;120(12):e2203352120. https://doi.org/10.1073/pnas.2203352120.

[15] Mulberry N, Stadler T. Strategies for resolving cellular phylogenies from sequential lineage tracing data; 2024. [Manuscript submitted for publication].

[16] Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology. 2019;15(4):e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

[17] Seidel S, Stadler T. TiDeTree: A Bayesian phylogenetic framework to estimate single-cell trees and population dynamic parameters from genetic lineage tracing data. Proceedings of the Royal Society B. 2022;289(1986):20221844. https://doi.org/10.1098/rspb.2022.1844.

[18] Seidel S, Zwaans A, Regalado S, Choi J, Shendure J, Stadler T. Sciphy: A Bayesian phylogenetic framework using sequential genetic lineage tracing data. bioRxiv. 2024;p. 2024–10. https://doi.org/10.1101/2024.10.01.615771.

[19] Kendall DG. On the Generalized "Birth-and-Death" Process. The Annals of Mathematical Statistics. 1948;19(1):1–15. https://doi.org/10.1214/aoms/1177730285.

[20] Stadler T. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. Journal of Theoretical Biology. 2009;261(1):58–66. https://doi.org/10.1016/j.jtbi.2009.07.018.

[21] Stadler T, Bonhoeffer S. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013;368(1614):20120198. https://doi.org/10.1098/rstb.2012.0198.

[22] Robinson DF, Foulds LR. Comparison of weighted labelled trees. In: Horadam AF, Wallis WD, editors. Combinatorial Mathematics VI. Berlin, Heidelberg: Springer Berlin Heidelberg; 1979. p. 119–126. Available from: https://doi.org/

10.1007/BFb0102690.

[23] Robinson DF, Foulds LR. Comparison of phylogenetic trees. Mathematical Biosciences. 1981;53(1-2):131–147. https://doi.org/10.1016/0025-5564(81)90043-2.

[24] Nee S, Holmes EC, May RM, Harvey PH. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 1994;344(1307):77–82. https://doi.org/10.1098/rstb.1994.0054.

[25] Paradis E. Analysis of diversification: combining phylogenetic and taxonomic data. Proceedings of the Royal Society of London Series B: Biological Sciences. 2003;270(1532):2499–2505. https://doi.org/10.1098/rspb.2003.2513.

[26] Helmstetter AJ, Glemin S, Käfer J, Zenil-Ferguson R, Sauquet H, de Boer H, et al. Pulled diversification rates, lineages-through-time plots, and modern macroevolutionary modeling. Systematic Biology. 2022;71(3):758–773. https://doi.org/10.1093/sysbio/syab083.

[27] Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35(3):526–528. https://doi.org/10.1093/bioinformatics/bty633.

[28] Stadler T. Simulating trees with a fixed number of extant species. Systematic Biology. 2011;60(5):676–684. https://doi.org/10.1093/sysbio/syr029.

[29] Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. Molecular Biology and Evolution. 2016;33(8):2102–2116. https://doi.org/10.1093/molbev/msw064.

[30] Scire J, Barido-Sottani J, Kühnert D, Vaughan TG, Stadler T. Robust phylodynamic analysis of genetic sequencing data from structured populations. Viruses. 2022;14(8):1648. https://doi.org/10.3390/v14081648.

[31] Vaughan TG, Stadler T. Bayesian phylodynamic inference of multi-type population trajectories using genomic data. bioRxiv. 2024;https://doi.org/10.1101/2024.11.26.625381.

[32] Packer JS, Zhu Q, Huynh C, Sivaramakrishnan P, Preston E, Dueck H, et al. A lineage-resolved molecular atlas of C. elegans embryogenesis at single-cell resolution. Science. 2019;365(6459):eaax1971. https://doi.org/10.1126/science.aax1971.

[33] Hormoz S, Singer ZS, Linton JM, Antebi YE, Shraiman BI, Elowitz MB. Inferring cell-state transition dynamics from lineage trees and endpoint single-cell measurements. Cell Systems. 2016;3(5):419–433. https://doi.org/10.1016/j.cels.2016.10.015.

[34] Chow KHK, Budde MW, Granados AA, Cabrera M, Yoon S, Cho S, et al. Imaging cell lineage with a synthetic digital recording system. Science. 2021;372(6538):eabb3099. https://doi.org/10.1126/science.abb3099.

[35] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al. Sustainable data analysis with Snakemake. F1000Research. 2021;10. https://doi.org/10.12688/f1000research.29032.2.

