PNAS
www.pnas.org

Supplementary Information for:

Heritable changes in division speed accompany the diversification of single T cell fate

Marten Plambeck[a,1], Atefeh Kazeroonian[a,1], Dirk Loeffler[b], Lorenz Kretschmer[a], Ciro Salinno[a], Timm Schroeder[b], Dirk H. Busch[a,2,3], Michael Flossdorf[a,2,3], Veit R. Buchholz[a,2,3]

[a] Institute for Medical Microbiology, Immunology and Hygiene, Technical University of Munich (TUM), Munich 81675, Germany
[b] Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich (ETH Zurich), 4058 Basel, Switzerland
[1] These authors contributed equally.
[2] D.H.B., M.F. and V.R.B. jointly supervised this work.
[3]To whom correspondence may be addressed:
Email: dirk.busch@tum.de, michael.flossdorf@tum.de, and veit.buchholz@tum.de

This PDF file includes:

Other supplementary materials for this manuscript include the following:

Supplementary Methods

0. Introduction

We developed a Bayesian inference framework for hypothesis-driven analysis of cell lineage trees. This inference scheme tests different hypotheses to infer underlying pathways of diversification in single-cell-derived populations. For a given model hypothesis, it uses lineage (family) trees obtained from live-cell imaging experiments to infer the model parameters. We used this framework to test various hypotheses about the existence of subsets with distinct inter-division time statistics in T cell family trees expanded *in vitro* and the corresponding diversification pathways. The scheme presented here is specifically designed for bright-field microscopy imaging data where no direct phenotypic measurements are available. However, it can be generalized to analyze live-cell imaging with phenotypic marker expression data. In the following, we describe different components of our computational analysis. In Sections 1 and 2, we outline our Bayesian inference scheme and Bayesian model comparison. In Section 3, we discuss the details of different model hypotheses we tested for analyzing T cell family trees. Section 4 briefly outlines the statistical features of the data that we aim to capture with mathematical modeling. Finally, Section 5 describes the simulation studies that were used to examine how well mathematical models recapitulated the statistical properties of the data.

1. Bayesian model inference scheme for the analysis of live-cell imaging data

Our Bayesian framework utilizes a Markov Chain Monte Carlo (MCMC) scheme with latent variables (Wilkinson, 2009; Wilkinson, 2011) tailored to tree-structured data, incorporating familial relationships of individual cells within a family tree. This framework takes family trees obtained via live-cell imaging, as well as a model hypothesis, as input and returns the posterior distribution of model parameters as output. It enables model comparison and hypothesis testing through the approximation of model evidences and corresponding Bayes factors. Different steps of our inference scheme are described in the following.

1.1. Input data. Family trees generated by The Tracking Tool (Hilsenbeck et al., 2016) from live-cell imaging experiments form the input data of the inference scheme (see Materials and Methods of the main text). This data encodes two types of information for every cell in a family tree: 1) the lifetime of that cell, and 2) the index of the cell within the family tree which yields familial relationships. Dataset D consists of the lifetimes of individual cells across different trees:

$$D = \{\tilde{t}_{im} | i = 1, \dots, n_{\text{tree}}, m = 1, \dots, n_{\text{cell}_i}\} \tag{1.1}$$

where n_{tree} is the total number of trees and n_{cell_i} is the total number of cells in tree i and $\tilde{t}_{im}$ is the lifetime of cell m in tree i. If cell m has divided in the course of the experiment, $\tilde{t}_{im}$ denotes its inter-division time; otherwise, it represents the time until which this cell had not divided indicating $\tilde{t}_{im}$ as a lower bound for the inter-division time. We assume that the interdivision times are log-normally distributed independent random variables. We further assume a minimum cell-cycle length τ to exclude un-physiologically short inter-division times. To enable robust numerical calculations, we first subtract the threshold τ and then log-transform the data in D to arrive at

$$\boldsymbol{y} = \left\{ t_{im} \middle| i = 1, \dots, n_{\text{tree}}, m = 1, \dots, n_{\text{cell}_i} \right\}, \quad \text{with } t_{im} = \log\left(\tilde{t}_{im} - \tau\right) \tag{1.2}$$

The dataset $\boldsymbol{y}$ is then used to infer the model structure and parameters.

1.2. Input model hypothesis. The inference scheme relies on a model hypothesis for the computational analysis of the data. This model hypothesis represents assumptions about the existence of distinct subsets in the population and the diversification pathways connecting these subsets. We used our inference scheme to analyze family trees in the absence of surface marker expression measurements. Therefore, no subsets defined based on marker combinations were considered in this study. Instead, subsets of cells which possess distinct inter-division time statistics were considered. However, if direct phenotypic measurements are available, they can be readily incorporated into the model hypothesis and the inference scheme.

In Section 3, we provide a detailed description of model topologies considered in this study. In all of the hypotheses, we assume that the inter-division times of cells belonging to a specific subset follow a log-normal distribution with specific but unknown mean and coefficient of variation. Other distribution assumptions can easily replace the lognormal assumption, and can be handled in the inference scheme the same way. Each model hypothesis is characterized by a set of unknown model parameters $\boldsymbol{\theta}$ such as the mean and coefficient of variation of the inter-division time distributions as well as the transition probabilities between different subsets. Furthermore, the subset that is assigned to a cell is modeled as a latent variable due to the unavailability of direct observation.

1.3. Bayesian Inference scheme. We use a Markov Chain Monte Carlo (MCMC) approach to generate samples from the posterior distribution of the model parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$. For all parameters, we assume uniform prior distributions with specified upper and lower bounds. Different steps of the iterative sampling scheme are described in the following sections, and a pseudo-code is given in Algorithm 1.

1.3.1. MCMC sampling for models with latent variables. As mentioned earlier, the underlying subset that a cell in the family tree belongs to, is not directly observed and therefore needs to be modeled as a latent variable. For a model with latent variables, we have the following factorization

$$\pi(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y}) = \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}) \tag{1.3}$$

where, $\boldsymbol{\theta}$ is the vector of model parameters, $\boldsymbol{x}$ is the vector of latent variables and $\boldsymbol{y}$ is the data. Here, $\boldsymbol{x}$ and $\boldsymbol{y}$ are respectively the assigned subsets and the lifetimes of individual cells. Since we are only interested in the posterior distribution of the model parameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$, we can use an arbitrary proposal $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$, to target this posterior using the following acceptance ratio:

$$A = \frac{\pi(\boldsymbol{\theta}^*)\, \pi(\boldsymbol{y}|\boldsymbol{\theta}^*)\, q(\boldsymbol{\theta}^*, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})\, \pi(\boldsymbol{y}|\boldsymbol{\theta})\, q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)} \tag{1.4}$$

In order to calculate the marginal likelihood of the data given the parameters $\pi(\boldsymbol{y}|\boldsymbol{\theta})$, we must integrate out the latent variables:

$$\pi(\boldsymbol{y}|\boldsymbol{\theta}) = \int_{\boldsymbol{X}} \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\, \pi(\boldsymbol{x}|\boldsymbol{\theta})\, d\boldsymbol{x} \tag{1.5}$$

Due to the intractability of the analytical calculation of this integral, we aim to use a Monte Carlo approximation of it

$$\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{n_x} \sum_{k=1}^{n_x} \pi(\boldsymbol{y}|\boldsymbol{x}_k, \boldsymbol{\theta}) \tag{1.6}$$

It can be shown that since this is an unbiased estimator of the likelihood, if we plug this approximation into the acceptance ratio (Eq. 1.4), we will exactly target the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ (Wilkinson, 2011):

$$A = \frac{\pi(\boldsymbol{\theta}^*)\ \hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}^*)\ q(\boldsymbol{\theta}^*, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})\ \hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta})\ q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)} \tag{1.7}$$

Here, $\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta})$ denotes our unbiased estimate of the marginal likelihood $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ obtained by the Monte Carlo integration above. We assume symmetric prior and proposal distributions, which cancels out the terms $\pi(\boldsymbol{\theta}^*)\ q(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta})\ q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ in the numerator and denominator of Eq. (1.7), resulting in the acceptance ratio

$$A = \frac{\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}^*)}{\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta})} \tag{1.8}$$

We propose parameter sets according to the proposal $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$, and based on the acceptance ratio Eq. 1.8, obtain samples from the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. In the study of T cell family trees, we used random walk and Gaussian proposals.

1.3.2. Unbiased estimates of the likelihood $\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}^*)$. We use the following Monte Carlo approximation based on the latent variable samples to obtain an unbiased estimate of the likelihood:

$$\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{n_x} \sum_{k=1}^{n_x} \pi(\boldsymbol{y}|\boldsymbol{x}_k, \boldsymbol{\theta}) \tag{1.9}$$

Since our data is comprised of several lineage trees, which are independent from each other, the data likelihood $\pi(\boldsymbol{y}|\boldsymbol{\theta})$ can be written as

$$\pi(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n_{\text{tree}}} \pi(\boldsymbol{y}_i|\boldsymbol{\theta}) \tag{1.10}$$

where $\boldsymbol{y}_i$ is the data belonging to the i^{th} lineage tree and $\pi(\boldsymbol{y}_i|\boldsymbol{\theta})$ is its likelihood. Moreover, the latent variables of every tree, i.e. the subsets of cells within that tree, are independent from other trees. Thus, we can apply the Monte Carlo approximation on every tree individually and reconstruct an unbiased estimate of the overall likelihood based on individual tree likelihoods. In this way we have

$$\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n_{\text{tree}}} \hat{\pi}(\boldsymbol{y}_i|\boldsymbol{\theta}) \tag{1.11}$$

where $\hat{\pi}(\boldsymbol{y}_i|\boldsymbol{\theta})$ is the unbiased estimate of the likelihood of the i^{th} tree. For numerical properties, it is favorable to work with the log-likelihood. Taking the log-transformation of Eq. 1.11, we obtain the following for the log-likelihood:

$$\hat{J}(\boldsymbol{\theta}) = \log\left(\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta})\right) = \sum_{i=1}^{n_{\text{tree}}} \log(\hat{\pi}(\boldsymbol{y}_i|\boldsymbol{\theta})). \tag{1.12}$$

We next need to sample latent variables in every tree to calculate the log-likelihoods $\log(\hat{\pi}(\boldsymbol{y}_i|\boldsymbol{\theta}))$.

1.3.3. Two-step sampling of the latent variables. To construct an unbiased estimate of the likelihood of a tree given the parameters, $\hat{\pi}(\boldsymbol{y}_i|\boldsymbol{\theta})$, we generate several samples of the latent variables, for each of which the likelihood can be directly evaluated. The number of possible configurations for latent variables increases exponentially with the number of cells. To increase the

efficiency and the acceptance ratio of our sampling scheme, we divide each tree into nine "subtrees" (Fig. S28). We perform sample generation and likelihood approximation for these subtrees in a modular fashion. The first subtree, denoted as T_0, includes all cells belonging to generations 1, 2 and 3. The remaining subtrees, denoted as $T_1, T_2, \dots, T_8$, are subtrees descending from the eight third-generation cells: T_1 includes all cells descending from cell c_1, T_2 includes all cells descending from cell c_2 and so forth (Fig. S28). In this way, cells $c_1, c_2, \dots, c_8$, which themselves belong to T_0, form the founder cells for subtrees $T_1, T_2, \dots, T_8$. The dependency between these subtrees is taken into account in the generation of the latent variable samples. Therefore, this modular division of the tree does not result in any loss of information and ensures that the likelihood of the whole tree is calculated at once. Subtree T_0—comprising of three cell divisions—is generally smaller than subtrees $T_1, T_2, \dots, T_8$ which can comprise of arbitrarily many cell divisions (in our data up to six divisions). Thus, there are considerably fewer possible configurations for T_0 than for $T_1, T_2, \dots, T_8$. To exploit this property for more efficient sampling, we propose a two-step sampling strategy: 1) first a set of latent samples for T_0 is generated and 2) based on each single sample, a second set of latent samples for $T_1, T_2, \dots, T_8$ is generated. In this way, the configuration of subtrees $T_1, T_2, \dots, T_8$ is sampled more often than T_0, increasing the chance of generating high-likelihood samples and thereby increasing the acceptance ratio of the sampling approach. This division of the tree is further encouraged by our data: in the T cell family trees, we observed an initial burst phase of semi-concordant cell divisions (corresponding to T_0), which is followed by more variable division speeds in the subsequent divisions (corresponding to $T_1, T_2, \dots, T_8$) (see the main text). This correspondence facilitates using different parameters for describing the first division phase (T_0) and the subsequent division phase ($T_1, T_2, \dots, T_8$), and lets us tune the adequate number of latent samples in each phase separately.

In the first step, we generate ns_1 samples of latent variables for all cells in subtree T_0. We denote the set of latent variables for T_0 by $\boldsymbol{x}_k^0$:

$$\boldsymbol{x}_k^0 = \left\{ x_k^{0,m} \middle| m = 1, \dots, n_{T_0} \right\} \quad \text{for} \quad k = 1, \dots, ns_1 \tag{1.13}$$

where n_{T_0} is the number of cells in subtree T_0. As explained in Section 3, we use a branching process framework (Harris, 1963) to randomly simulate latent variables $x_k^{0,m}$ according to the assumed model topology and its parameters. In this way, the latent variable of every cell is only dependent on the latent variable of its mother cell. Therefore, the configuration of subtree T_0, $\boldsymbol{x}_k^0$, is independent from the configuration of subtrees $T_1, T_2, \dots, T_8$. On the contrary, the latent states of $T_1, T_2, \dots, T_8$ depend on the their corresponding founder cells $c_1, c_2, \dots, c_8$.

In the second step, for each latent sample of $c_1, c_2, \dots, c_8$, we generate ns_2 latent samples for subtrees $T_1, T_2, \dots, T_8$. We denote the latent states of subtree T_j by $\boldsymbol{x}_{k,l}^j$:

$$\boldsymbol{x}_{k,l}^j = \left\{ x_{k,l}^{j,m} \middle| m = 1, \dots, n_{T_j} \right\} \quad \text{for} \quad k = 1, \dots, ns_1, \quad l = 1, \dots, ns_2, \quad j = 1, \dots, 8 \tag{1.14}$$

where n_{T_j} is the number of cells in subtree T_j. Each latent sample $\boldsymbol{x}_{k,l}^j$ for subtree T_j depends on the latent sample $\boldsymbol{x}_k^0$ of subtree T_0 through the state of the founder cells $c_1, c_2, \dots, c_8$. Since the lifetimes of individual cells are assumed to be independent random variables, the only dependence between subtrees is through topological dependence of latent variables. Thus, once latent variables are sampled, the data of different subtrees are independent from each other. For a given latent variable sample $\boldsymbol{x}^*$, the likelihood of the whole tree can be written as

$$\pi(\boldsymbol{y}|\boldsymbol{x}^*, \boldsymbol{\theta}) = \prod_{j=0}^{8} \pi\left(\boldsymbol{y}^j \middle| \boldsymbol{x}^*, \boldsymbol{\theta}\right) \tag{1.15}$$

For the simplicity of notation, we have dropped the index of tree in this section and use $\boldsymbol{y}$ to denote the data belonging to a given tree. $\boldsymbol{y}^j$ represents the data corresponding to subtree T_j. We then use the following Monte Carlo approximation to integrate out the two-layer latent variable samples and estimate the tree likelihood given the parameters:

$$\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}) = \frac{1}{ns_1}\sum_{k=1}^{ns_1}\hat{\pi}(\boldsymbol{y}|\boldsymbol{x}_k^0,\boldsymbol{\theta}) = \frac{1}{ns_1}\sum_{k=1}^{ns_1}\left(\pi(\boldsymbol{y}^0|\boldsymbol{x}_k^0,\boldsymbol{\theta})\ \prod_{j=1}^{8}\hat{\pi}(\boldsymbol{y}^j|\boldsymbol{x}_k^0,\boldsymbol{\theta})\right)$$
$$= \frac{1}{ns_1}\sum_{k=1}^{ns_1}\left(\pi(\boldsymbol{y}^0|\boldsymbol{x}_k^0,\boldsymbol{\theta})\ \prod_{j=1}^{8}\frac{1}{ns_2}\sum_{l=1}^{ns_2}\pi(\boldsymbol{y}^j|\boldsymbol{x}_{k,l}^j,\boldsymbol{\theta})\right) \tag{1.16}$$

As we need $\log(\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}))$ for the calculation of overall data log-likelihood in Eq. 1.12, we take the logarithm of the equation above. For better readability, we drop the dependence on $\boldsymbol{\theta}$ in the rest of the equations in this section, however it is implied that all the data likelihoods are conditioned on the parameter set $\boldsymbol{\theta}$:

$$\log(\hat{\pi}(\boldsymbol{y}))\ = \log\left(\frac{1}{ns_1}\right) + \ \log\left(\sum_{k=1}^{ns_1}\left(\pi(\boldsymbol{y}^0|\boldsymbol{x}_k^0)\ \prod_{j=1}^{8}\frac{1}{ns_2}\sum_{l=1}^{ns_2}\pi(\boldsymbol{y}^j|\boldsymbol{x}_{k,l}^j)\right)\right) \tag{1.17}$$

The logarithm cannot be pushed inside the sum in the second term of Eq. 1.17; therefore, we use the following numerical trick to robustly calculate this term based on the individual subtree log-likelihoods. More details about this robust numerical evaluation can be found in (Loos, 2016; Loos et al., 2016). We define q_k as the logarithm of the summand in the second term of Eq. 1.17 :

$$q_k = \ \log\left(\pi(\boldsymbol{y}^0|\boldsymbol{x}_k^0)\ \prod_{j=1}^{8}\frac{1}{ns_2}\sum_{l=1}^{ns_2}\pi(\boldsymbol{y}^j|\boldsymbol{x}_{k,l}^j)\right) \tag{1.18}$$

q^* denotes the maximum of q_k:

$$q^* = \ \arg\max_k q_k \tag{1.19}$$

The second term in Eq. 1.17 can then be evaluated via

$$\log\left(\sum_{k=1}^{ns_1}\left(\pi(\boldsymbol{y}^0|\boldsymbol{x}_k^0)\ \prod_{j=1}^{8}\frac{1}{ns_2}\sum_{l=1}^{ns_2}\pi(\boldsymbol{y}^j|\boldsymbol{x}_{k,l}^j)\right)\right) = \log\left(\sum_{k=1}^{ns_1}e^{(q_k - q^*)}\right) + \log(q^*) \tag{1.20}$$

Inserting 1.20 into 1.17, we obtain the following for the log-likelihood of a tree:

$$\log(\hat{\pi}(\boldsymbol{y}))\ = \log\left(\frac{1}{ns_1}\right) + \ \log\left(\sum_{k=1}^{ns_1}e^{(q_k - q^*)}\right) + \log(q^*) \tag{1.21}$$

According to Eq. 1.18, q_k is given as

$$q_k = \ \log\left(\pi(\boldsymbol{y}^0|\boldsymbol{x}_k^0)\right) + \ 8\ \log\left(\frac{1}{ns_2}\right) + \sum_{j=1}^{8}\log\left(\sum_{l=1}^{ns_2}\pi(\boldsymbol{y}^j|\boldsymbol{x}_{k,l}^j)\right) \tag{1.22}$$

To calculate the summands of the last term in Eq. 1.22, we use the same numerical trick as above. We define p_l^j as

$$p_l^j = \ \log\left(\pi(\boldsymbol{y}^j|\boldsymbol{x}_{k,l}^j)\right) \tag{1.23}$$

and denote the maximum of p_l^j as $p^{*,j}$:

$$p^{*,j} = \arg\max_l p_l^j \tag{1.24}$$

We then evaluate the logarithm of the summation in Eq. 1.22 as:

$$\log\left(\sum_{l=1}^{ns_2} \pi\left(\boldsymbol{y}^j \middle| \boldsymbol{x}_{k,l}^j\right)\right) = \log\left(\sum_{l=1}^{ns_2} e^{(p_l^j - p^{*,j})}\right) + \log\left(p^{*,j}\right) \tag{1.25}$$

Inserting Eq. 1.25 into Eq. 1.22, we obtain

$$q_k = \log\left(\pi\left(\boldsymbol{y}^0 \middle| \boldsymbol{x}_k^0\right)\right) + 8\log\left(\frac{1}{ns_2}\right) + \sum_{j=1}^{8}\left(\log\left(\sum_{l=1}^{ns_2} e^{(p_l^j - p^{*,j})}\right) + \log\left(p^{*,j}\right)\right) \tag{1.26}$$

To evaluate the log-likelihood for a given tree based on Equations 1.21, 1.23 and 1.26, we finally need to calculate $\log\left(\pi\left(\boldsymbol{y}^0 \middle| \boldsymbol{x}_k^0\right)\right)$ and $\log\left(\pi\left(\boldsymbol{y}^j \middle| \boldsymbol{x}_{k,l}^j\right)\right)$. For every latent sample $\boldsymbol{x}_k^0$, we calculate the log-likelihood of subtree T_0 by:

$$J_k^0 = \log\left(\pi\left(\boldsymbol{y}^0 \middle| \boldsymbol{x}_k^0\right)\right) = \sum_{m=1}^{n_{T_0}} \log\left(\pi\left(t_m^0 \middle| x_k^{0,m}\right)\right) = \sum_{m=1}^{n_{T_0}} J_k^{0,m} \tag{1.27}$$

Similarly, for each latent sample $\boldsymbol{x}_{k,l}^j$ dependent on $\boldsymbol{x}_k^0$, we calculate the log-likelihood of subtree T_j by:

$$J_{k,l}^j = \log\left(\pi\left(\boldsymbol{y}^j \middle| \boldsymbol{x}_{k,l}^j\right)\right) = \sum_{m=1}^{n_{T_j}} \log\left(\pi\left(t_m^j \middle| x_{k,l}^{j,m}\right)\right) = \sum_{m=1}^{n_{T_j}} J_{k,l}^{j,m} \tag{1.28}$$

The likelihood of individual cell lifetimes $\pi\left(t_m^0 \middle| x_k^{0,m}\right)$ and $\pi\left(t_m^j \middle| x_{k,l}^{j,m}\right)$ are determined based on the model hypothesis. The likelihoods for the models tested in this study are described in Section 3.