[36] Stadler T, Magnus C, Vaughan T, Barido-Sottani J, Bošková V, Huisman JS, et al. Decoding Genomes: From Sequences to Phylodynamics. Pečerska J, editor. ETH Zurich; 2024. Available from: https://decodinggenomes.org.

[37] Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proceedings of the National Academy of Sciences. 2013;110(1):228–233. https://doi.org/10.1073/pnas.1207965110.

[38] Nye TM, Lio P, Gilks WR. A novel algorithm and web-based tool for comparing two alternative phylogenetic trees. Bioinformatics. 2006;22(1):117–119. https://doi.org/10.1093/bioinformatics/bti720.

[39] Smith MR. Information theoretic generalized Robinson–Foulds metrics for comparing phylogenetic trees. Bioinformatics. 2020;36(20):5007–5013. https://doi.org/10.1093/bioinformatics/btaa614.

[40] Smith MR.: TreeDist: Distances between Phylogenetic Trees. R package version 2.9.1. Available from: https://doi.org/10.5281/zenodo.3528124.

[41] Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27(4):592–593. https://doi.org/10.1093/bioinformatics/btq706.

[42] Wang LG, Lam TTY, Xu S, Dai Z, Zhou L, Feng T, et al. treeio: an R package for phylogenetic tree input and output with richly annotated and associated data. Molecular Biology and Evolution. 2020;37:599–603. https://doi.org/10.1093/molbev/msz240.

[43] Yu G. Data Integration, Manipulation and Visualization of Phylogenetic Trees. 1st ed. Chapman and Hall/CRC; 2022. Available from: https://yulab-smu.top/treedata-book/.

[44] Bilderbeek RJ, Etienne RS. babette: BEAUti 2, BEAST 2 and Tracer for R. Methods in Ecology and Evolution. 2008;https://doi.org/10.1111/2041-210X.13032.

[45] Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686. https://doi.org/10.21105/joss.01686.

Appendix A Supplementary Tables

Table A1: Edit probabilities for simulating CRISPR lineage recordings

Index k	Scarring probability s_k	Insert probability i_k
1	0.0021	0.0109
2	0.0078	0.0266
3	0.0812	0.0690
4	0.0105	0.0018
5	0.0039	0.0707
6	0.0309	0.0803
7	0.0025	0.2657
8	0.0513	0.0038
9	0.0021	0.0445
10	0.0731	0.0106
11	0.0441	0.1265
12	0.0038	0.1935
13	0.0195	0.0961
14	0.0455	
15	0.0072	
16	0.0210	
17	0.0136	
18	0.0116	
19	0.0065	
20	0.0014	
21	0.0225	
22	0.0056	
23	0.0067	
24	0.0370	
25	0.0189	
26	0.0281	
27	0.0070	
28	0.0132	
29	0.0002	
30	0.0099	
31	0.0397	
32	0.0190	
33	0.0099	
34	0.0017	
35	0.0160	
36	0.0359	
37	0.0001	
38	0.0063	
39	0.0006	
40	0.0494	
41	0.0396	
42	0.0027	
43	0.0148	
44	0.0391	
45	0.0421	
46	0.0360	
47	0.0032	
48	0.0242	
49	0.0004	
50	0.0306	

Appendix B Supplementary Figures

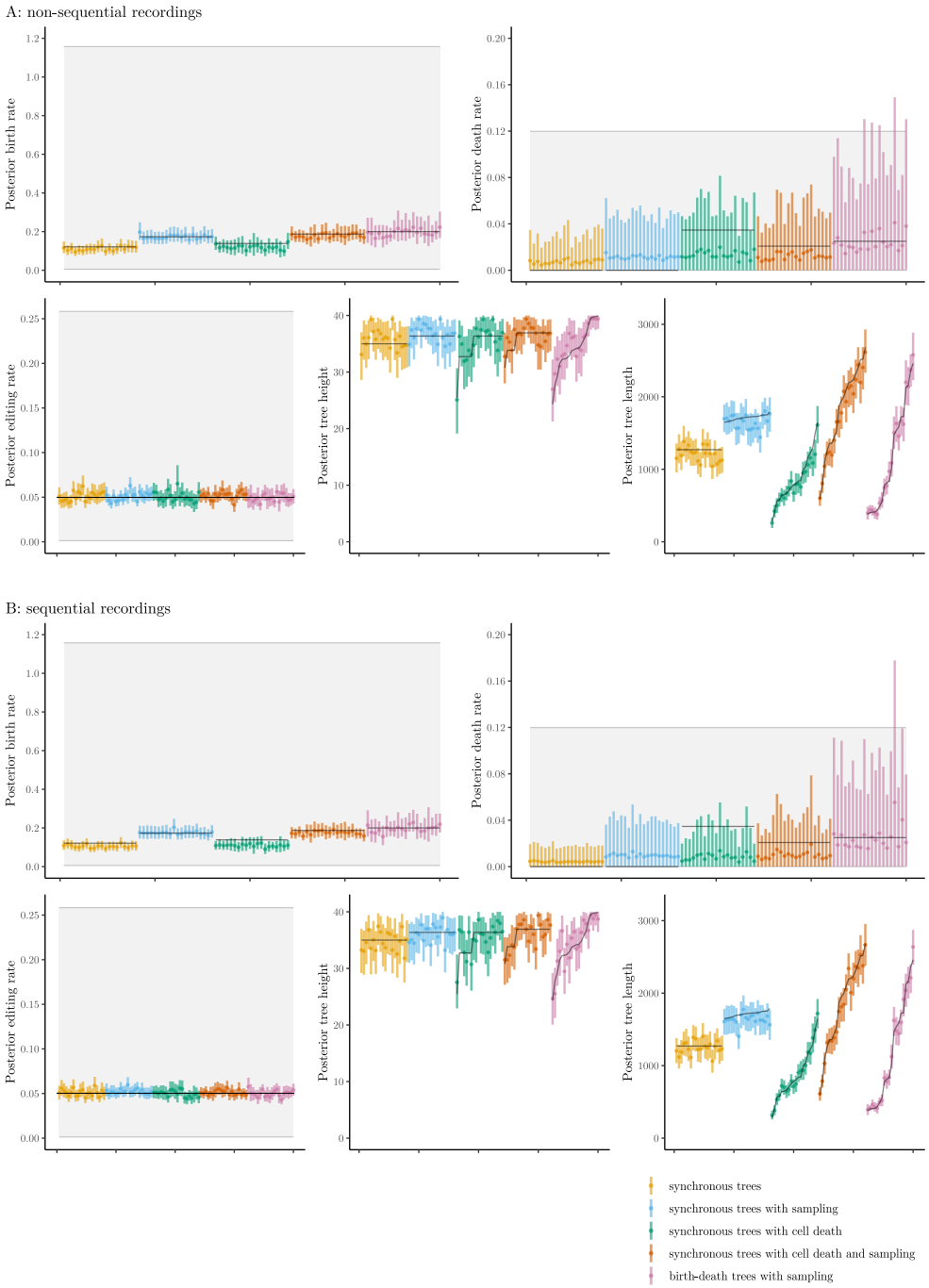

Fig. B1: Parameters inferred from (A) non-sequential and (B) sequential CRISPR lineage recordings at baseline. In each graph, points indicate the medians and bars show the 95% HPD intervals of the parameter estimates per simulation (x-axis). Colors represent the different tree generating processes. Black lines display the true parameters. Grey shaded areas represent the 95% HPD intervals of prior distributions.