1.4. Output. This inference scheme returns samples from the posterior distribution of parameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ as output. It also provides the corresponding likelihood value for every sampled parameter set. The latter is used for the calculation of model evidence and model comparison.

2. Bayesian model comparison.

To compare the ability of different model hypotheses for explaining the lineage tree data, we perform Bayesian model comparison based on Bayes factors. We use the *Posterior Harmonic Mean estimator* (Vyshemirsky et al., 2007) to approximate the likelihood of the data given a model hypothesis M:

$$\pi(\boldsymbol{y}|M) \approx \left(\frac{1}{N}\sum_{i=1}^{N} \frac{1}{\pi\left(\boldsymbol{y}\middle|M, \boldsymbol{\theta}^{(i)}\right)}\right)^{-1}, \qquad \boldsymbol{\theta}^{(i)} \sim \pi(\boldsymbol{\theta}|\boldsymbol{y}, M) \tag{2.1}$$

where $\boldsymbol{\theta}^{(i)}$ are samples from the parameter posterior and $\pi\left(\boldsymbol{y}\middle|M, \boldsymbol{\theta}^{(i)}\right)$ are the corresponding likelihoods. We approximate the Bayes factor comparing model M_1 to model M_2 using the model evidences:

$$B_{12} = \frac{\pi(\boldsymbol{y}|M_1)}{\pi(\boldsymbol{y}|M_2)} \tag{2.2}$$

To quantify the uncertainty of these approximation, we drew 10,000 sets of samples from the parameter posterior distribution of every model. This yielded a distribution for the approximated model evidences and Bayes factors.

Algorithm 1: Bayesian inference scheme for tree-structured data

Data: A set of single cell lifetimes $\boldsymbol{y} = \{t_{im} | i = 1, \dots, n_{\text{tree}}, m = 1, \dots, n_{\text{cell}_i}\}$ from n_{tree} lineage trees each having n_{cell_i} cells; a model hypothesis; parameter prior $\pi(\boldsymbol{\theta})$; number of iterations N

Results: A set of samples $\{\boldsymbol{\theta}^{(n)}\}_{n=1}^{N}$ from the posterior distribution of parameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$

Initialization;

Sample a set of parameter values from the prior: $\boldsymbol{\theta}^{(0)} \sim \pi(\boldsymbol{\theta})$;

Calculate the log-likelihood of the data given the initial parameter set: $\hat{J}^{(0)} = \log\left(\hat{\pi}\left(\boldsymbol{y}|\boldsymbol{\theta}^{(0)}\right)\right)$

Initialize the chain of sampled parameter values and the corresponding log-likelihood values: $\boldsymbol{\theta}_{\text{chain}} \coloneqq [\boldsymbol{\theta}^{(0)}]$, $J_{\text{chain}} \coloneqq [\hat{J}^{(0)}]$;

Iterate over the parameter samples;

for $n = 1 \dots N$ **do**

  Generate a proposed set of parameter values $\boldsymbol{\theta}^*$ using the transition kernel (proposal distribution) $q\left(\boldsymbol{\theta}^{(n-1)}, \boldsymbol{\theta}^*\right)$: $\boldsymbol{\theta}^* \sim q\left(\boldsymbol{\theta}^{(n-1)}, \boldsymbol{\theta}^*\right) = f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(n-1)})$.

  for $i = 1 \dots n_{\text{tree}}$ **do**

    for $k = 1 \dots ns_1$ **do**

      Simulate a sample of latent variables of all cells in subtree T_0, according to the model topology and proposed parameters: $\boldsymbol{x}_k^0 = \{x_k^{0,m} | m = 1, \dots, n_{T_0}\}$

      Calculate the log-likelihood of subtree T_0 given $\boldsymbol{x}_k^0$ and the proposed parameter values $\boldsymbol{\theta}^*$:

      $$J_k^0 = \log\left(\pi\left(\boldsymbol{y}_i^0 | \boldsymbol{x}_k^0, \boldsymbol{\theta}^*\right)\right) = \sum_{m=1}^{n_{T_0}} \log\left(\pi\left(t_m^0 | x_k^{0,m}, \boldsymbol{\theta}^*\right)\right) = \sum_{m=1}^{n_{T_0}} J_k^{0,m}$$

      for $l = 1 \dots ns_2$ **do**

        for $j = 1 \dots 8$ **do**

          Simulate a sample of latent variables of all cells in subtree T_j, according to the model topology and proposed parameters: $\boldsymbol{x}_{k,l}^j = \left\{x_{k,l}^{j,m} \middle| m = 1, \dots, n_{T_j}\right\}$

          Calculate the log-likelihood of subtree T_j given $\boldsymbol{x}_{k,l}^j$ and the proposed parameter values $\boldsymbol{\theta}^*$:

          $$p_l^j = J_{k,l}^j = \log\left(\pi\left(\boldsymbol{y}_i^j | \boldsymbol{x}_{k,l}^j, \boldsymbol{\theta}^*\right)\right) = \sum_{m=1}^{n_{T_j}} \log\left(\pi\left(t_m^j | x_{k,l}^{j,m}, \boldsymbol{\theta}^*\right)\right) = \sum_{m=1}^{n_{T_j}} J_{k,l}^{j,m}$$

Calculate the log-likelihood of tree i given the latent variable given $\boldsymbol{x}_k^0$ and the proposed parameter values $\boldsymbol{\theta}^*$:

$$q_k = J_k^0 + 8\,\log\left(\frac{1}{ns_2}\right) + \sum_{j=1}^{8}\left(\log\left(\sum_{l=1}^{ns_2} e^{(p_l^j - p^{*,j})}\right) + \log\left(p^{*,j}\right)\right)$$

Calculate the log-likelihood of tree i given the proposed parameter values $\boldsymbol{\theta}^*$:

$$\hat{J}_i = \log\left(\hat{\pi}(\boldsymbol{y}_i|\boldsymbol{\theta}^*)\right) = \log\left(\frac{1}{ns_1}\right) + \log\left(\sum_{k=1}^{ns_1} e^{(q_k - q^*)}\right) + \log(q^*)$$

Calculate the overall data log-likelihood given the proposed parameter values $\boldsymbol{\theta}^*$:

$$\hat{J}^* = \hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}^*) = \exp\left(\sum_{i=1}^{n_{\text{tree}}} \hat{J}_i\right)$$

Evaluate the acceptance probability of the proposed parameter values $\boldsymbol{\theta}^*$**:**

$$\alpha\left(\boldsymbol{\theta}^{(n-1)}, \boldsymbol{\theta}^*\right) = \min\left\{1, \frac{\hat{\pi}(\boldsymbol{y}|\boldsymbol{\theta}^*)}{\hat{\pi}\left(\boldsymbol{y}|\boldsymbol{\theta}^{(n-1)}\right)}\right\} = \min\left\{1, \frac{\hat{J}^*}{\hat{J}^{(n-1)}}\right\}.$$

If uniform random variable $r \sim U(0,1) < \alpha\left(\boldsymbol{\theta}^{(n-1)}, \boldsymbol{\theta}^*\right)$ **then**

Accept the proposed parameter values and add to the chain: $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}^*, \hat{J}^{(n)} = \hat{J}^*$.

else

Reject the proposed parameter values and keep the previous parameter values: $\boldsymbol{\theta}^{(n)} = \boldsymbol{\theta}^{(n-1)}, \hat{J}^{(n)} = \hat{J}^{(n-1)}$.

Construct a sample of parameters from the posterior distribution and the corresponding log-likelihoods: $\left\{\left(\boldsymbol{\theta}^{(n)}|\boldsymbol{y}\right); \hat{J}^{(n)}\right\}_{n=1}^{N}$.

---

3. Model specifications and corresponding likelihoods.

For analyzing the T cell lineage trees, we considered several model hypotheses about the diversification of single-cell-derived T cell populations into distinct subsets. We characterized different subsets with their distinct division speeds and modeled them as latent variables in our approach. We used a branching process framework (Harris, 1963) to model this diversification: we assumed that upon every cell division, each of the two daughter cells can (independently) either adopt the subset of the mother cell, or change into a different subset according to the model topology. Each of these choices can occur with probabilities that are part of the model parameters. In the following, we first define these model hypotheses, and then describe the simulation of latent variables and likelihood calculation for them.

3.1. Model hypotheses tested in this study. A schematic of the following models is shown in Figure 2C and 2J of the main text:

1. *Model #1*. We assume that no diversification occurs during the expansion of the T cell families and therefore the whole cell population follows the same inter-division time statistics. This "homogeneous" subset is denoted by H. The only model parameters are the mean and the CV

of the inter-division time distribution: $\boldsymbol{\theta} = \{\mu_H, \mathrm{CV}_H\}$. Since all cells belong to one subset, there are no latent variables in this model.

2. *Model #2.* This model assumes that after an initial phase of semi-concordant divisions, the cells differentiate into another subset with different inter-division time statistics. The subset in the early expansion phase is referred to as Early-Activated (EA) and the subsequent subset is denoted by H. Upon every division of EA cells, each of the daughter cells can either stay in the EA subset with a certain probability or differentiate into H otherwise. The model allows differentiation probabilities to vary between the phase 1 (subtree T_0 in Fig. S28) and phase 2 (subtrees $T_1, T_2, \dots, T_8$ in Fig. S28) of expansion; see Model #3 for the justification of this variation. The model parameters are the mean and CV of the inter-division time distribution of the subsets EA and H ($\mu_{EA}, \mathrm{CV}_{EA}, \mu_H, \mathrm{CV}_H$), and the probability of remaining in subset EA in the two expansion phases ($p_{\mathrm{EA},1}, p_{\mathrm{EA},2}$): $\boldsymbol{\theta} = \{p_{\mathrm{EA},1}, p_{\mathrm{EA},2}, \mu_{EA}, \mathrm{CV}_{EA}, \mu_H, \mathrm{CV}_H\}$. The latent variable (i.e. subset) for a cell m in model #2 can have two values: $x_m \in \{EA, H\}$.

3. *Model #3.* Similar to model #2, this model assumes an initial phase of semi-concordant divisions (subset EA). However, model #3 assumes a bifurcating pathway of diversification where the cells can differentiate out of EA subset into two distinct subsets of fast and slowly dividing cells (subsets F and S respectively). This differentiation can happen at every EA cell division with specific probabilities for staying in the EA subset or differentiating into F or S. In our T cell trees, the first few divisions are considerably less variable than the subsequent divisions (see the main text). To mimic this aspect, model #3 allows the differentiation probabilities to vary between the two expansion phases: the probability of differentiating from EA into S or F in phase 2 (subtrees $T_1, T_2, \dots, T_8$) is constrained to be higher than 90%, while it is unconstrained in phase 1 (subtree T_0). The ratio between the differentiation probabilities into S and F subsets ($\mathrm{ratio}_{\mathrm{Slow}} = \frac{\text{differentiation probabilities into } S}{\text{differentiation probabilities into } F}$) is set to be the same in the two phases. To directly capture the distinction between the F and S subsets, the difference between their mean inter-division times is modeled as a parameter. The model parameters are the mean and CV of the inter-division time distribution of the *EA* and *F* subsets ($\mu_{EA}, \mathrm{CV}_{EA}$, μ_F, CV_F), the difference in the mean inter-division time of subset S w.r.t. subset F, and the CV of its inter-division time distribution ($d_{\mu_S} = \mu_S - \mu_F, \mathrm{CV}_S$), the probability of cells remaining in the EA subset in the two expansion phases ($p_{\mathrm{EA},1}, p_{\mathrm{EA},2}$) and the ratio between the differentiation probabilities into S and F subsets ($\mathrm{ratio}_{\mathrm{Slow}}$): $\boldsymbol{\theta} = \{p_{\mathrm{EA},1}, p_{\mathrm{EA},2}, \mathrm{ratio}_{\mathrm{Slow}}, \mu_{EA}, \mathrm{CV}_{EA}, \mu_F, \mathrm{CV}_F, d_{\mu_S}, \mathrm{CV}_S\}$. The latent variable (i.e. subset) for a cell m in this model can have three values: $x_m \in \{EA, S, F\}$.

4. Mixture model. This model assumes three subsets as in model #3. Initially cells belong to the EA subset; upon every division, each daughter cell can stay in the EA subset with a certain probability or change into a Mixture type (M) otherwise. The latter represents a mixture of two distinct subsets of fast and slowly dividing cells (subsets F and S respectively). Like model #3, different differentiation probabilities for the two expansion phases are assumed. The subset of M cells is not certainly known; instead, it is assumed that they belong to subset S with probability $\mathrm{ratio}_{\mathrm{Slow}}$ and to subset F with probability $1 - \mathrm{ratio}_{\mathrm{Slow}}$. The model parameters are the mean and CV of the inter-division time distribution of the *EA* and *F* subsets ($\mu_{EA}, \mathrm{CV}_{EA}$, μ_F, CV_F), the difference in the mean inter-division time of subset S w.r.t. subset F, and the CV

of its inter-division time distribution (d_{μ_S}, CV_S), the probability of cells remaining in the EA subset in the two expansion phases ($p_{\mathrm{EA},1}, p_{\mathrm{EA},2}$) and the probability that the cells in the mixture type possess subset S ($\mathrm{ratio}_{\mathrm{Slow}}$): $\boldsymbol{\theta} = \left\{p_{\mathrm{EA},1}, p_{\mathrm{EA},2}, \mathrm{ratio}_{\mathrm{Slow}}, \mu_{EA}, \mathrm{CV}_{EA}, \mu_F, \mathrm{CV}_F, d_{\mu_S}, \mathrm{CV}_S\right\}$. The latent variable (i.e. type) for a cell m can have two values: $x_m \in \{EA, M\}$.

We emphasize that this "mixture model" does not intend to model an actual topology for the diversification of T cells. Rather, it provides a mathematical model alternative in which the same number of subsets and parameters as in model #3 is assumed. We compare this model against model #3 to test whether the superior performance of model #3 over models #1 and #2 is merely due to its additional flexibility, or whether the topological pathway encoded in model #3 is indeed essential.

5. *Model #4.* This model assumes the same diversification pathway as in model #3 and adds additional variability between different family trees (interfamily variation). It assumes that the mean inter-division time of subset F is distributed across different trees according to a log-normal distribution with unknown mean and coefficient of variation ($\tilde{\mu}_F, \widetilde{\mathrm{CV}}_F$):

$$\mu_F \sim \log N\left(\tilde{\mu}_F, \widetilde{\mathrm{CV}}_F\right) \tag{3.1}$$

It further assumes that the difference between the mean inter-division time of different subsets is constant in all trees. This implies that if subset F in a tree is slower than the average, then subsets EA and S are also slower with the same distance to the average. In this way, all subset-specific mean inter-division times—for subsets EA, F and S—are log-normally distributed across different trees. The model parameters are the mean and CV for the log-normal distribution of μ_F ($\tilde{\mu}_F, \widetilde{\mathrm{CV}}_F$), the differences in the mean inter-division time between the subsets ($d_{\mu_{EA}} = \mu_{EA} - \mu_F, d_{\mu_S} = \mu_S - \mu_F$), the subset-specific CVs of the inter-division time distribution ($\mathrm{CV}_{EA}, \mathrm{CV}_F, \mathrm{CV}_S$), and the differentiation probabilities as in model #3: $\boldsymbol{\theta} = \left\{p_{\mathrm{EA},1}, p_{\mathrm{EA},2}, \mathrm{ratio}_{\mathrm{Slow}}, \tilde{\mu}_F, \widetilde{\mathrm{CV}}_F, \mathrm{CV}_F, d_{\mu_{EA}}, \mathrm{CV}_{EA}, d_{\mu_S}, \mathrm{CV}_S\right\}$. The latent variable (i.e. subset) for a cell m in model #4 can have three values: $x_m \in \{EA, S, F\}$.

6. *Model #1 with mother-daughter correlation.* This model assumes that no diversification occurs during the expansion of the T cell families and therefore the whole cell population follows the same inter-division time statistics described by a single log-normal distribution. This "homogeneous" subset is denoted by H. However, it further assumes that the interdivision of mother and daughter cells are correlated. Specifically, if we denote the log-transformed inter-division time of the mother cell as t_m and the log-transformed inter-division time of the daughter cell as t_d, then the pair $\begin{bmatrix} t_m \\ t_d \end{bmatrix}$ is distributed according to the bivariate normal distribution $N\left(\begin{bmatrix} \mu_H \\ \mu_H \end{bmatrix}, \begin{bmatrix} \sigma_H^2 & \rho_{m,d}\,\sigma_H^2 \\ \rho_{m,d}\,\sigma_H^2 & \sigma_H^2 \end{bmatrix}\right)$. μ_H and σ_H^2 are the mean and variance of the inter-division time distribution of all cells, and $\rho_{m,d}$ is the correlation coefficient between the inter-division time of mother and daughter cells. The model parameters are: $\boldsymbol{\theta} = \left\{\mu_H, \mathrm{CV}_H, \rho_{m,d}\right\}$ with CV_H being the coefficient of variation of the inter-division time distribution. Since all cells belong to one subset, there are no latent variables in this model.

7. *Model #3* + $\Delta\mathrm{CD62L}^{+/-}$. This model is the same as model #3, except that the difference in the mean inter-division time of subset S w.r.t. subset F ($d_{\mu_S} = \mu_S - \mu_F$) is <u>not</u> a model parameter. Instead, d_{μ_S} is set to the difference between the mean inter-division time of $\mathrm{CD62L}^{+}$and $\mathrm{CD62L}^{-}$ cells in the experimental data. The model parameters are, therefore, the parameters of model #3 excluding d_{μ_S}: $\boldsymbol{\theta} = \{p_{\mathrm{EA},1}, p_{\mathrm{EA},2}, \mathrm{ratio}_{\mathrm{Slow}}, \mu_{EA}, \mathrm{CV}_{EA}, \mu_F, \mathrm{CV}_F, \mathrm{CV}_S\}$. The latent variable (i.e. subset) for a cell m in this model can have three values: $x_m \in \{EA, S, F\}$.