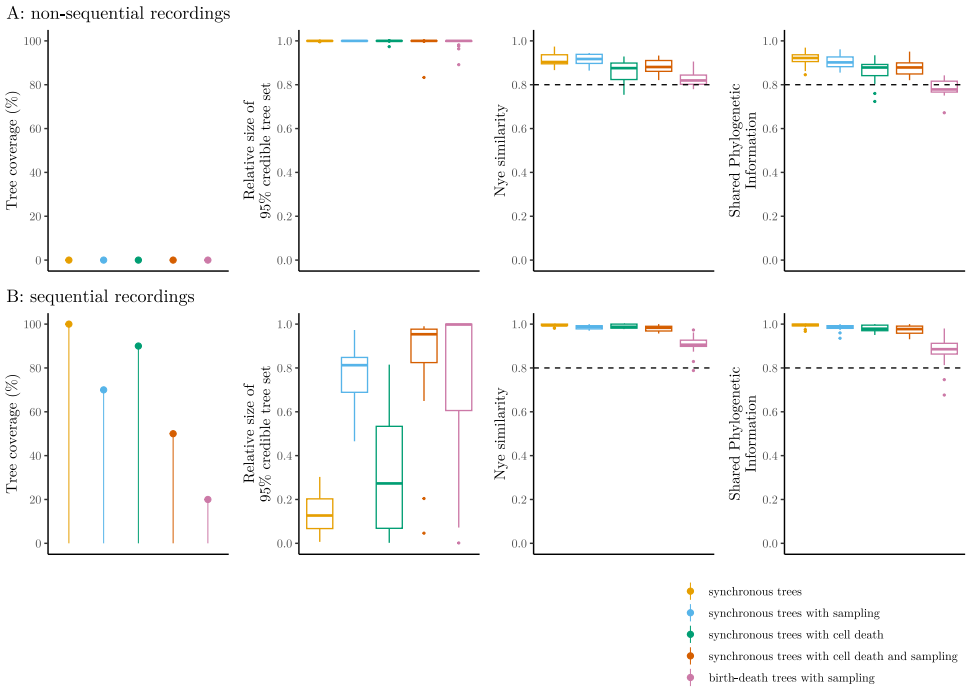

Fig. B2: Additional metrics for assessing phylogeny reconstruction from (A) non-sequential and (B) sequential CRISPR lineage recordings. The graphs show from left to right: the tree coverage, the relative size of the 95% credible tree set, Nye similarity, and Shared Phylogenetic Information between the inferred and true trees. The metrics are summarized across 20 simulations per tree generating process (indicated by colors). Dashed lines display thresholds to facilitate visual comparability.

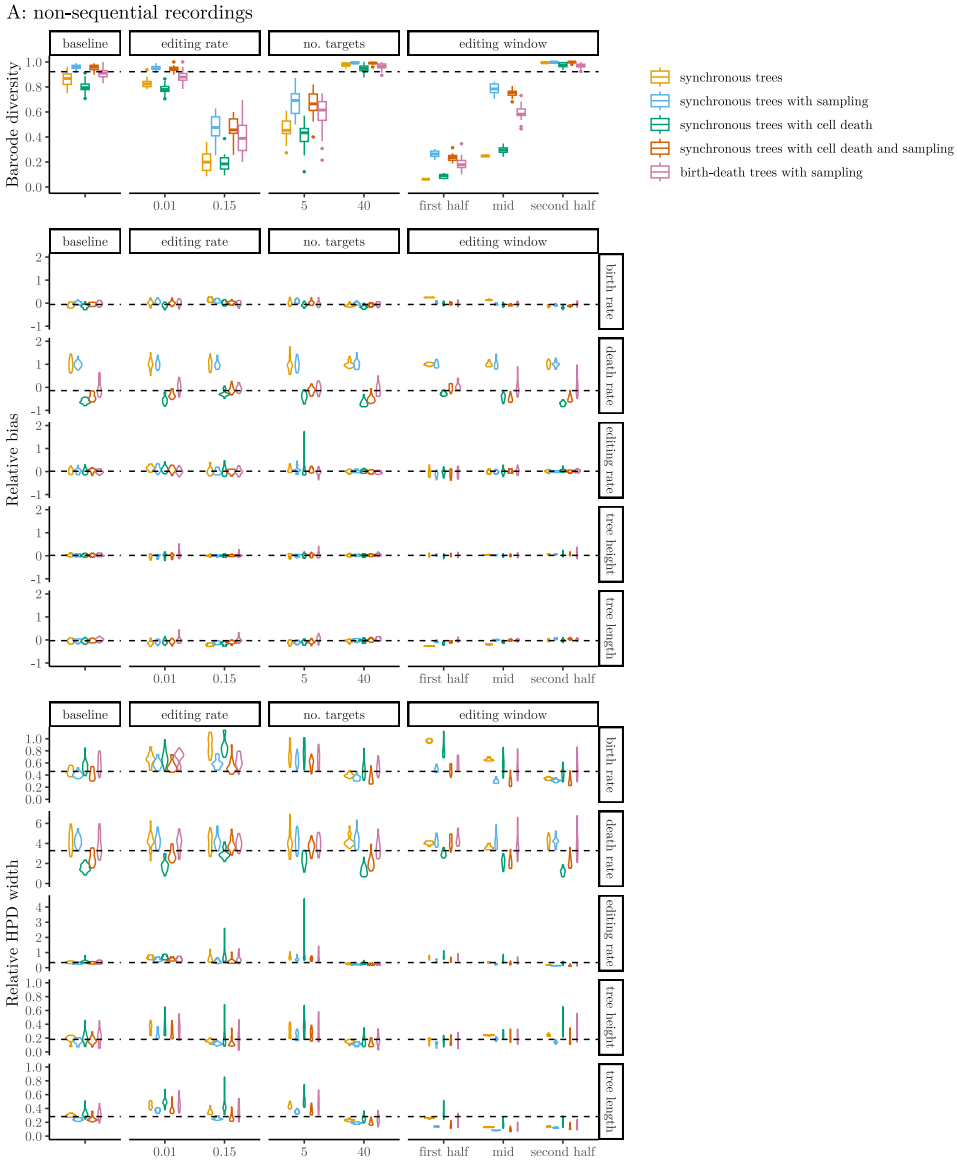

Fig. B3: Additional metrics for characterizing barcodes and assessing inference performance from (A) non-sequential and (B) sequential CRISPR lineage recordings with varying experimental parameters. Vertical facets indicate the varied experimental parameters, horizontal facets indicate the inferred parameters. At baseline, the editing rate was 0.05, the recorders carried 20 targets or 20 tapes of length 5, respectively, and editing occurred throughout the entire experiment. Dashed lines display the baseline medians (across trees). The graphs show from top to bottom: barcode diversity, the relative bias, and relative HPD widths of the inferred parameters. The metrics are summarized across 20 simulations per tree generating process (indicated by colors) per setting.

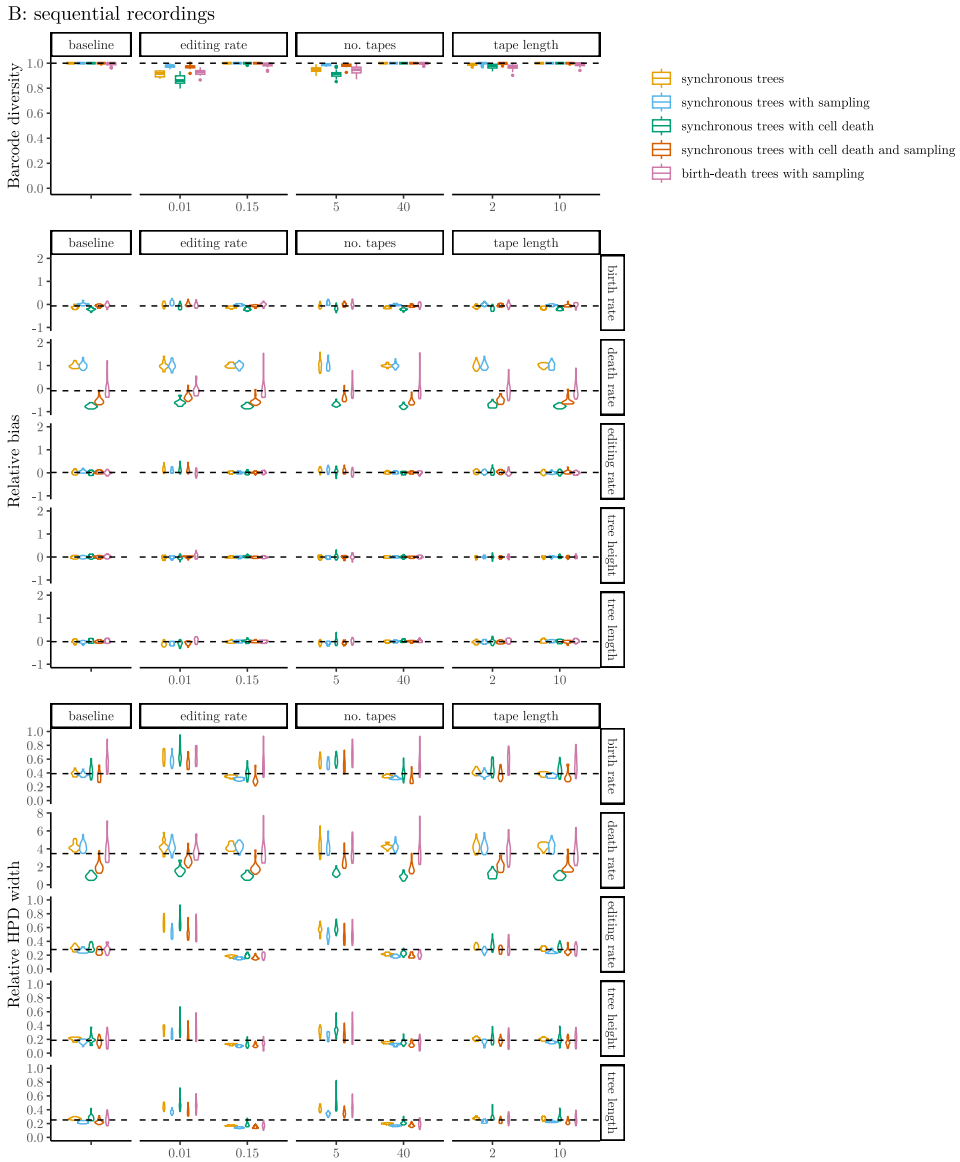

Fig. B3: Additional metrics for characterizing barcodes and assessing inference performance from (A) non-sequential and (B) sequential CRISPR lineage recordings with varying experimental parameters (cont.)