3.2. Simulation of the latent variables. To calculate the likelihood of the data we need to simulate samples from the latent variables (Section 1.3.3). In model #1 all cells belong to the same subset and therefore no latent variables exist. For the rest of the models, we use the model topology and parameters to simulate the subsets of cells along a family tree. We assign the first founder cell of the tree to the Early-Activated subset as this subset is the starting point of all the considered pathways. We move along the family tree and upon every cell division randomly assign the subset of the daughter cells. Given the simulated subset of the mother cell, the possible subsets for each daughter cell are determined based on the model topology. The probability of a daughter cell to adopt any of these possible subsets is given by the model parameters (i.e. the differentiation probabilities). In this way, the tree structure and mother-daughter relationships are taken into account in simulating the latent variables.
For example in model #3, if a mother cell belongs to subset EA in the first expansion phase, each daughter cell is independently assigned to subset EA with probability $p_{EA,1}$, subset S with probability $\mathrm{ratio}_{\mathrm{Slow}}(1 - p_{EA,1})$ and subset F with probability $(1 - \mathrm{ratio}_{\mathrm{Slow}})(1 - p_{EA,1})$. However, if the mother cell belongs to either of the S and F subsets, both daughter cells directly inherit the subset of the mother cell.

3.3. Calculation of the likelihood for models #1-4. We need to calculate the likelihood of every cell given the latent variables and model parameters to obtain the overall data likelihood (see Equations 1.27 and 1.28 in Section 1.3.3). We assume that the inter-division of cells are log-normally distributed with subset-specific mean and coefficient of variation. If we denote the latent variable corresponding to cell m by x_m, describing the subset to which cell m belongs, the log-transformed inter-division time of cell m is distributed according to the normal distribution $N(\mu_{x_m}, \mathrm{CV}_{x_m})$; μ_{x_m} and CV_{x_m} are determined by the model parameters. The likelihood of the lifetime of cell m, t_m, is then

$$\begin{cases} \pi(t_m | \mu_{x_m}, \mathrm{CV}_{x_m}) = \dfrac{1}{\mu_{x_m}\ \mathrm{CV}_{x_m}\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{t_m - \mu_{x_m}}{\mu_{x_m}\ \mathrm{CV}_{x_m}}\right)^2}, & \text{if cell } m \text{ divided in the course of the experiment} \\ \pi(t_m | \mu_{x_m}, \mathrm{CV}_{x_m}) = 1 - \displaystyle\int_0^{t_m} \frac{1}{\mu_{x_m}\ \mathrm{CV}_{x_m}\sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{s - \mu_{x_m}}{\mu_{x_m}\ \mathrm{CV}_{x_m}}\right)^2} ds, & \text{if cell } m \text{ did not divide in the course of the experiment} \end{cases}$$

(3.2)

According to Eq. 3.2, if a cell has divided, its lifetime represents its inter-division time and therefore the PDF of the normal distribution $N(\mu_{x_m}, \mathrm{CV}_{x_m})$ is used to calculate its likelihood. But if a cell has not divided in the course of the experiment, its lifetime only provides a lower-bound for its inter-division time and therefore, the CDF of the corresponding normal distribution is used.

3.4. Calculation of the likelihood for model #1 with mother-daughter correlation. In this model, the inter-division times of individual cells are not independent, but every cell's division time depends on the division time of its mother. If we denote the log-transformed inter-division time of a mother cell and its daughter cell respectively by t_m and t_d, then the likelihood of the pair $\begin{bmatrix} t_m \\ t_d \end{bmatrix}$ is described by a bivariate normal distribution $N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho_{m,d}\,\sigma^2 \\ \rho_{m,d}\,\sigma^2 & \sigma^2 \end{bmatrix}\right)$. The overall data likelihood is calculated as the product of the conditional likelihoods $\pi(t_d|t_m)$ over all cells in the tree. For every cell, this conditional probability is calculated according to the above-mentioned bivariate distribution.

3.5. Calculation of the likelihood for the mixture model. In the mixture model, the subset of cells in the mixture type (M) is not directly known; instead, it is assumed that they could belong to subset S with probability $\text{ratio}_{\text{Slow}}$ and to subset F with probability $1 - \text{ratio}_{\text{Slow}}$. Therefore Eq. 3.2, that assumes the subset of cells is known given the latent variables, cannot be used for likelihood calculation. Instead, we use the following equation for cells of type M based on the likelihood for a mixture model (Pyne et al., 2009):

$$\begin{cases} \pi(t_m) = \text{ratio}_{\text{Slow}} \dfrac{1}{\mu_S\ \text{CV}_S\ \sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{t_m-\mu_S}{\mu_S\ \text{CV}_S}\right)^2} \quad + \\ (1-\text{ratio}_{\text{Slow}})\ \dfrac{1}{\mu_F\ \text{CV}_F\ \sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{t_m-\mu_F}{\mu_F\ \text{CV}_F}\right)^2}, & \text{if cell } m \text{ divided in the course of the experiment} \\ \pi(t_m) = \text{ratio}_{\text{Slow}} \left(1 - \int_0^{t_m} \dfrac{1}{\mu_S\ \text{CV}_S\ \sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{s-\mu_S}{\mu_S\ \text{CV}_S}\right)^2} ds\right) + \\ (1-\text{ratio}_{\text{Slow}}) \left(1 - \int_0^{t_m} \dfrac{1}{\mu_F\ \text{CV}_F\ \sqrt{2\pi}}\ e^{-\frac{1}{2}\left(\frac{s-\mu_F}{\mu_F\ \text{CV}_F}\right)^2} ds\right), & \text{if cell } m \text{ did not divide in the course of the experiment} \end{cases}$$

(3.3)

Equation 3.3 describes the likelihood of the lifetime of a cell as the weighted sum of the likelihoods if it belonged to subsets S and F respectively. For cells in the EA subset, Eq. 3.2 is used to calculate their likelihood.

3.6. Inference results for different models. We split our data into eight groups of five trees and performed parameter inference and model comparison on all eight data groups. The inferred parameter posteriors, model evidences and Bayes factors for every data group are shown in Figures S4, S6, S7, S12, S13 and S17-20. For better comparability, we reported base-10 logarithm of model evidences and Bayes factors. In addition to the group-wise metrics, we calculated average model evidences and Bayes factors from the results of all data groups together. The average log10-model evidence was calculated as $\frac{1}{8}\sum_{i=1}^{8} ME_i$, where ME_i is the median of the log10-model evidence calculated in data group i. The average log10-Bayes factor was calculated as $\frac{1}{8}\sum_{i=1}^{8} BF_i$, where BF_i is the median of the log10-Bayes factor calculated in data group i. The average model evidences and Bayes factors are shown in Figures 2D, 2L, S3 and S5.

4. Statistical features of the experimental data.

4.1. Contribution of intrafamily and interfamily variability in the overall variation of inter-division times. We calculated the total variance of the inter-division times and the contribution of intrafamily and interfamily sources in the experimental data. Intrafamily variance was calculated as the weighted mean of the variances of inter-division times within different families:

$$\text{Intrafamily variance} = \frac{\sum_{i=1}^{n_{\text{tree}}} n_{\text{cell}_i}\, \sigma_i^2}{\sum_{i=1}^{n_{\text{tree}}} n_{\text{cell}_i}} \tag{4.1}$$

where n_{cell_i} and σ_i^2 are respectively the number of cells and the variance of inter-division times in the i^{th} tree. Interfamily variance was calculated as the weighted variance of the family mean inter-division times:

$$\text{Interfamily variance} = \frac{\sum_{i=1}^{n_{\text{tree}}} n_{\text{cell}_i}\, (\mu_i - \mu^*)^2}{\sum_{i=1}^{n_{\text{tree}}} n_{\text{cell}_i} - 1}\ , \quad \mu^* = \frac{\sum_{i=1}^{n_{\text{tree}}} n_{\text{cell}_i}\, \mu_i}{\sum_{i=1}^{n_{\text{tree}}} n_{\text{cell}_i}} \tag{4.2}$$

where μ_i is the mean inter-division time in the i^{th} tree, and μ^* is the weighted mean of μ_i for $i = 1, \dots, n_{\text{tree}}$. These variance terms for the experimental data are shown in Figure 1G.

4.2. Percentage of the trees whose four branches have significantly distinct inter-division times. For every tree in the experimental data, we compared the inter-division time distribution of the four branches of the tree as depicted in Figure 1I. We performed one-way ANOVA to test whether the four branches have different mean inter-division times. We then estimated positive false discovery rates (pFDR) for multiple hypothesis testing based on the Benjamini and Hochberg method (Benjamini and Hochberg, 1995) across all trees. We selected the trees that had a pFDR smaller than 0.05 as trees whose four branches have significantly distinct inter-division times. These selected trees are highlighted in Figure 1J.

5. Simulation studies for model validation.

To inspect how well the studied models were matching to the experimental data—beyond measures provided by model evidence and Bayes factors—we performed the following simulation studies. We examined how well the simulated data based on different models captured the statistical features of the experimental data.

5.1. Overall distribution of the inter-division times. We simulated 10,000 family trees based on the model topologies #3 and #4, and the inferred posterior distribution for the corresponding model parameters. We obtained the distribution of the inter-division times of all cells regardless of their subset from all simulated trees together; we compared this against the distribution of inter-division times in the experimental data. We further obtained the distribution of inter-division times of cells in the subsets EA, S and F in the simulated data for illustration purposes. The results are shown in Figures S8 and S14 for model #3 and model #4 respectively.

5.2. Contribution of intrafamily and interfamily variability in the overall variation of inter-division times. We simulated 500 datasets of each 44 trees based on model topologies #3 and #4, and the inferred posterior distribution for the corresponding model parameters. For each dataset, we calculated the total variance of the inter-division times and the contribution of intrafamily and interfamily sources. Intrafamily variance was calculated as the weighted mean of the variances of inter-division times within different families:

$$\text{Intrafamily variance} = \frac{\sum_{i=1}^{44} \hat{n}_i \, \hat{\sigma}_i^2}{\sum_{i=1}^{44} \hat{n}_i} \tag{5.1}$$

where $\hat{n}_i$ and $\hat{\sigma}_i^2$ are respectively the number of cells and the variance of inter-division times in the i^{th} simulated tree. Interfamily variance was calculated as the weighted variance of the family mean inter-division times:

$$\text{Interfamily variance} = \frac{\sum_{i=1}^{44} \hat{n}_i \, (\hat{\mu}_i - \bar{\mu})^2}{\sum_{i=1}^{44} \hat{n}_i - 1} \;, \quad \bar{\mu} = \frac{\sum_{i=1}^{44} \hat{n}_i \, \hat{\mu}_i}{\sum_{i=1}^{44} \hat{n}_i} \tag{5.2}$$

where $\hat{\mu}_i$ is the mean inter-division time in the i^{th} simulated tree, and $\bar{\mu}$ is the weighted mean of $\hat{\mu}_i$ for $i = 1, \ldots, 44$. We then obtained the 90%-confidence interval for the total variance, intrafamily variance and interfamily variance in the 500 simulated datasets. We examined whether these confidence intervals contained the corresponding variance terms calculated for the experimental data. The results for the simulations based on model #3 and model #4 are shown in Figures S9 and S15 respectively.

5.3. Percentage of the trees whose four branches have significantly distinct inter-division times. We used the same simulated data as in Section 5.2. For every tree in a simulated dataset, we compared the inter-division time distribution of the four branches of the tree as depicted in Figure 1I. We performed one-way ANOVA to test whether the four branches have different mean inter-division times. We then estimated positive false discovery rates (pFDR) for multiple hypothesis testing based on the Benjamini and Hochberg method (Benjamini and Hochberg, 1995) across all 44 trees in the dataset. We selected the trees that had a pFDR smaller than 0.05 as trees whose four branches have significantly distinct inter-division times, and in every dataset calculated the percentage of these selected trees. We then obtained the distribution of this percentage across the 500 simulated datasets and examined whether this distribution contained the percentage calculated from the experimental data. The results for the simulations based on models #3 and #4 are shown in Figures S10 and S16 respectively.

5.4. Simulation of T Cell family responses to mimic an *in vivo* response. We simulated 500 single-cell-derived T cell families according to model #4 up to day 7 post infection to mimic the peak of *in vivo* T cell responses. The model parameters were set to the median of the parameter posteriors obtained from fitting model #4 to the continuous stimulation data (Fig. S13, group 7). Since the activation and recruitment of cells into the *in vivo* T cell response does not occur as concerted as upon *in vitro* stimulation, we modeled the distribution of first cell division times as in Gerlach et al. (2013, Fig. S7). We then compared the statistics of the simulated responses at the peak of expansion to those of the *in vivo* data in Buchholz et al. (2013). Of note: Buchholz et al. (2013) constitutes a very similar experimental model as Gerlach et al. (2013). We assumed that the "Slow" cells in the simulated data correspond to the CMp cells in the *in vivo* data.

SI References:

1. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Royal Stat. Soc.* 57:289–300.

2. Harris T (1963) *The Theory of Branching Processes*. Springer, Berlin.

3. Hilsenbeck O, et al. (2016) Software tools for single-cell tracking and quantification of cellular and molecular properties. *Nat Biotechnol* 34(7):703–706.

4. Loos C (2016) *Analysis of single-cell data: ODE-constrained mixture modeling and approximate Bayesian computation.* Best Masters. Springer.

5. Loos C, et al. (2016) Parameter Estimation for Reaction Rate Equation Constrained Mixture Models. In: Bartocci E., Lio P., Paoletti N. (eds) *Computational Methods in Systems Biology*. CMSB 2016. Lecture Notes in Computer Science, vol 9859. Springer, Cham.

6. Pyne S, et al. (2009) Automated high-dimensional flow cytometric data analysis. *Proc. Natl. Acad. Sci.* USA, 106, 8519–8124.

7. Vyshemirsky V, Girolami MA (2008) Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6):833–839.

8. Wilkinson DJ (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat. Rev. Genet.* 10:122-133.

9. Wilkinson DJ (2011) *Stochastic Modelling for Systems Biology*. Second Edition. CRC Press.

10. Gerlach C, et al. (2013) Heterogeneous differentiation patterns of individual CD8+ T cells. *Science* 340(6132):635–639.

11. Buchholz VR, et al. (2013) Disparate individual fates compose robust CD8+ T cell immunity. *Science* 340(6132):630–635.

SI Figure Legends:

Figure S1:
IL-2 levels are stable throughout the experiment. Cells were activated for 24 h. Subsequently, 0, 1, 10 or 100 cells were transferred into wells with 25 U/mL IL-2 and cultured for further three days. IL-2 in the supernatant was detected by ELISA. As a control, culture medium without cells was measured, containing no or 25 U/mL freshly added IL-2. One-way ANOVA and Dunnett's multiple comparison test: **** $p<0.0001$.

Figure S2:
Comparison of division speeds in the two branches emerging after the first cell division. (A) Analogue to Figure 1I: A representative family tree was divided into two branches (red and green) after the first cell division. **(B)** Inter-division time of cells in the two branches starting from the first generation are color-coded for all trees in Figure 1F. In 15 out of 43 trees (~35%, tress highlighted in gray), inter-division times differed significantly between the two branches. For every tree, a p-value was calculated based on one-way ANOVA. We then estimated positive false discovery rates (pFDR) for multiple hypothesis testing based on these p-values, and used a cutoff of $pFDR<0.05$ for significance.

Figure S3:

Model comparison summary from the results of all data groups. The model evidences for models #1, #2, #3, #4 (Fig. 2 C and J), the mixture model and model #1 with mother-daughter correlation are shown. These evidences indicate the following hierarchy where model #4 explains the data best and model #1 is least matching to the data: model #4 > model #3 > mixture model > model #1 with mother-daughter correlation > model #2 > model #1. The circles show the mean of model evidences of eight data groups (Supplementary Methods). The error bars show the standard error of the mean.

Figure S4:

Model comparison results in individual groups. The model evidences for models #1, #2, #3, #4 (respectively M 1, M2, M 3, M 4), the mixture model (Mix. M), and the model including mother-daughter correlation (M 1 + corr.) for the eight data groups. The boxplots show the distribution of the model evidences calculated based on 10000 bootstrapped samples of parameter posteriors (Supplementary Methods).

Figure S5:

Bayes factors summary from the results of all data groups. (A) The log10-Bayes factor of model #4 (Fig. 2J) compared to models #1, #2, #3 (Fig. 2C), the mixture model and model #1 with mother-daughter correlation is shown. The circles show the mean of log10-Bayes factors calculated in the eight data groups. The error bars show the standard error of the mean. The dashed grey line shows the cutoff value for "strong evidence" (log10-Bayes factor = 1) and the dashed black line shows the cutoff value for "decisive evidence" (log10-Bayes factor = 2) for model #4. **(B)** Same as (A) where the log10-Bayes factors are calculated for model #3 compared to models #2 and #1, the mixture model and the model #1 with mother-daughter correlation.

Figure S6:

Bayes factors in individual groups with respect to model #3. The log10-Bayes factor of model #3 (Fig. 2C) compared to models #1, #2 (M 1 and M 2 respectively), the mixture model (Mix. M) and the model including mother-daughter correlation (M 1 + corr.) for the eight data groups. The boxplots show the distribution of the log10-Bayes factors calculated based on 10000 bootstrapped samples of parameter posteriors (Supplementary Methods). The dashed grey line shows the cutoff value for "strong evidence" (log10-Bayes factor = 1) and the dashed black line shows the cutoff value for "decisive evidence" (log10-Bayes factor = 2) for model #3.

Figure S7:

Inferred parameter values for model #3 in the eight data groups. The circles and the error bars respectively show the median and the 95% credible intervals of the parameter posterior distributions.

Figure S8:

The distribution of inter-division times in simulated data based on model #3. The distribution of inter-division times in 10000 simulated trees based on model #3 compared to that of the experimental data (grey) for the eight data groups. In every group, the parameter posteriors inferred in that group are used for simulating the trees. The red histogram shows the overall distribution in the simulated data, while the black, blue and green histograms show the distribution of "Early-activated", "Fast-dividing" and "Slow-dividing" subsets respectively.

Figure S9:

Variance of inter-division times in the simulated data based on model #3. Total variance of the inter-division times and the contribution of intrafamily and interfamily sources as observed in the experimental data (grey bars) and the simulated data (red boxes) for the eight data groups. The simulated data consists of 500 datasets of each 44 trees simulated based on model #3 and the parameter posteriors of every data group. Intrafamily variance is calculated as the weighted mean of the variances of inter-division times within different families. Interfamily variance is calculated as the weighted variance of the family mean inter-division times (Supplementary Methods).

Figure S10:

Percentage of trees with significantly distinct branches in the simulated data based on model #3. Percentage of the trees whose four branches (as in Fig. 1I) have significantly distinct inter-division times is shown. The grey line shows the experimental data and the red histogram shows the distribution of this percentage in the simulated data for the eight data groups. The simulated data consists of 500 datasets of each 44 trees simulated based on model #3 and the parameter posteriors of every data group. (Supplementary Methods).

Figure S11:

Variation of subset-specific mean inter-division times between different families. Inferred distribution of subset-specific mean inter-division times based on model #4 for the eight data groups. The black, blue and green curves show the distribution for the "Early-activated", "Fast-dividing" and "Slow-dividing" subsets respectively.