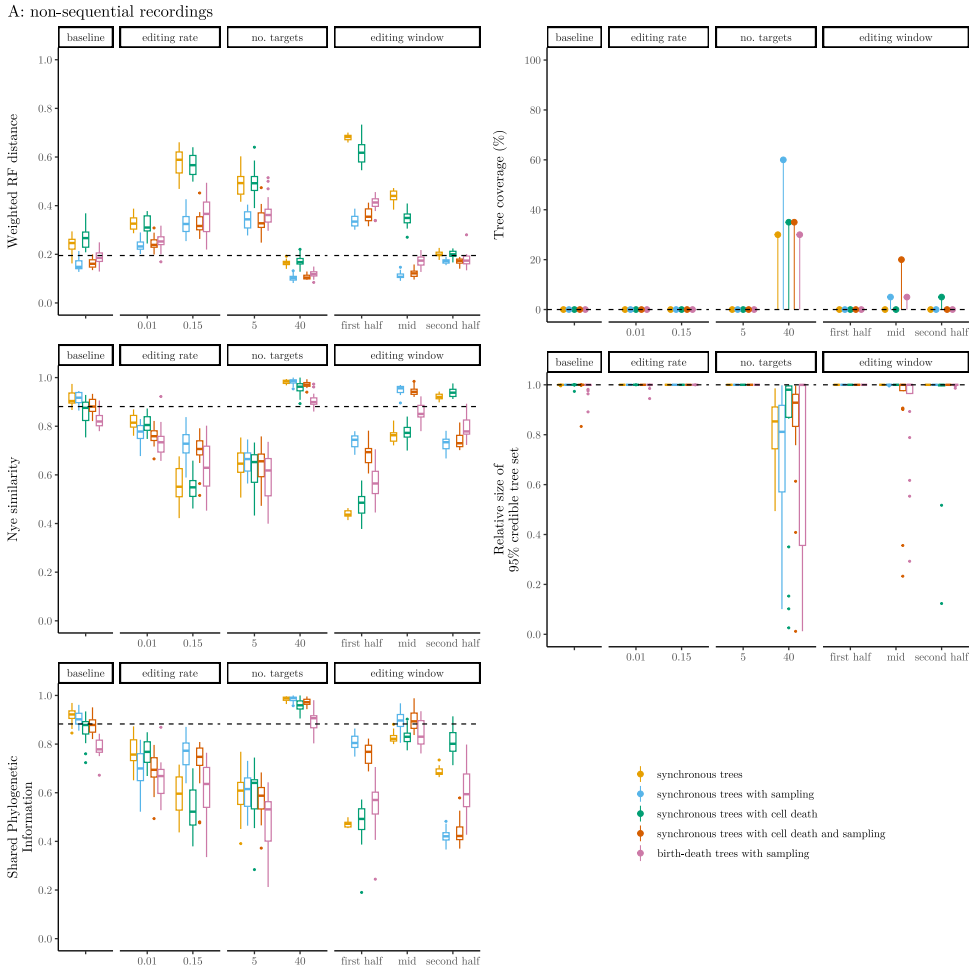

Fig. B4: Additional metrics for assessing phylogeny reconstruction from (A) non-sequential and (B) sequential CRISPR lineage recordings with varying experimental parameters. Facets indicate the varied experimental parameters. At baseline, the editing rate was 0.05, the recorders carried 20 targets or 20 tapes of length 5, respectively, and editing occurred throughout the entire experiment. Dashed lines display the baseline medians (across trees). The graphs in the left column show the tree coverage (top) and the relative size of the 95% credible tree set (bottom). The graphs in the right column show from top to bottom: weighted RF distance, Nye similarity, and Shared Phylogenetic Information between the inferred and true trees. The metrics are summarized across 20 simulations per tree generating process (indicated by colors) per setting.