Figure S12:

Bayes factors in individual groups with respect to model #4. The log10-Bayes factor of model #4 (Fig. 2J) compared to models #1, #2, #3 (respectively M 1, M2, M 3), the mixture model (Mix. M) and the model including mother-daughter correlation (M 1 + corr.) for the eight data groups. The boxplots show the distribution of the log10-Bayes factors calculated based on 10000 bootstrapped samples of parameter posteriors (Supplementary Methods). The dashed grey line shows the cutoff value for "strong evidence" (log10-Bayes factor = 1) and the dashed black line shows the cutoff value for "decisive evidence" (log10-Bayes factor = 2) for model #4.

Figure S13:

Inferred parameter values for model #4 in the eight data groups. The circles and the error bars respectively show the median and the 95% credible intervals of the parameter posterior distributions.

Figure S14:

The distribution of inter-division times in simulated data based on model #4. The distribution of inter-division times in 10000 simulated trees based on model #4 compared to that of the experimental data (grey) for the eight data groups. In every group, the parameter posteriors inferred in that group are used for simulating the trees. The red histogram shows the overall distribution in the simulated data, while the black, blue and green histograms show the distribution of "Early-activated", "Fast-dividing" and "Slow-dividing" subsets respectively.

Figure S15:

Variance of inter-division times in the simulated data based on model #4. Total variance of the inter-division times and the contribution of intrafamily and interfamily sources as observed in the experimental data (grey bars) and the simulated data (red boxes) for the eight data groups. The simulated data consists of 500 datasets of each 44 trees simulated based on model #4 and the parameter posteriors of every data group. Intrafamily variance is calculated as the weighted mean of the variances of inter-division times within different families. Interfamily variance is calculated as the weighted variance of the family mean inter-division times (Supplementary Methods).

Figure S16:

Percentage of trees with significantly distinct branches in the simulated data based on model #4. Percentage of the trees whose four branches (as in Fig. 1I) have significantly distinct inter-division times is shown. The grey line shows the experimental data and the red histogram shows the distribution of this percentage in the simulated data for the eight data groups. The simulated

data consists of 500 datasets of each 44 trees simulated based on model #4 and the parameter posteriors of every data group. (Supplementary Methods).

Figure S17:

Inferred parameter values for model #1 in the eight data groups. The circles and the error bars respectively show the median and the 95% credible intervals of the parameter posterior distributions.

Figure S18:

Inferred parameter values for model #2 in the eight data groups. The circles and the error bars respectively show the median and the 95% credible intervals of the parameter posterior distributions.

Figure S19:

Inferred parameter values for the mixture model in the eight data groups. The circles and the error bars respectively show the median and the 95% credible intervals of the parameter posterior distributions.

Figure S20:

Inferred parameter values for model #1 with mother-daughter correlation in the eight data groups. The circles and the error bars respectively show the median and the 95% credible intervals of the parameter posterior distributions.

Figure S21:

Simulated *in vivo* response based on the *in vitro* model. Model #4 is used to generate simulated single-cell-derived responses at day 7 post infection. We used the median of the parameter posteriors obtained from fitting model #4 to the continuous stimulation data to simulate 500 family trees (Supplementary Methods). The statistics of these simulated data are compared to those of the *in vivo* data from (Buchholz et al. 2013). **(A)** The family sizes for the simulated data. The mean, median and CV of the family sizes are shown for the model (black) and the *in vivo* data (orange). **(B)** The correlation of the family sizes and the percentage of "Slow" cells for the simulated data. The Spearman's rank correlation coefficient is shown for the model (black) and the *in vivo* data (orange). The $CD62L^+$ cells from the *in vivo* data are considered equivalent to the simulated "Slow" cells.

Figure S22:

Model comparison results for the "brief" and "brief + IL12" groups. The model evidences for models #1, #2, #3, and #4 are shown. The boxplots show the distribution of the model evidences

calculated based on 10000 bootstrapped samples of parameter posteriors (Supplementary Methods).

Figure S23:

Bayes factors in the "brief" and "brief + IL12" groups. The log10-Bayes factor of model #4 compared to models #1, #2 and #3 is shown. The boxplots show the distribution of the log10-Bayes factors calculated based on 10000 bootstrapped samples of parameter posteriors (Supplementary Methods). The dashed grey lines show the cutoff value for "strong evidence" (log10-Bayes factor = 1) and the dashed black lines show the cutoff value for "decisive evidence" (log10-Bayes factor = 2) for model #4.

Figure S24:

Impact of antibodies on proliferation *in vitro*. 100 naïve T cells were sorted per well (anti-CD3/CD28 coated). Anti-CD62L or anti-CD25 antibodies were added to the culture medium 1:100, 1:1000 or 1:10 000 diluted. After 5 days the cell number per well was measured using flow cytometry. Of note: For live cell imaging dilutions of 1:10 000 and 1:20 000 were used for anti-CD62L and anti-CD25, respectively.

Figure S25:

CD25 is enriched in distinct branches and correlates with division speed. Cells were stimulated with anti-CD3/CD28 and imaged in the presence of anti-CD25-APC. **(A)** A representative tree and its heattree plot as in Fig. 3F. Two very distinct (in terms of their CD25 expression) branches (A and B) and the mother cell from which these branches were derived are marked. **(B)** The inter-division times of all cells in (A) are plotted against their CD25 expression levels. **(C)** The cells in the marked branches from (A). Blue numbers (1-4) indicate the number of cell divisions starting from the division of the common mother cell. The inter-division times of these cells are compared within their respective generation (1-4, blue) between branch A and branch B **(D).** Numbers with x indicate the factor between the mean inter-division times of the cells in branches A and B. Student's t-test: ****: $p<0.0001$.

Figure S26:

Duration of the initial TCR stimulus influences CD62L phenotype and division activity. (**A**) Scheme of the experimental setup. Briefly, naïve $CD8^+$ T cells were isolated and labeled with CTV before transfer to wells coated with anti-CD3/CD28 and containing 25U/mL IL-2. After 6h, 12h, 18h or 24h cells were transferred to new wells containing only anti-CD28, IL-2 and IL-12. (**B**) Overlaid histograms showing the CTV-profiles of cells stimulated for the indicated time frames at day 3 after activation. (**C**) Bar graph showing the percentage of cells in each CTV-division peak for the indicated time frames. The ANOVA-test was used to assess the statistical significance between the different stimulation conditions: ns: p-value > 0.05, *: p-value ≤ 0.05, **: p-value ≤ 0.01, ***: p-value ≤ 0.001, ****: p-value ≤ 0.0001. (**D**) Representative pseudo-color plots showing the CTV/CD62L profile of T cells for each stimulation condition. (**E**) Corresponding bar graph showing the percentage of CD62L negative cells. The ANOVA-test was used to assess the

statistical significance between the different stimulation conditions: *: p-value ≤ 0.05, **: p-value ≤ 0.01.

Figure S27:

Model comparison on the "brief" and "brief + IL12" groups with and without incorporating information from CD62L expression. Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log10-Bayes factors are shown for model #3 and model #3 + $\Delta CD62L^{+/-}$ in the three data groups under the "brief" and "brief+IL12" stimulation conditions. Model #3 + $\Delta CD62L^{+/-}$ sets the difference in the mean inter-division time of "Slow" and "Fast" subsets to the difference between the mean inter-division of $CD62L^{+}$ and $CD62L^{-}$ cells (Supplementary Methods). The AIC and BIC show support for model #3 + $\Delta CD62L^{+/-}$ compared to model #3.

Figure S28:

Modular division of a family tree for efficient sampling of latent variables. A family tree is divided into nine "subtrees". The first subtree (T_0) includes all cells belonging to generations 1, 2 and 3. The remaining subtrees ($T_1, T_2, \ldots, T_8$) are subtrees descending from the eight third-generation cells: T_1 includes all cells descending from cell c_1, T_2 includes all cells descending from cell c_2 and so forth.

SI Movie Legends:

Movie S1:

A single naïve OT-I cell was sorted on anti-CD3/CD28 (as in Fig. 1) and imaged for approximately four days. The numbers in the upper right corner indicate the time after start of the experiment (d – hh:mm:ss). The white circles or dots on the cells are virtual markers that were set manually to track the cells.

Movie S2:

A single activated OT-I cell was sorted into a "macrowell" without anti-CD3 coating and was imaged for approximately four days. Under these conditions, the T cells distributed over the complete well.

Movie S3:

A single naïve OT-I cell was sorted on anti-CD3/CD28 (as in Fig. 1) and imaged for approximately four days. The numbers in the upper right corner indicate the time after start of the experiment (d – hh:mm:ss). Every cell is marked with a colored tracing line to visualize the movement of the cells.

Movie S4:

A single naïve OT-I cell was sorted on anti-CD3/CD28 (as in Fig. 1) and imaged for approximately five days in the presence of anti-CD25-APC. The numbers in the upper right corner indicate the time after start of the experiment (d – hh:mm:ss). The bright dots on the cells are virtual markers that were set manually to track the cells. The bright field image is shown on the left side. The corresponding APC signal is shown on the right side. To reduce bleaching and phototoxicity, the APC channel was not acquired as often as the bright field. Thus, the APC channel does not always update together with the bright field image.

Fig. S1

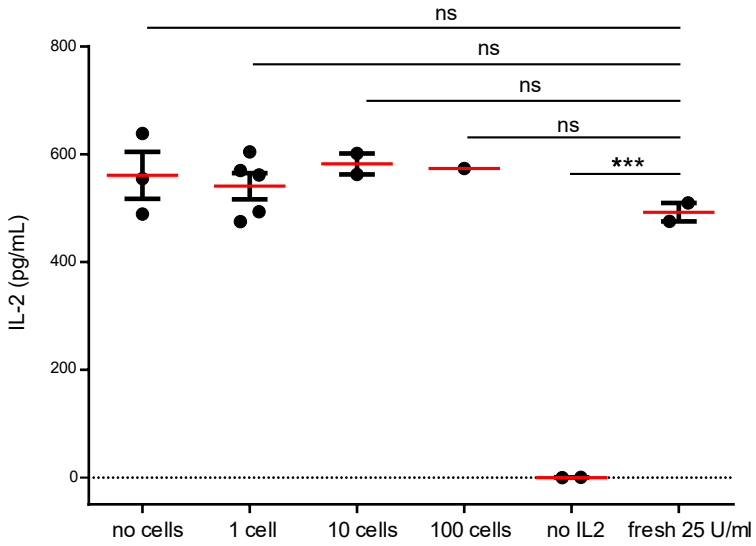

ns
ns
ns
ns
***
800
600
400
200
0
IL-2 (pg/mL)
no cells
1 cell
10 cells
100 cells
no IL2
fresh 25 U/ml

Fig. S2

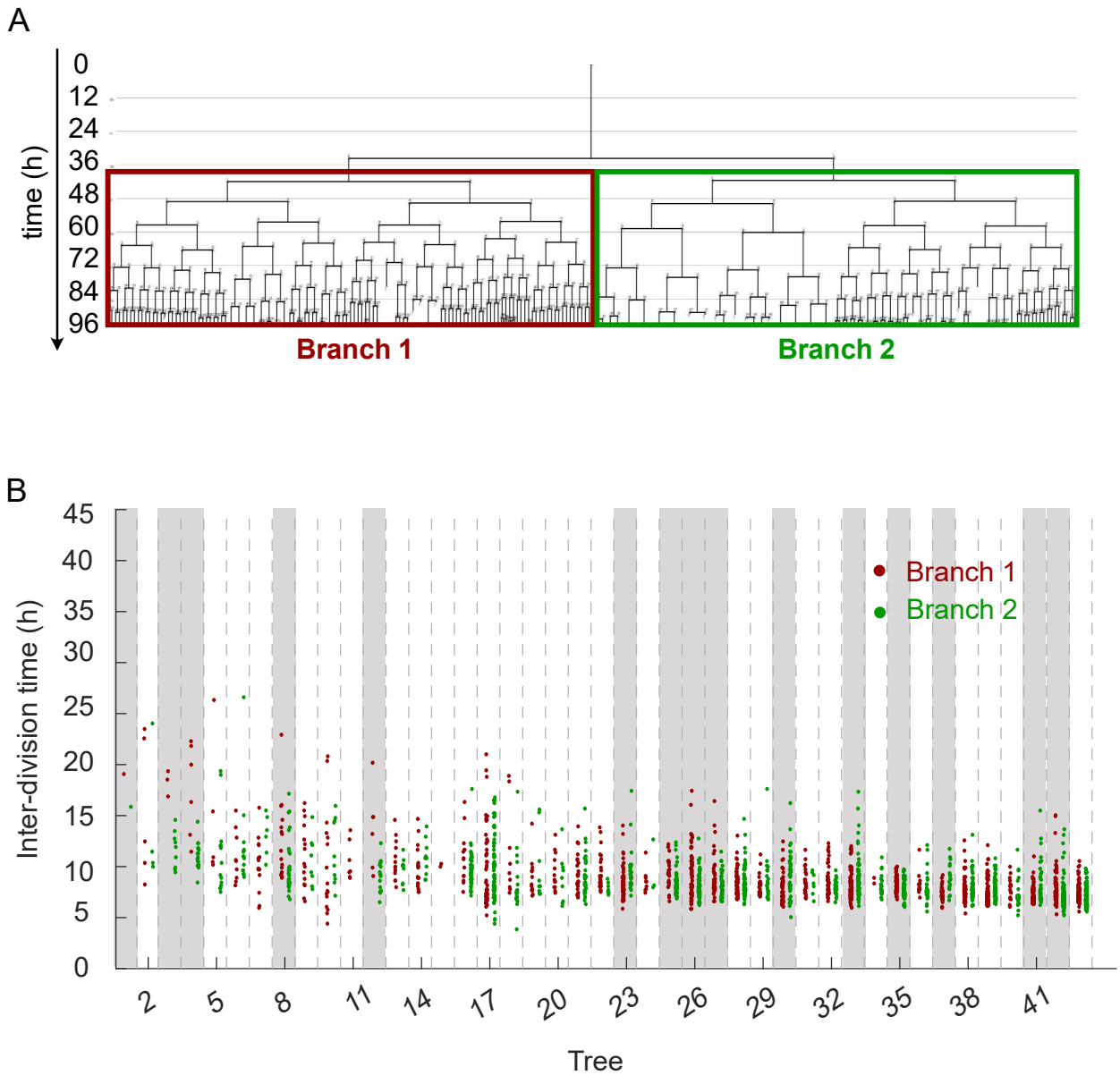

A
time (h)
0
12
24
36
48
60
72
84
96
Branch 1
Branch 2
B
Inter-division time (h)
45
40
35
30
25
20
15
10
5
0
Branch 1
Branch 2
2
5
8
11
14
17
20
23
26
29
32
35
38
41
Tree

Fig. S3

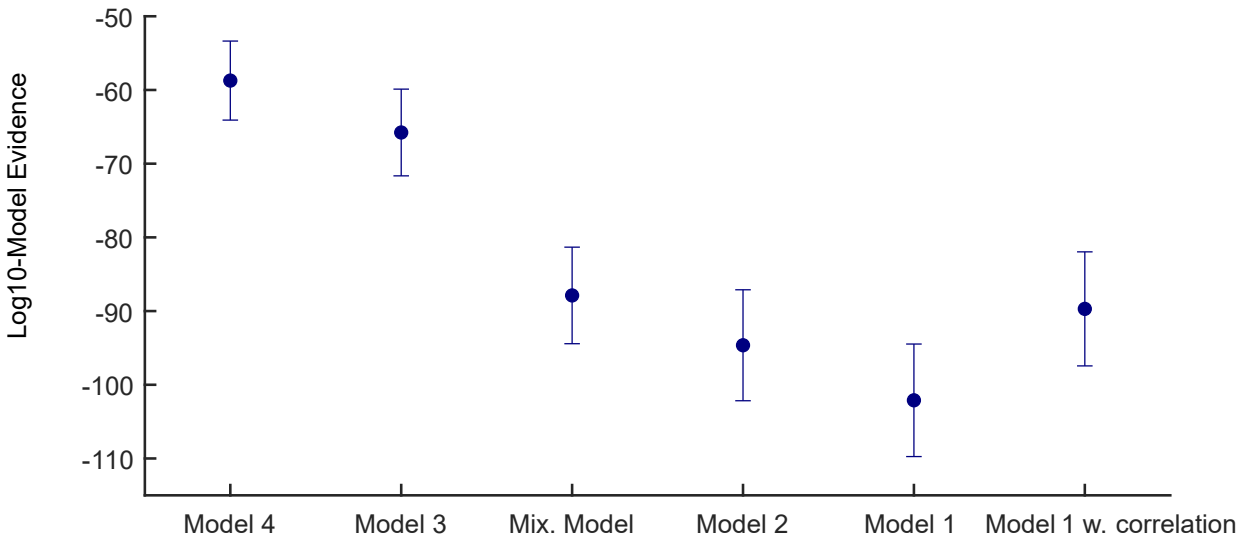

Log10-Model Evidence
-50
-60
-70
-80
-90
-100
-110
Model 4
Model 3
Mix. Model
Model 2
Model 1
Model 1 w. correlation

Fig. S4

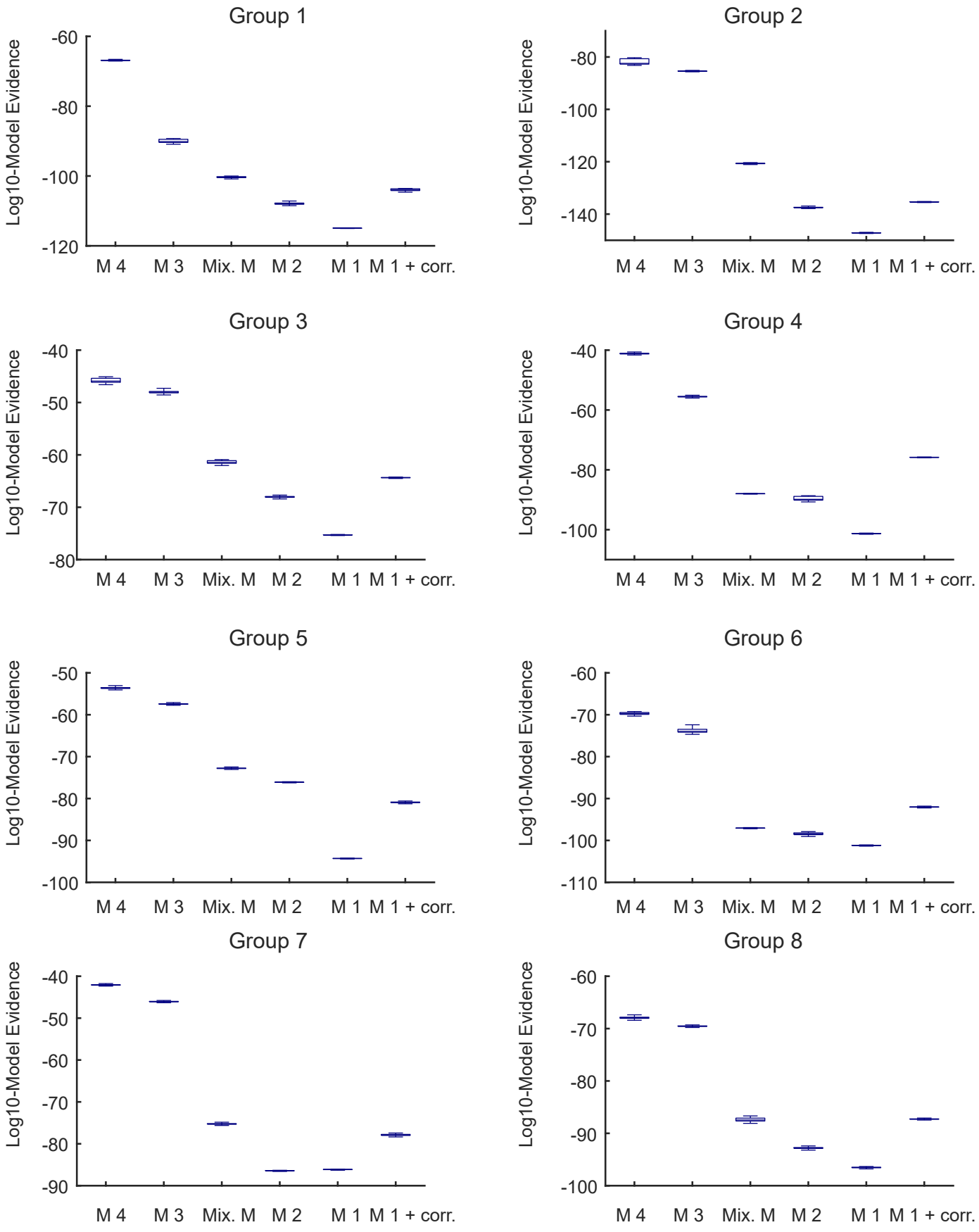

Group 1
Log10-Model Evidence
-60
-80
-100
-120
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 2
Log10-Model Evidence
-80
-100
-120
-140
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 3
Log10-Model Evidence
-40
-50
-60
-70
-80
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 4
Log10-Model Evidence
-40
-60
-80
-100
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 5
Log10-Model Evidence
-50
-60
-70
-80
-90
-100
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 6
Log10-Model Evidence
-60
-70
-80
-90
-100
-110
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 7
Log10-Model Evidence
-40
-50
-60
-70
-80
-90
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.
Group 8
Log10-Model Evidence
-60
-70
-80
-90
-100
M 4
M 3
Mix. M
M 2
M 1
M 1 + corr.

Fig. S5

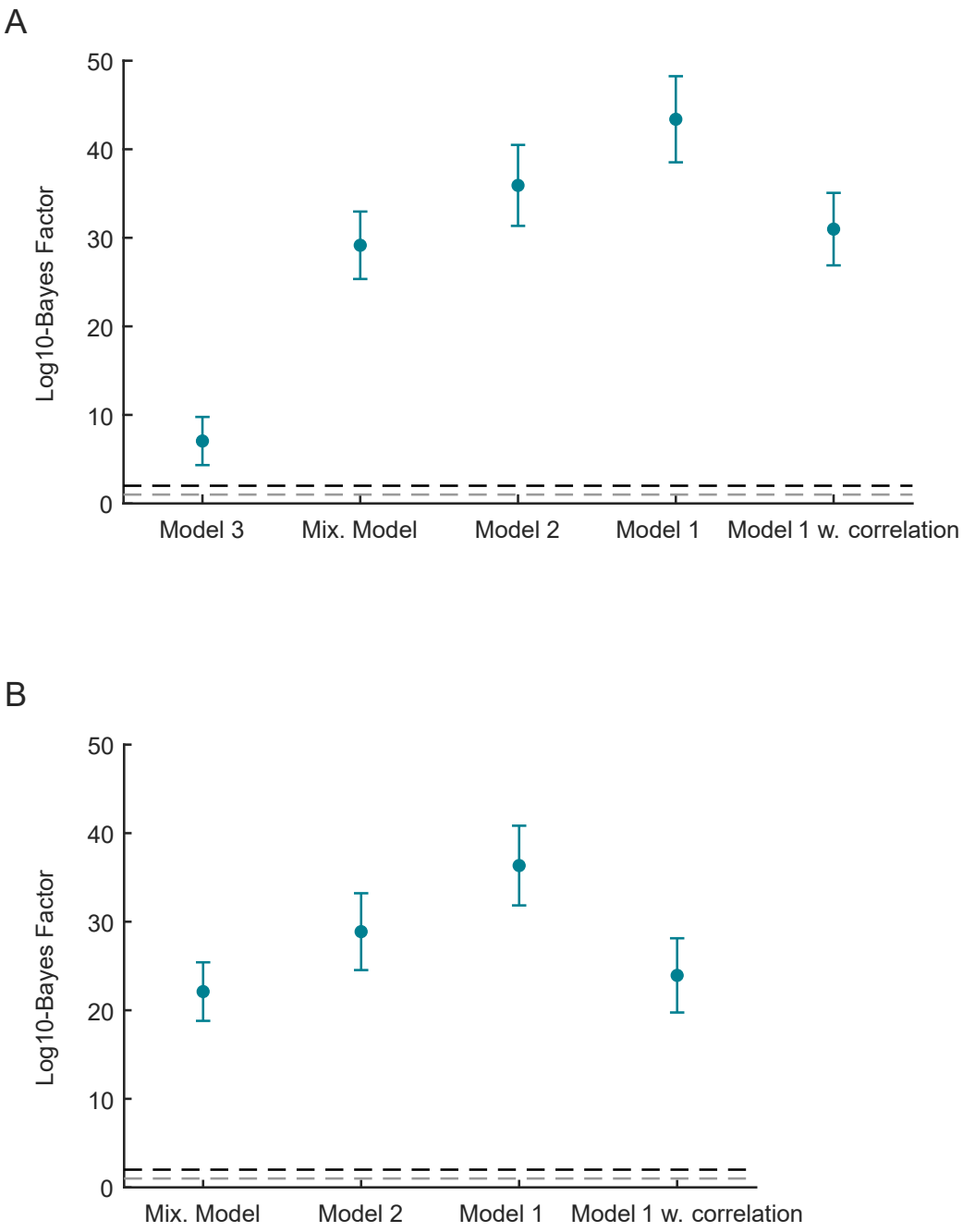

A
50
40
30
20
10
0
Log10-Bayes Factor
Model 3
Mix. Model
Model 2
Model 1
Model 1 w. correlation
B
50
40
30
20
10
0
Log10-Bayes Factor
Mix. Model
Model 2
Model 1
Model 1 w. correlation

Fig. S6

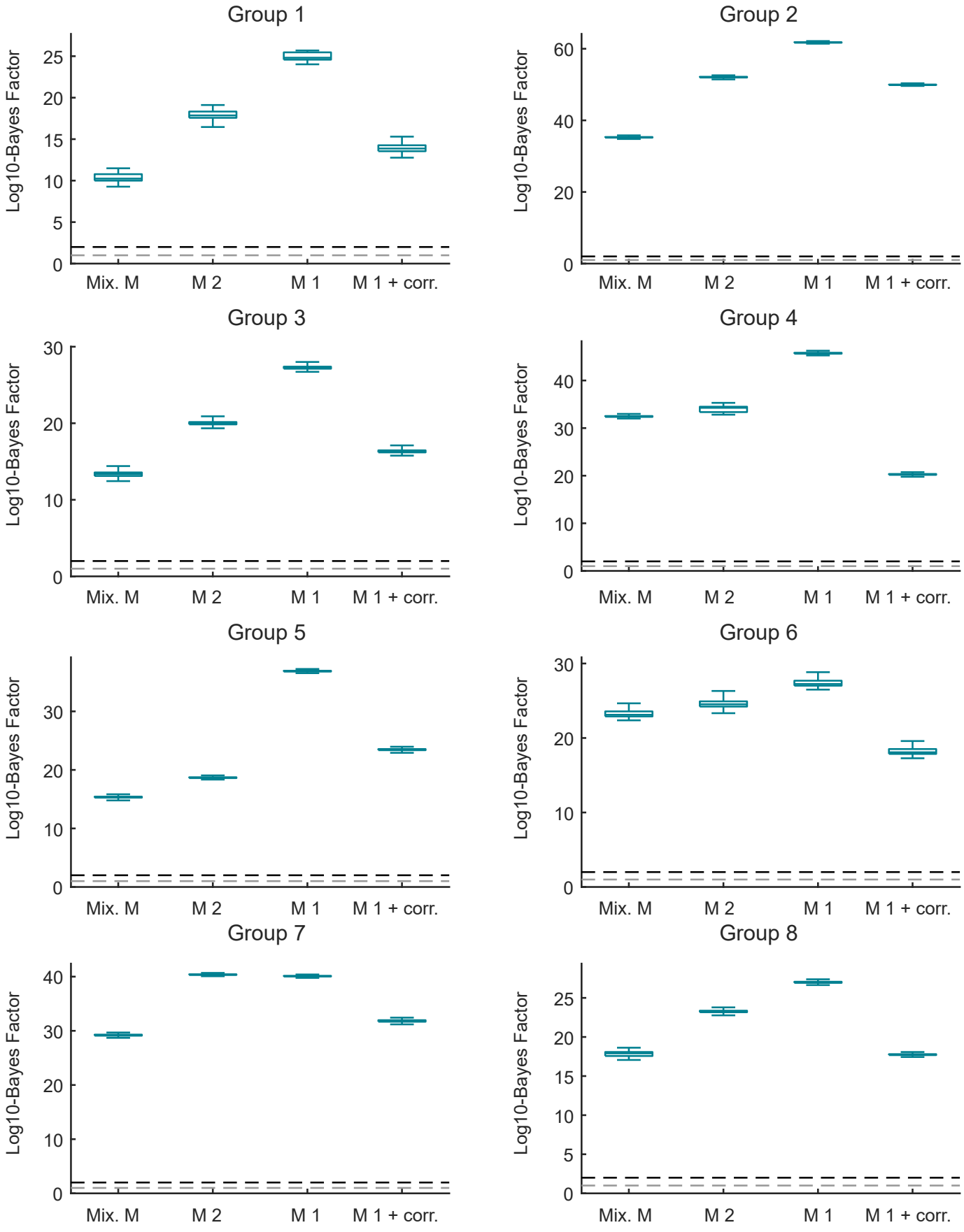

Group 1
Log10-Bayes Factor
0
5
10
15
20
25
Mix. M
M 2
M 1
M 1 + corr.
Group 2
Log10-Bayes Factor
0
20
40
60
Mix. M
M 2
M 1
M 1 + corr.
Group 3
Log10-Bayes Factor
0
10
20
30
Mix. M
M 2
M 1
M 1 + corr.
Group 4
Log10-Bayes Factor
0
10
20
30
40
Mix. M
M 2
M 1
M 1 + corr.
Group 5
Log10-Bayes Factor
0
10
20
30
Mix. M
M 2
M 1
M 1 + corr.
Group 6
Log10-Bayes Factor
0
10
20
30
Mix. M
M 2
M 1
M 1 + corr.
Group 7
Log10-Bayes Factor
0
10
20
30
40
Mix. M
M 2
M 1
M 1 + corr.
Group 8
Log10-Bayes Factor
0
5
10
15
20
25
Mix. M
M 2
M 1
M 1 + corr.

Fig. S7

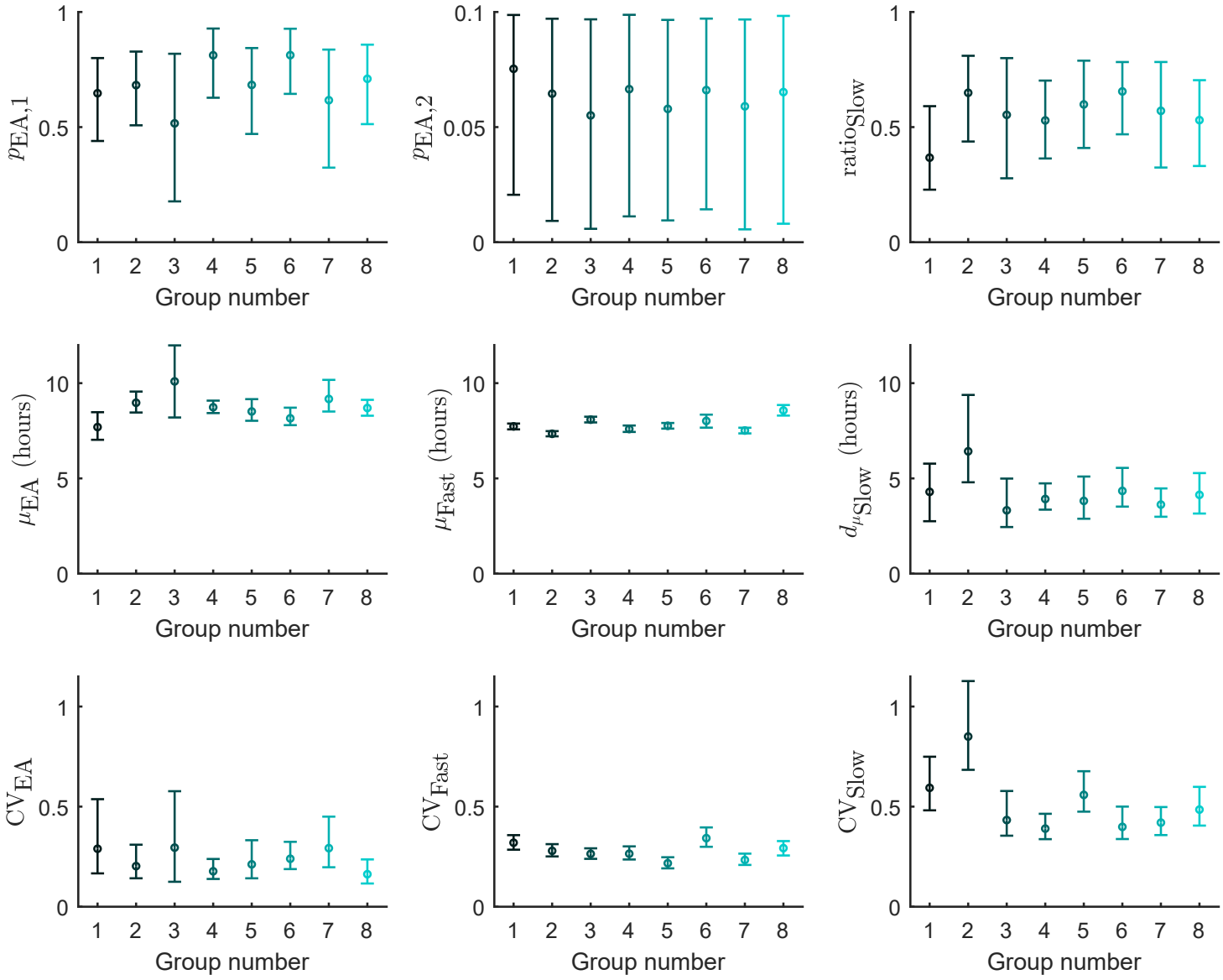

$p_{\mathrm{EA},1}$
$p_{\mathrm{EA},2}$
$\mathrm{ratio}_{\mathrm{Slow}}$
μ_{EA} (hours)
μ_{Fast} (hours)
$d_{\mu_{\mathrm{Slow}}}$ (hours)
$\mathrm{CV}_{\mathrm{EA}}$
$\mathrm{CV}_{\mathrm{Fast}}$
$\mathrm{CV}_{\mathrm{Slow}}$
Group number

Fig. S8

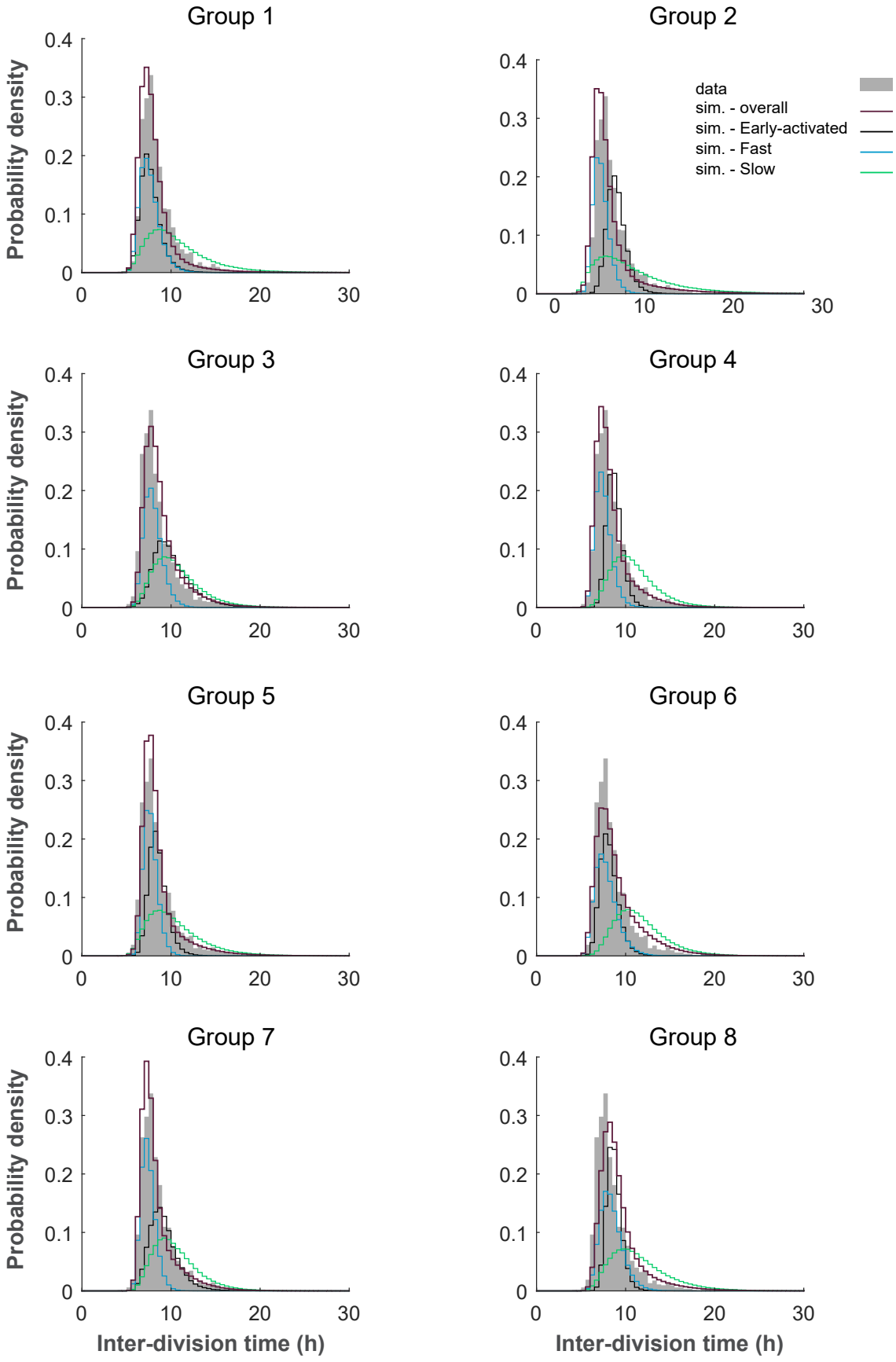

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Probability density
Inter-division time (h)
data
sim. - overall
sim. - Early-activated
sim. - Fast
sim. - Slow
0
0.1
0.2
0.3
0.4
10
20
30