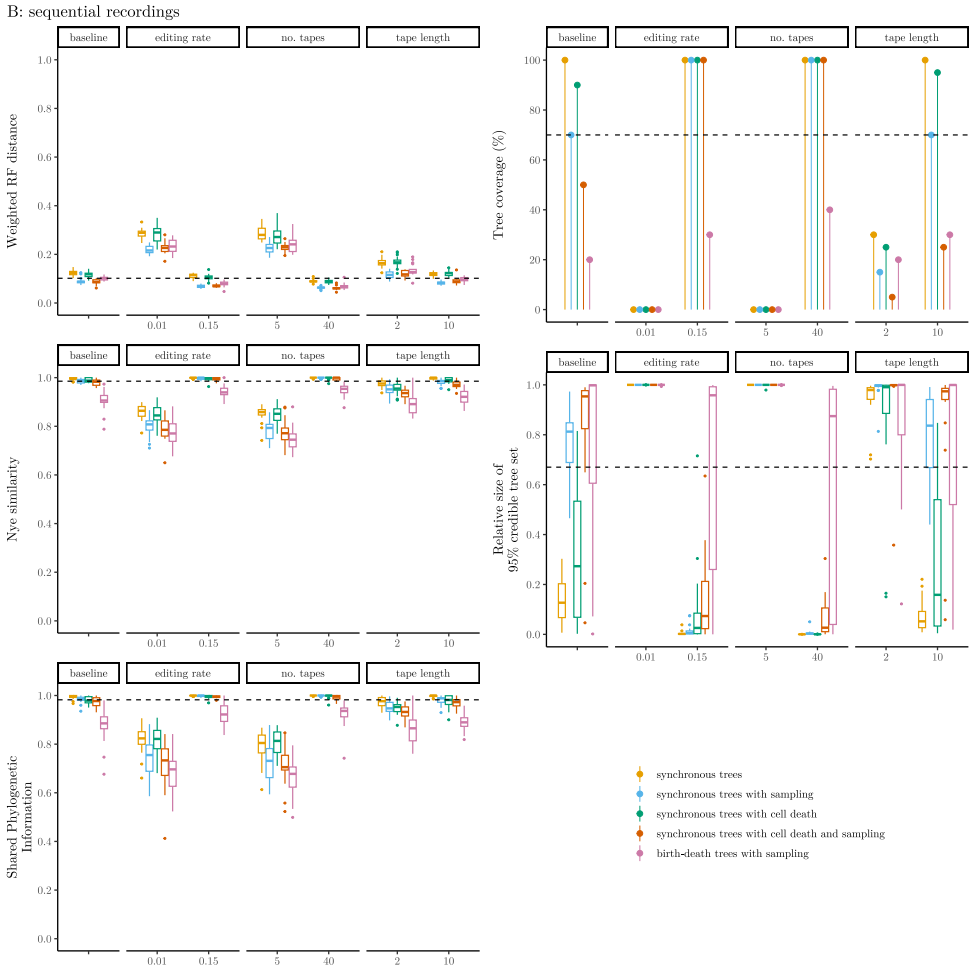

Fig. B4: Additional metrics for assessing phylogeny reconstruction from (A) non-sequential and (B) sequential CRISPR lineage recordings with varying experimental parameters (cont.)