Fig. S9

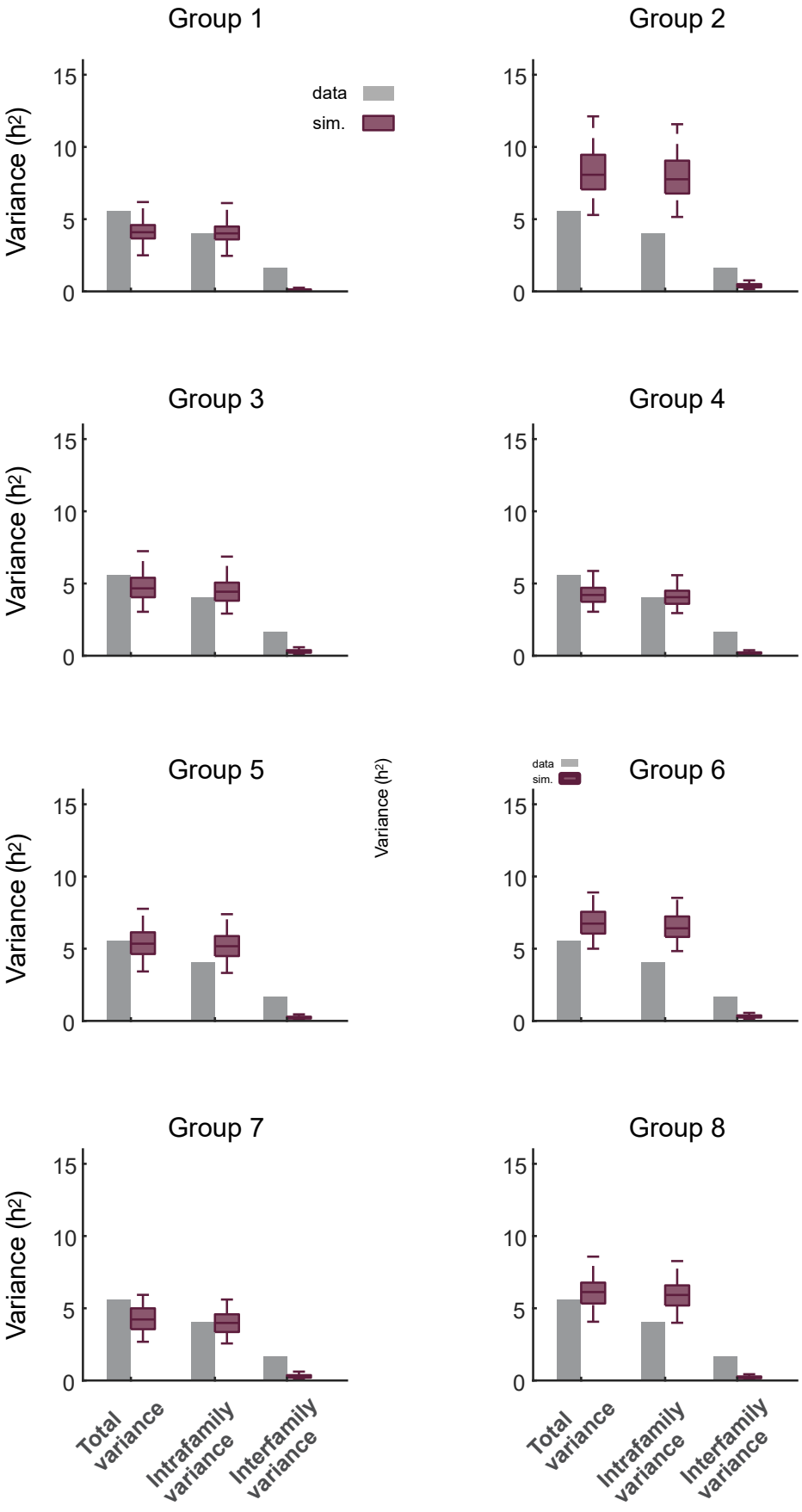

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
data
sim.
Variance (h²)
0
5
10
15
Total variance
Intrafamily variance
Interfamily variance

Fig. S10

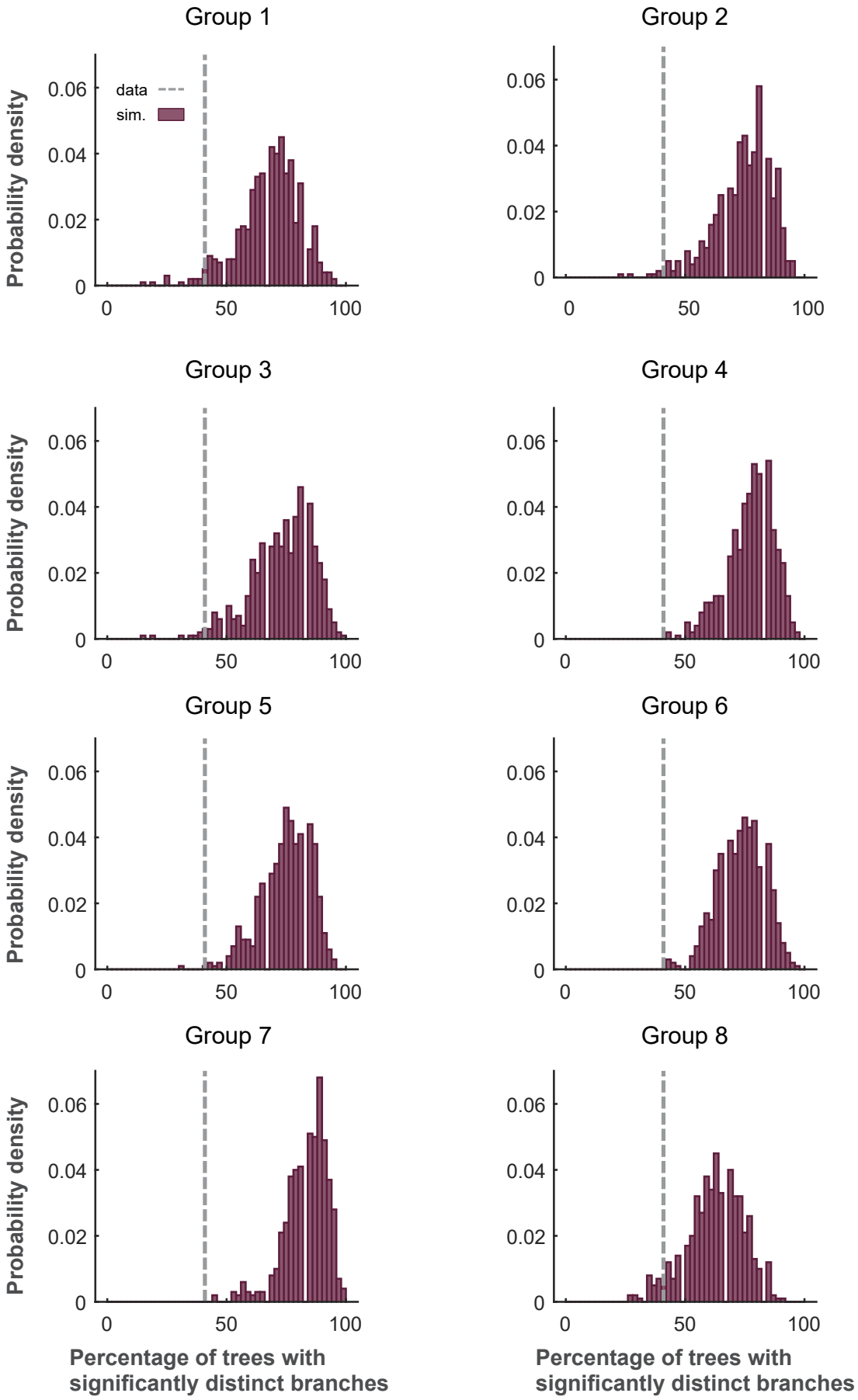

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
data
sim.
Probability density
0.06
0.04
0.02
0
0
50
100
Percentage of trees with significantly distinct branches

Fig. S11

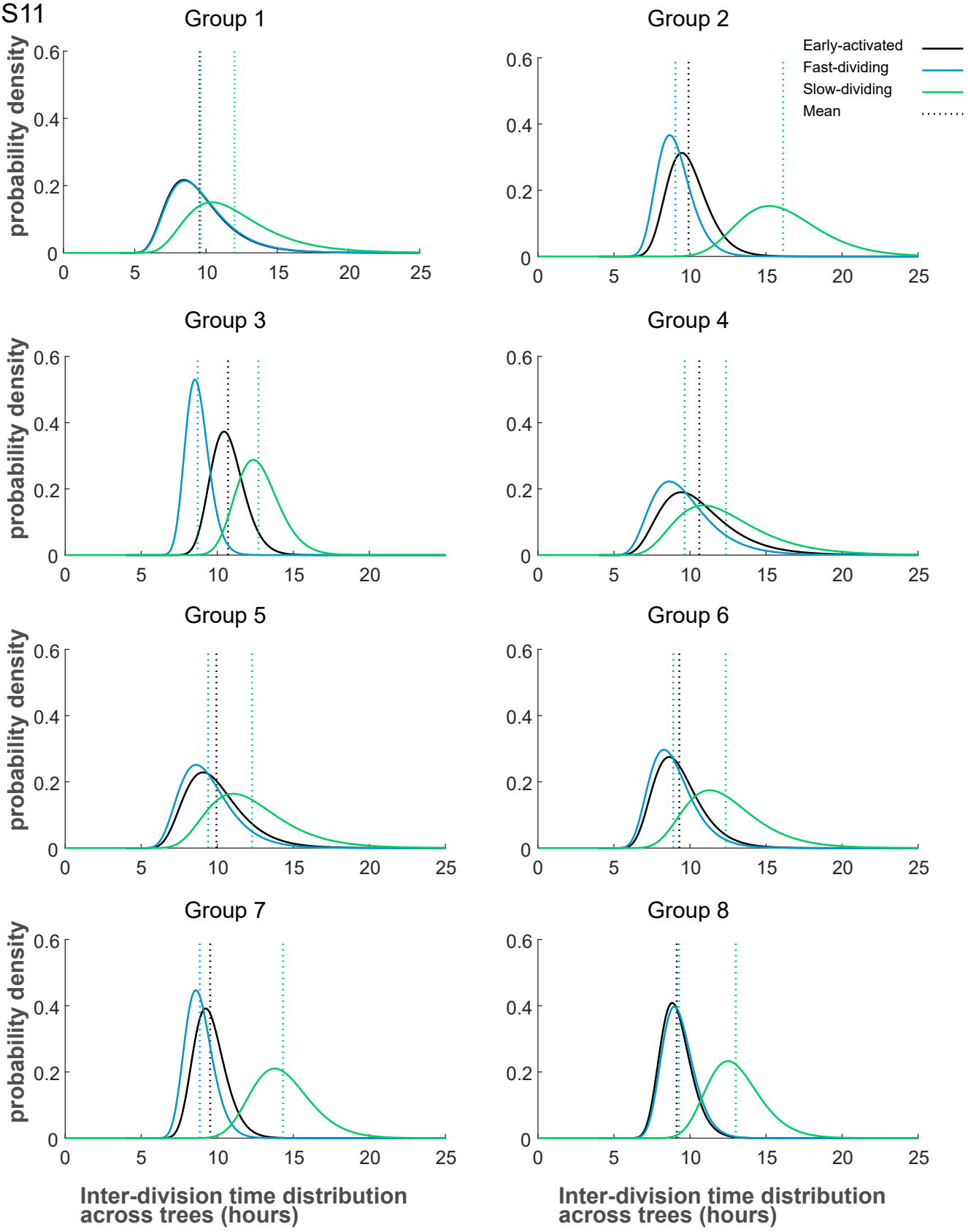

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Early-activated
Fast-dividing
Slow-dividing
Mean
probability density
0
0.2
0.4
0.6
0
5
10
15
20
25
Inter-division time distribution across trees (hours)

Fig. S12

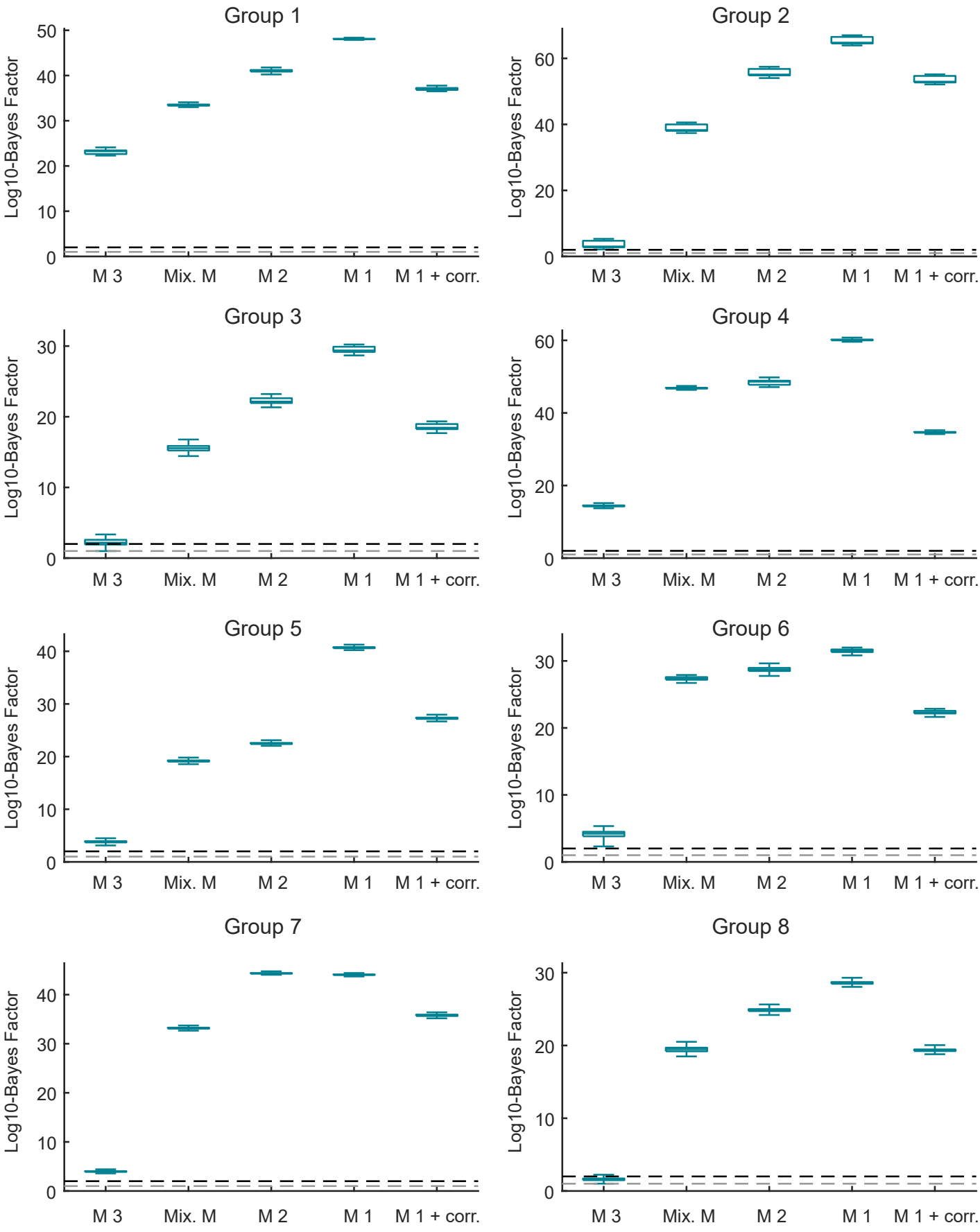

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Log10-Bayes Factor
M 3
Mix. M
M 2
M 1
M 1 + corr.

Fig. S13

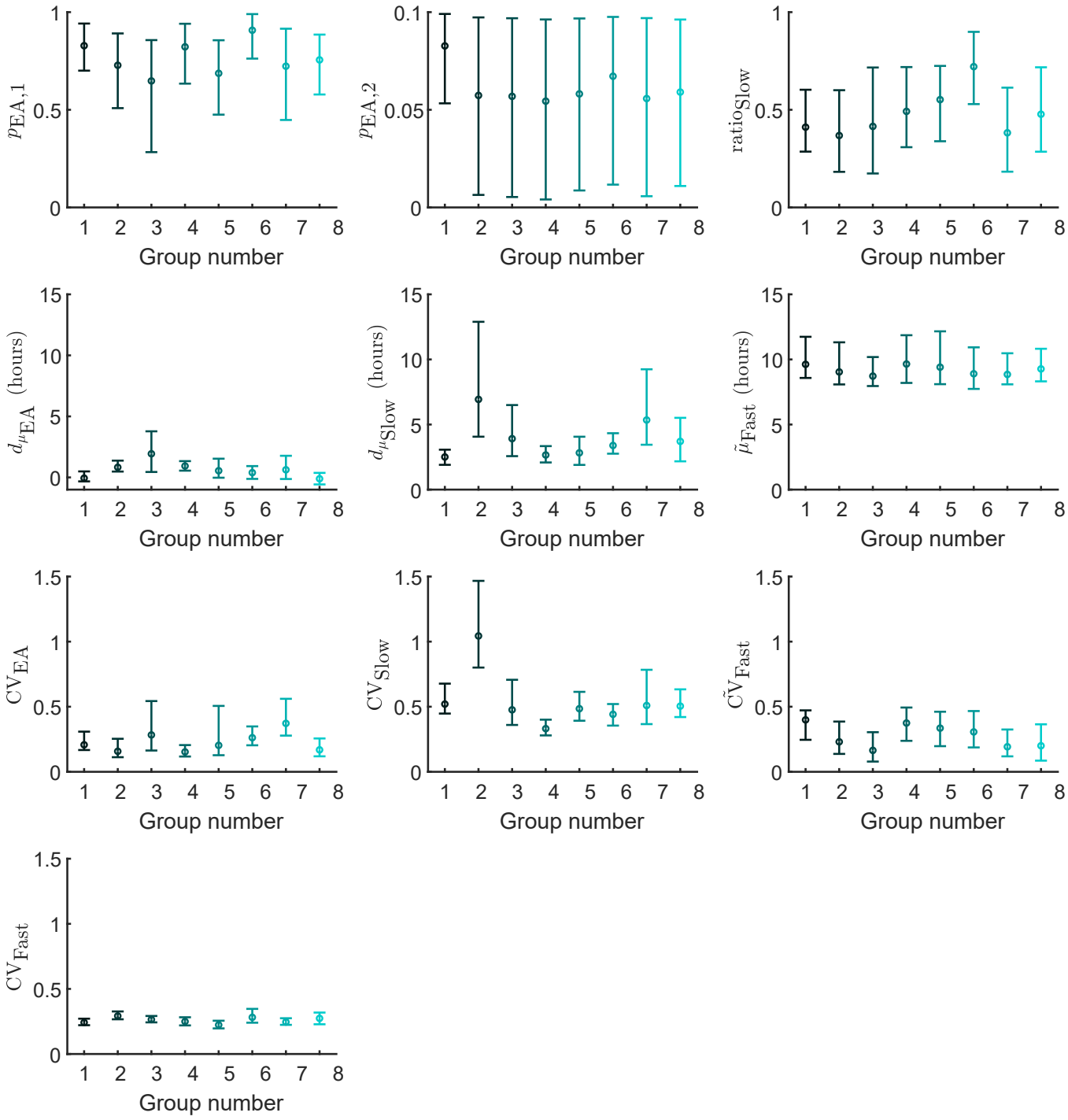

$p_{\mathrm{EA},1}$
$p_{\mathrm{EA},2}$
$\mathrm{ratio}_{\mathrm{Slow}}$
$d_{\mu_{\mathrm{EA}}}$ (hours)
$d_{\mu_{\mathrm{Slow}}}$ (hours)
$\tilde{\mu}_{\mathrm{Fast}}$ (hours)
$\mathrm{CV}_{\mathrm{EA}}$
$\mathrm{CV}_{\mathrm{Slow}}$
$\tilde{\mathrm{C}}\mathrm{V}_{\mathrm{Fast}}$
$\mathrm{CV}_{\mathrm{Fast}}$
Group number

Fig. S14

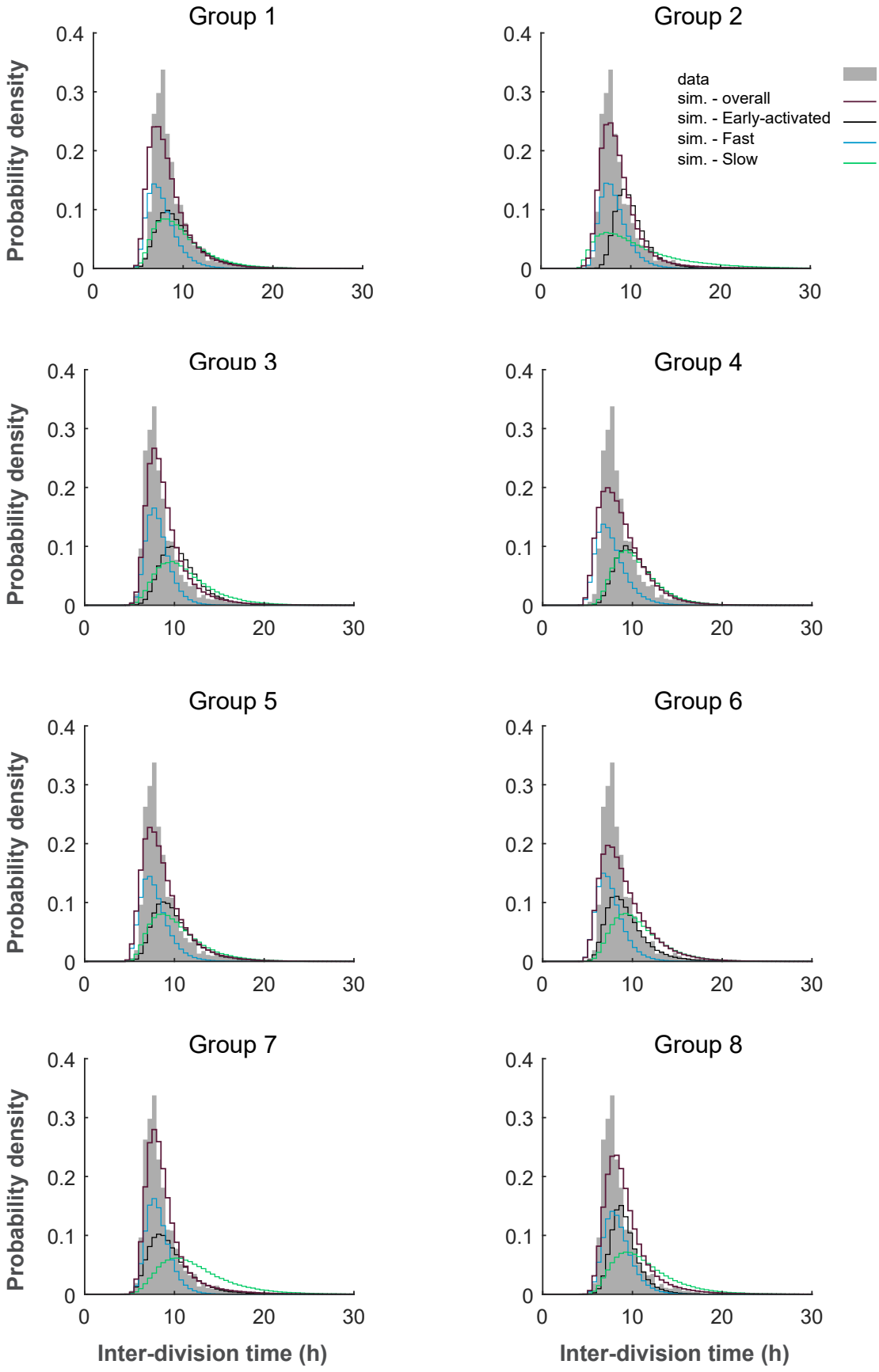

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
Probability density
Inter-division time (h)
data
sim. - overall
sim. - Early-activated
sim. - Fast
sim. - Slow
0
0.1
0.2
0.3
0.4
10
20
30

Fig. S15

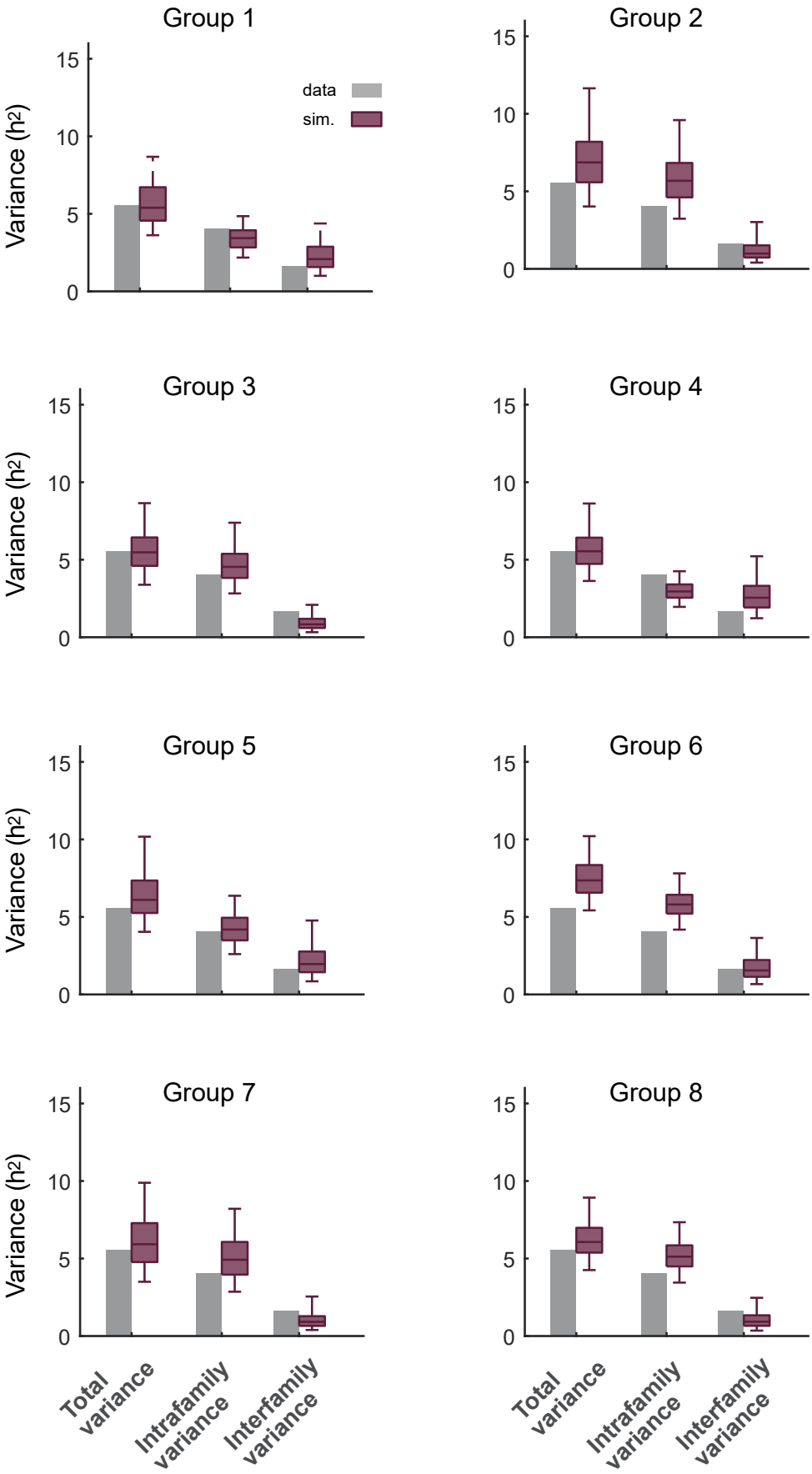

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
data
sim.
Variance (h²)
0
5
10
15
Total variance
Intrafamily variance
Interfamily variance

Fig. S16

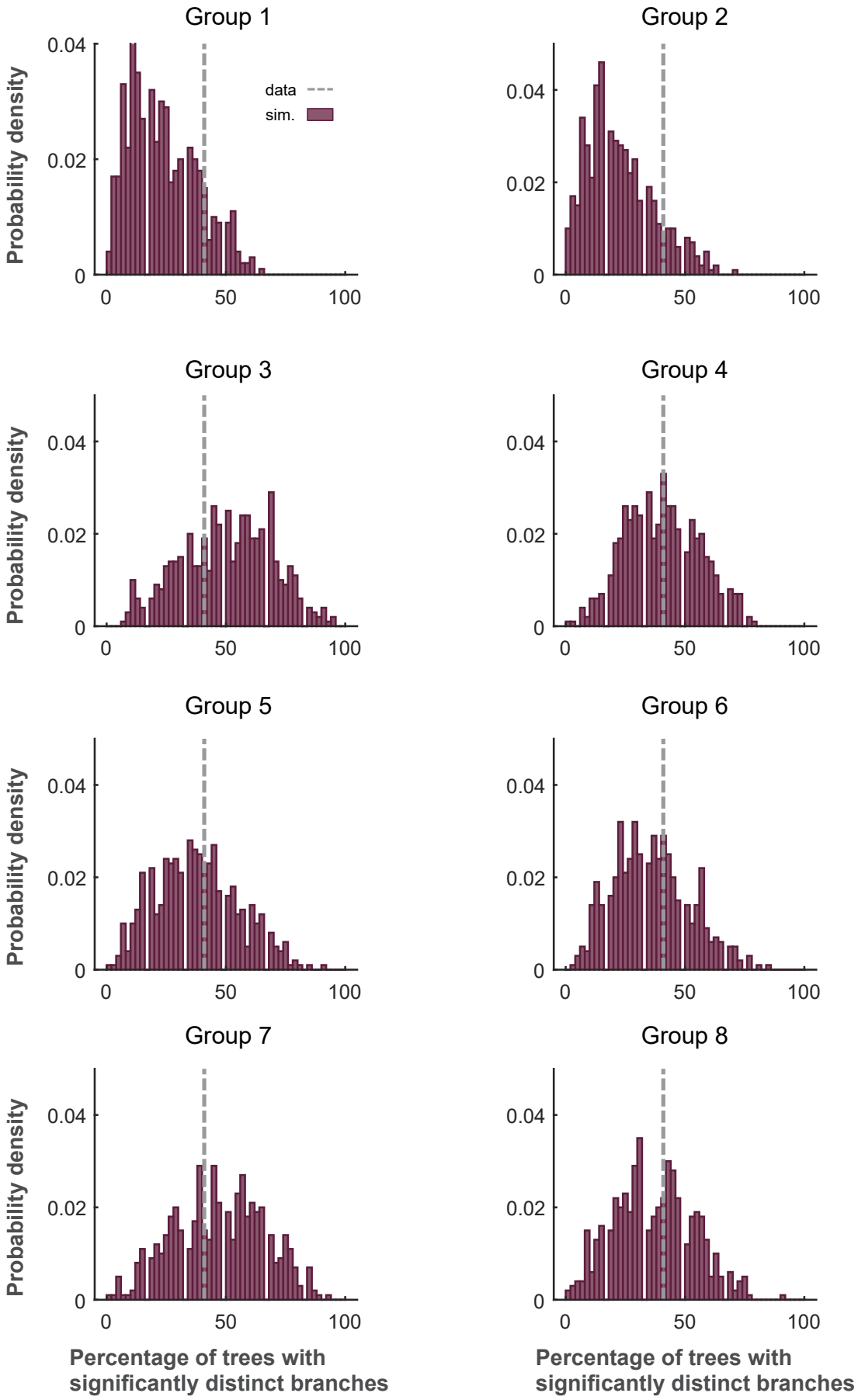

Group 1
Group 2
Group 3
Group 4
Group 5
Group 6
Group 7
Group 8
data
sim.
Probability density
0.04
0.02
0
0
50
100
Percentage of trees with significantly distinct branches

Fig. S17

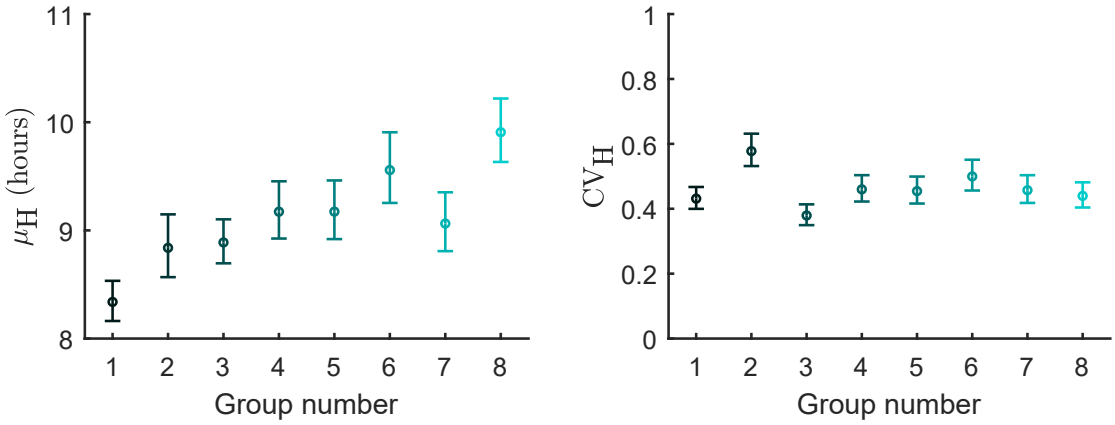

11
10
9
8
μ_H (hours)
1 2 3 4 5 6 7 8
Group number
1
0.8
0.6
0.4
0.2
0
CV_H
1 2 3 4 5 6 7 8
Group number

Fig. S18

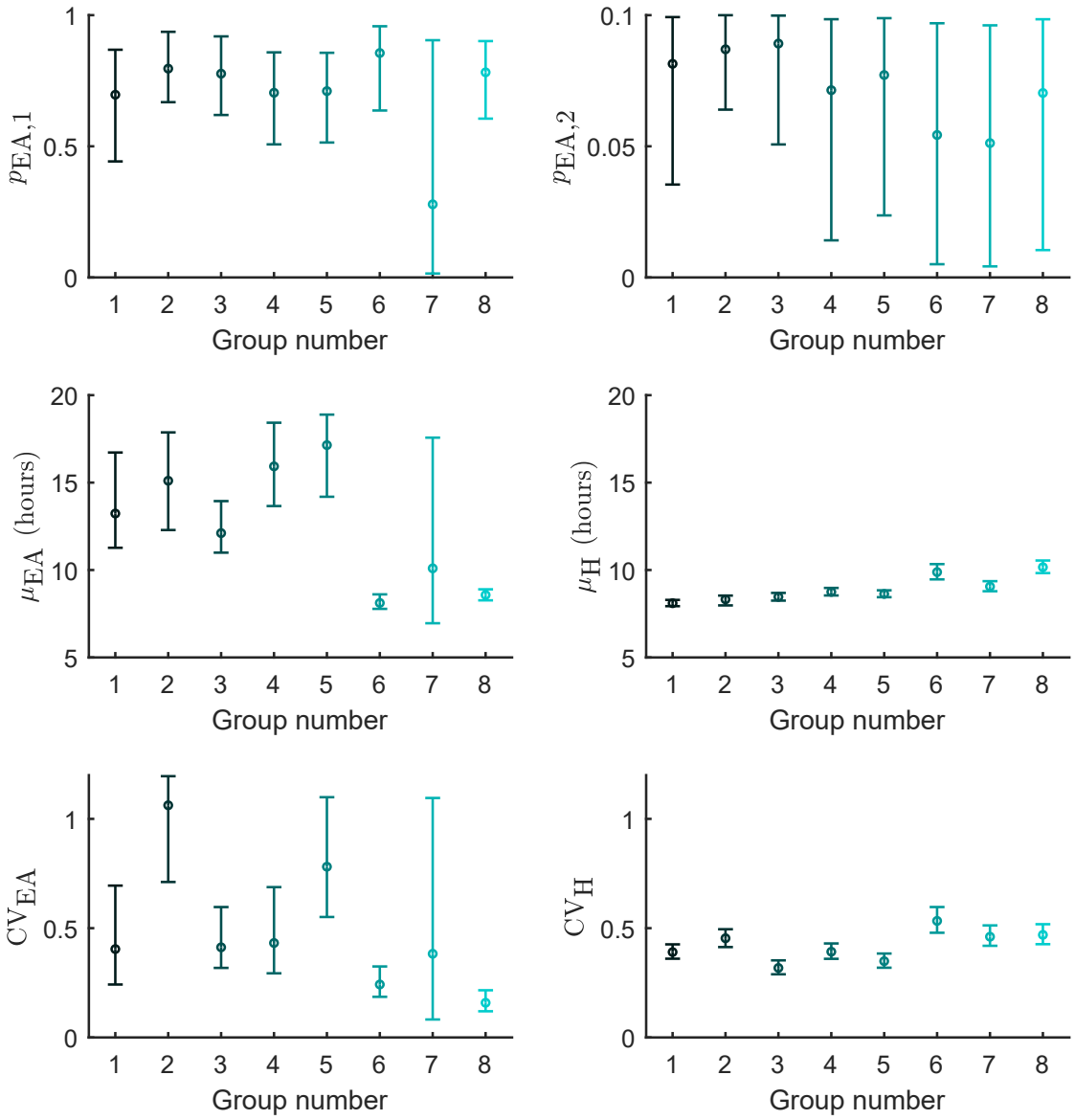

$p_{\mathrm{EA},1}$
$p_{\mathrm{EA},2}$
μ_{EA} (hours)
μ_{H} (hours)
$\mathrm{CV}_{\mathrm{EA}}$
CV_{H}
Group number

Fig. S19

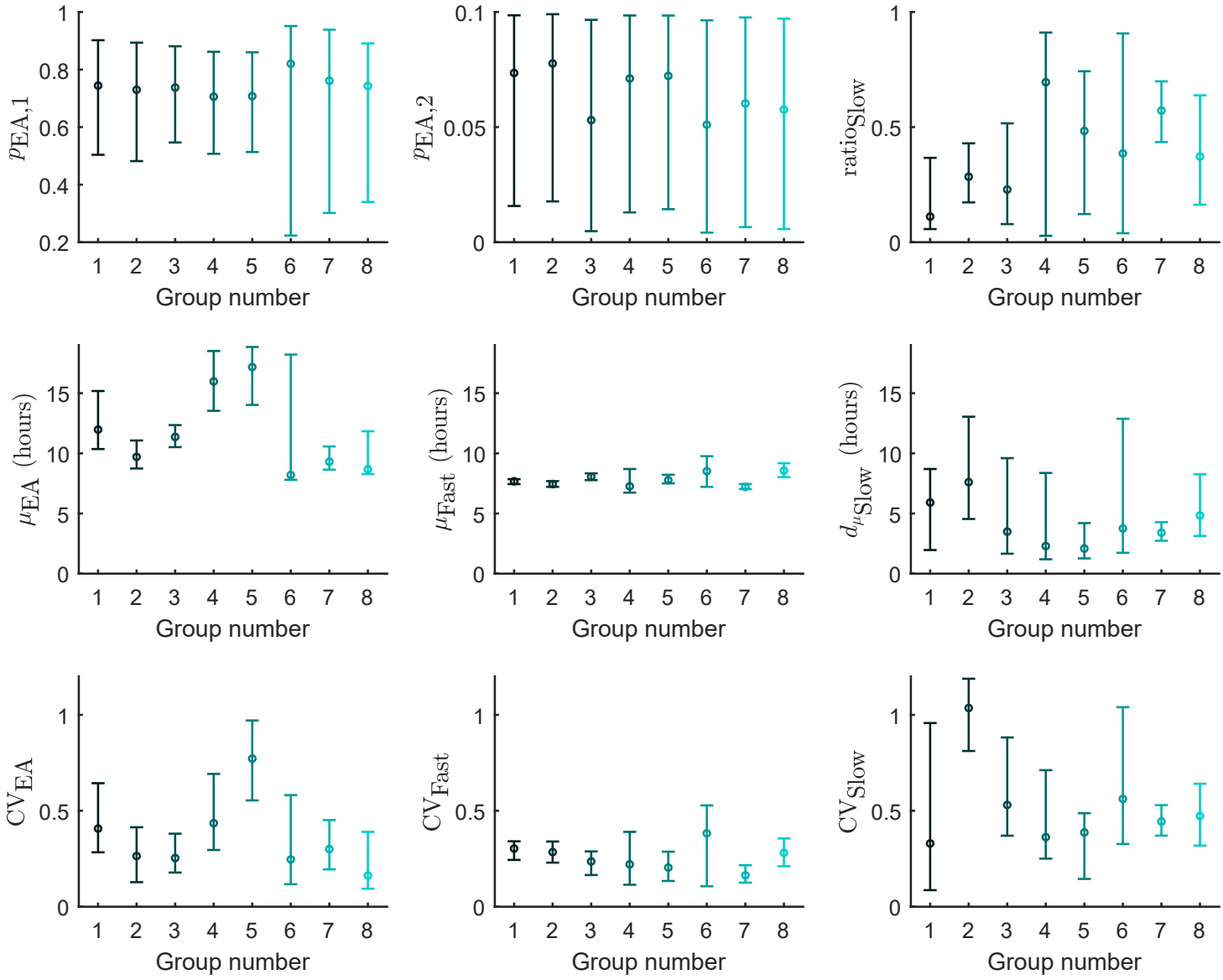

$p_{\mathrm{EA},1}$
$p_{\mathrm{EA},2}$
$\mathrm{ratio}_{\mathrm{Slow}}$
μ_{EA} (hours)
μ_{Fast} (hours)
$d_{\mu_{\mathrm{Slow}}}$ (hours)
$\mathrm{CV}_{\mathrm{EA}}$
$\mathrm{CV}_{\mathrm{Fast}}$
$\mathrm{CV}_{\mathrm{Slow}}$
Group number