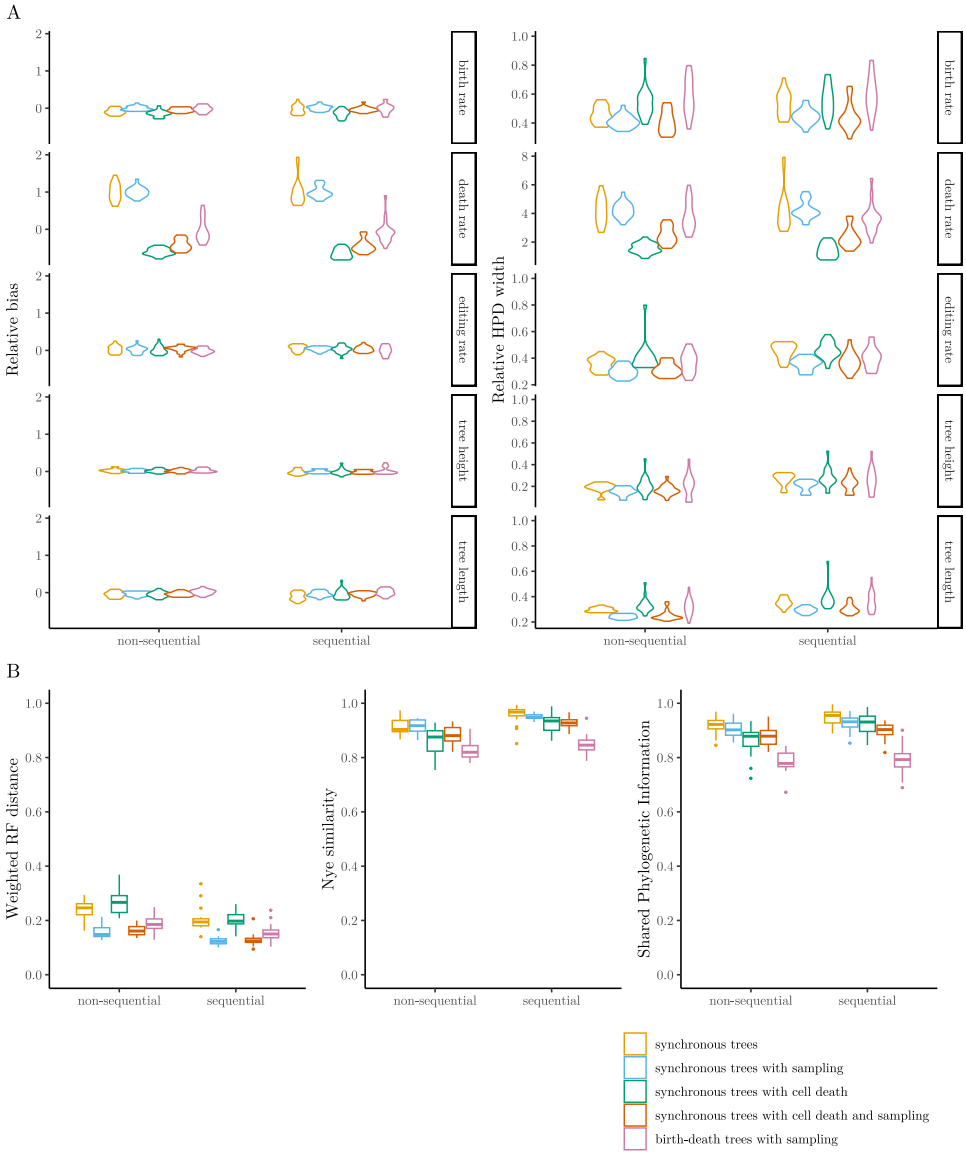

Fig. B5: Comparison of the sequential and non-sequential accumulation of edits in CRISPR recorders. Panel (A) shows the relative bias (left) and relative HPD width (right) of the inferred parameters. Panel (B) shows from left to right: the weighted RF distance, Nye similarity and Shared Phylogenetic Information between the inferred and true trees. The results are summarized across 20 simulations per tree generating process (indicated by colors) per setting.

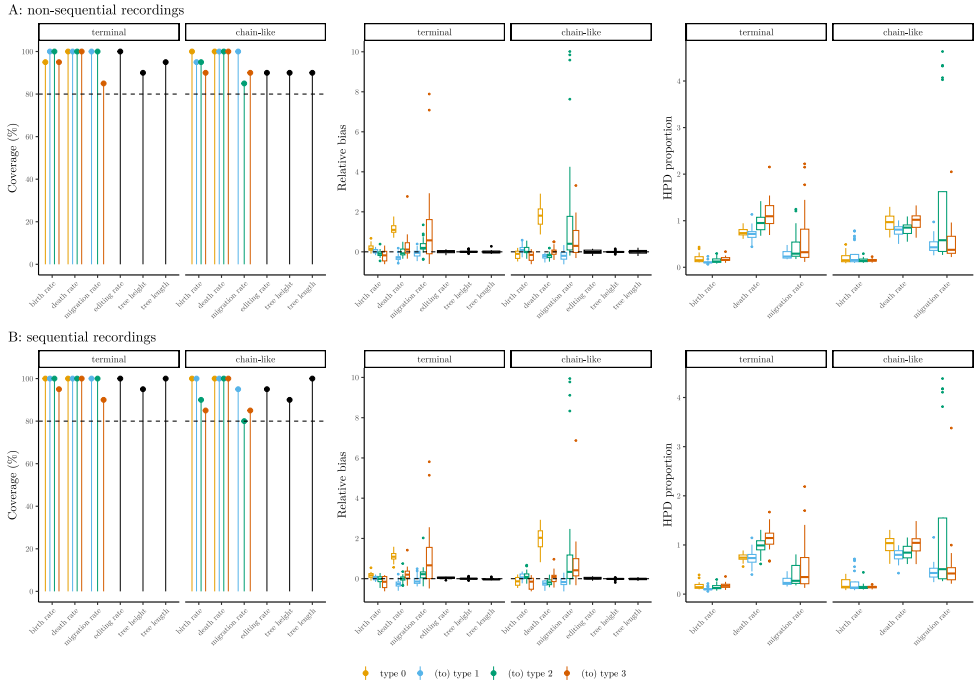

Fig. B6: Inference performance on (A) non-sequential and (B) sequential CRISPR lineage recordings simulated along multi-type birth-death trees. The graphs show from left to right: the coverage, relative bias and HPD proportion of the inferred parameters. Colors indicate for which cell type the parameters hold. The inference results are summarized across 20 simulations per cell type transition dynamics. Dashed lines display thresholds for comparability.