Fig. S20

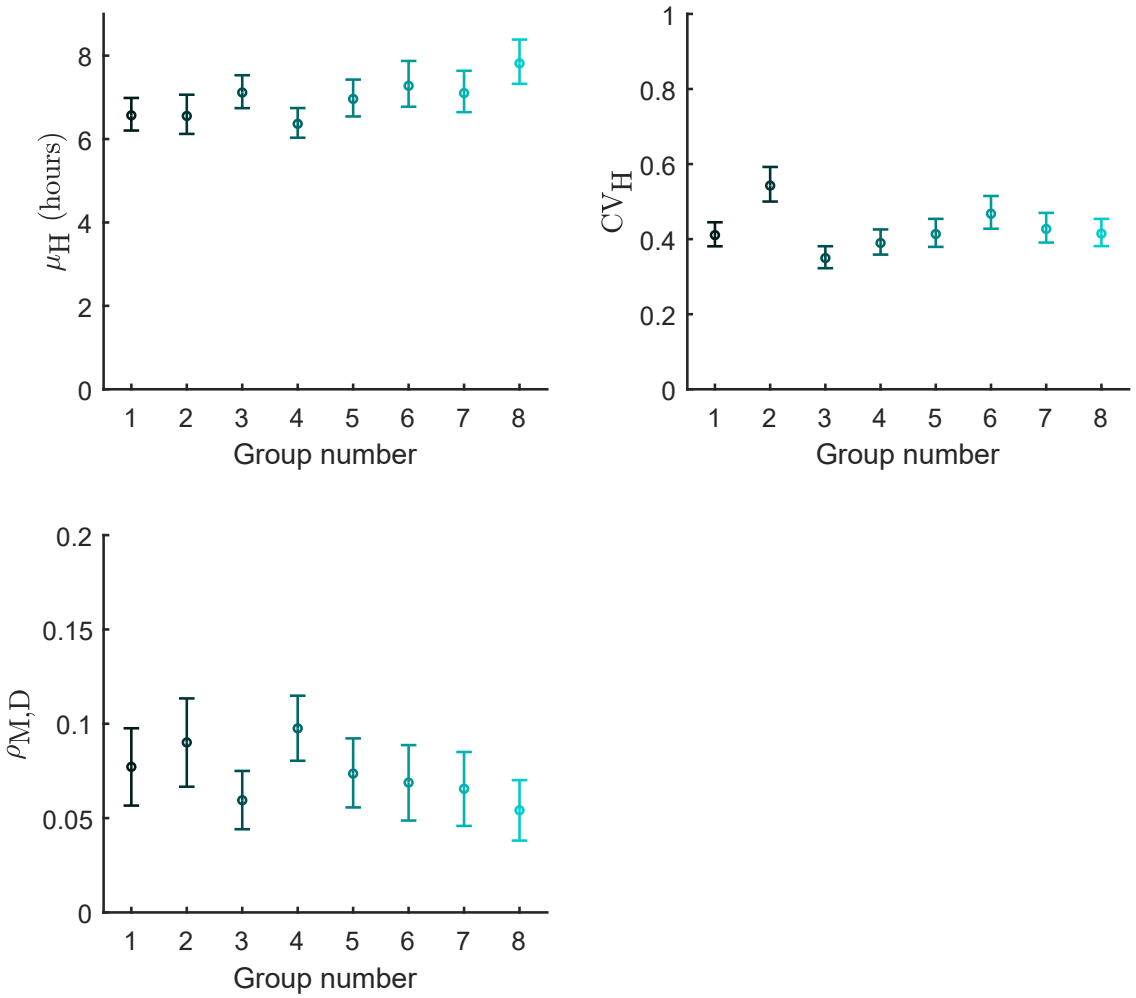


Fig. S21

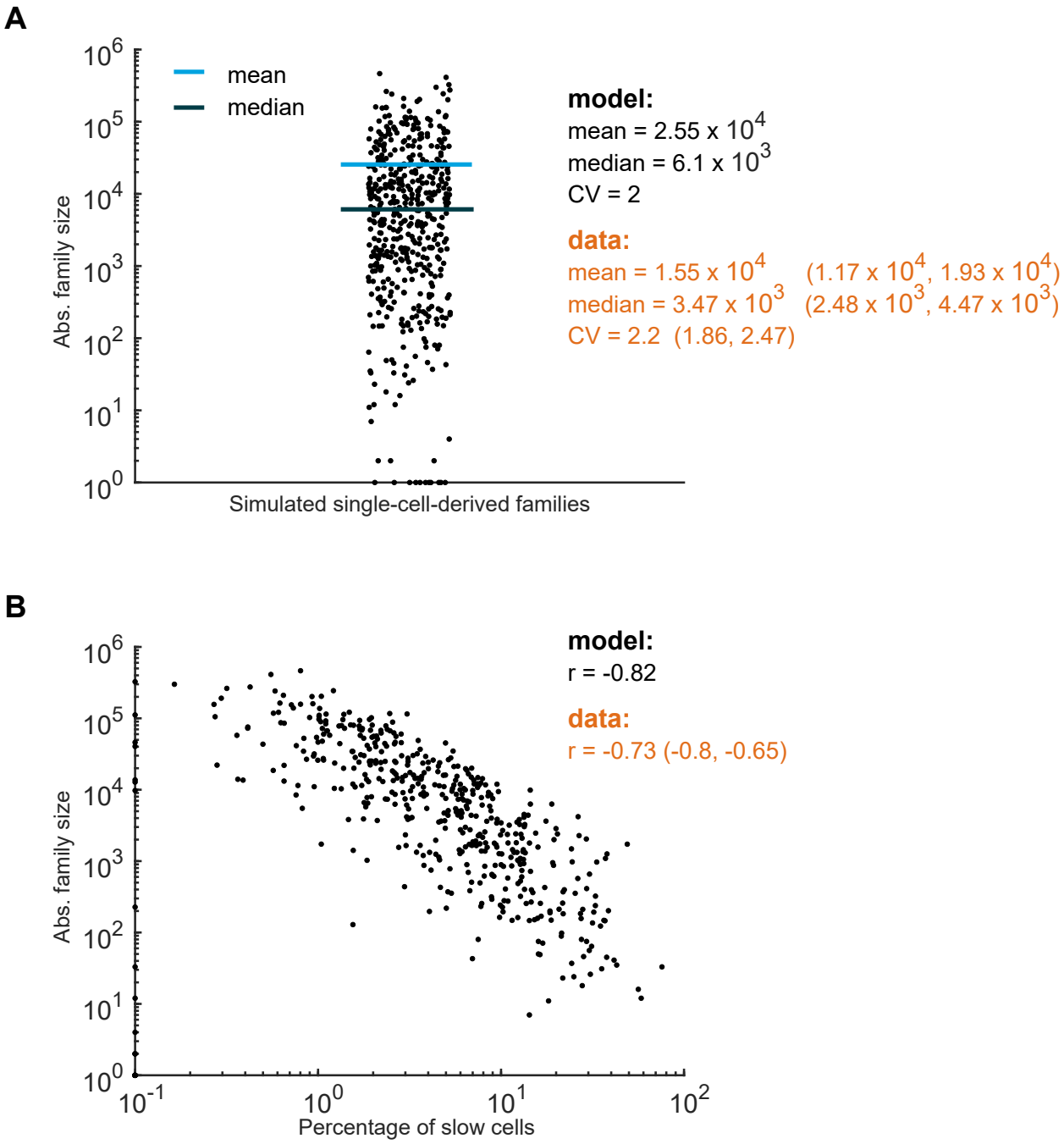

A
mean
median
model:
mean = 2.55 x 10^4
median = 6.1 x 10^3
CV = 2
data:
mean = 1.55 x 10^4 (1.17 x 10^4, 1.93 x 10^4)
median = 3.47 x 10^3 (2.48 x 10^3, 4.47 x 10^3)
CV = 2.2 (1.86, 2.47)
Abs. family size
Simulated single-cell-derived families
B
model:
r = -0.82
data:
r = -0.73 (-0.8, -0.65)
Abs. family size
Percentage of slow cells

Fig. S22

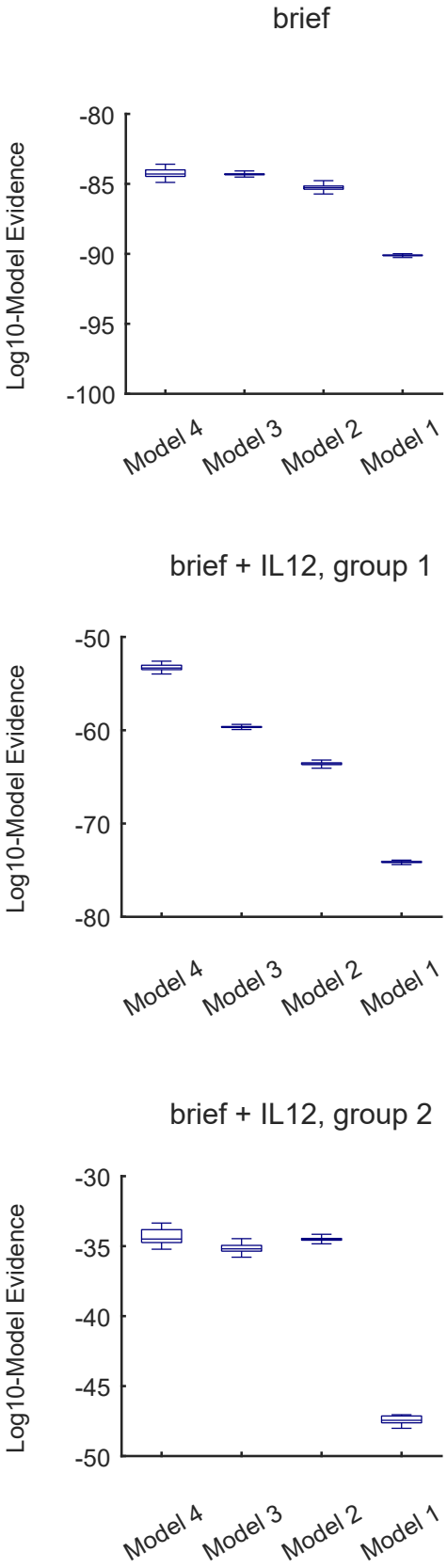

brief
Log10-Model Evidence
-80
-85
-90
-95
-100
Model 4
Model 3
Model 2
Model 1
brief + IL12, group 1
Log10-Model Evidence
-50
-60
-70
-80
Model 4
Model 3
Model 2
Model 1
brief + IL12, group 2
Log10-Model Evidence
-30
-35
-40
-45
-50
Model 4
Model 3
Model 2
Model 1

Fig. S23

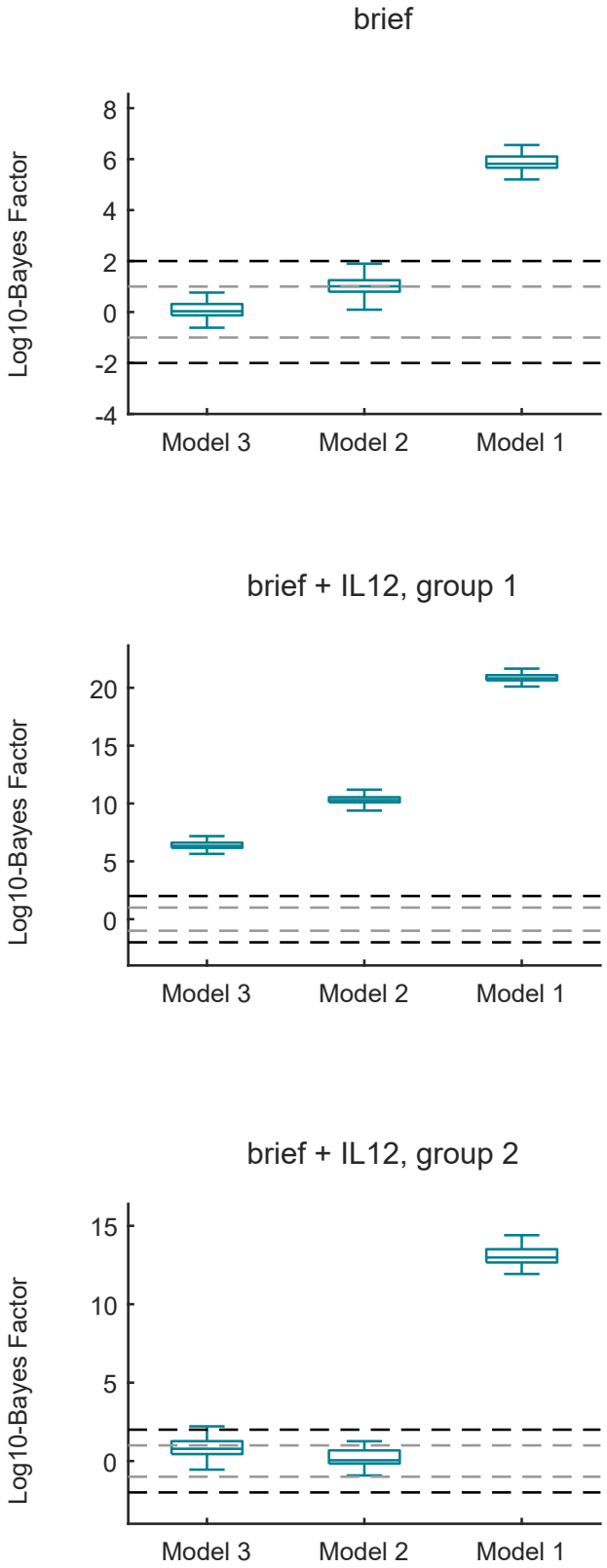

brief
Log10-Bayes Factor
8
6
4
2
0
-2
-4
Model 3
Model 2
Model 1
brief + IL12, group 1
Log10-Bayes Factor
20
15
10
5
0
Model 3
Model 2
Model 1
brief + IL12, group 2
Log10-Bayes Factor
15
10
5
0
Model 3
Model 2
Model 1

Fig. S24

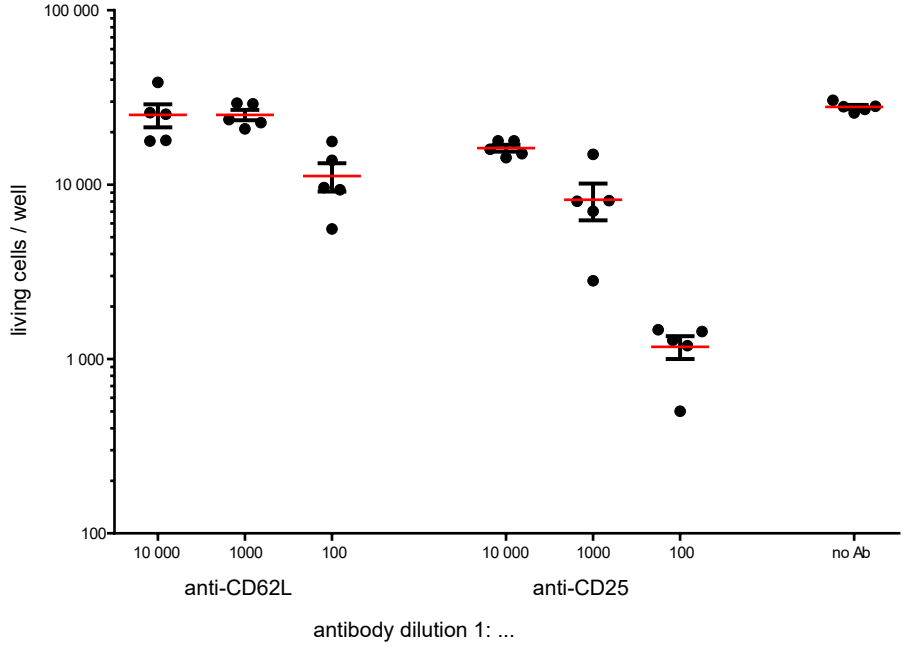
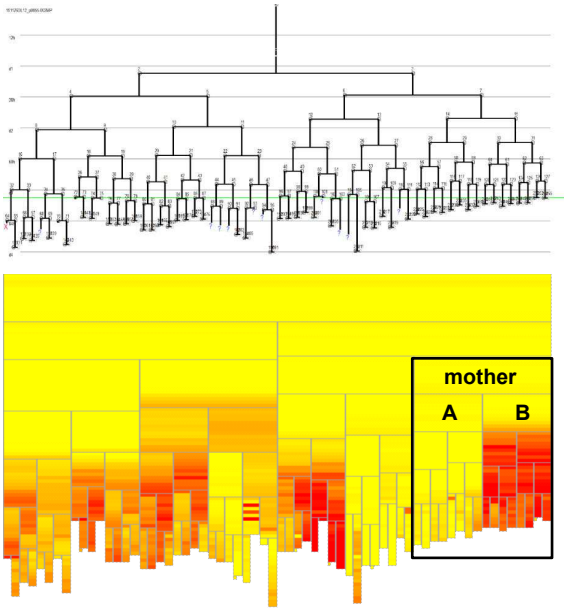
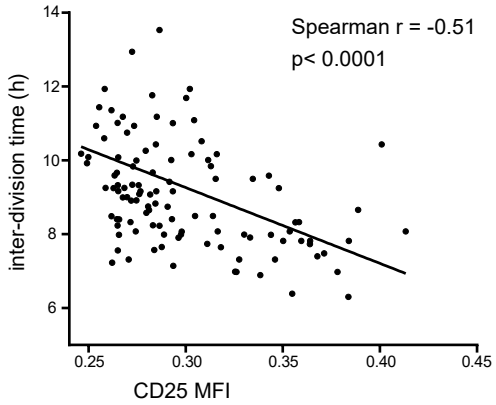

100 000
10 000
1 000
100
living cells / well
10 000
1000
100
10 000
1000
100
no Ab
anti-CD62L
anti-CD25
antibody dilution 1: ...

Fig. S25

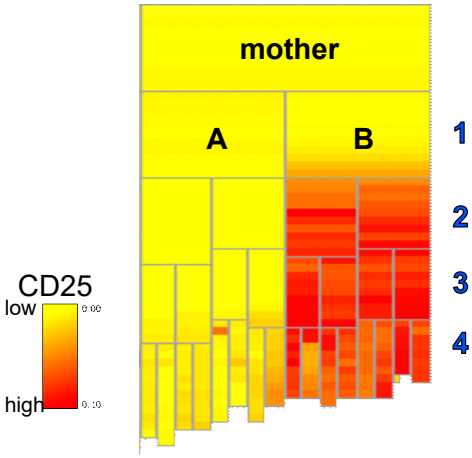
A


mother
A
B

B


Spearman r = -0.51
p< 0.0001
inter-division time (h)
CD25 MFI

C


CD25
low
high
mother
A
B
1
2
3
4

D

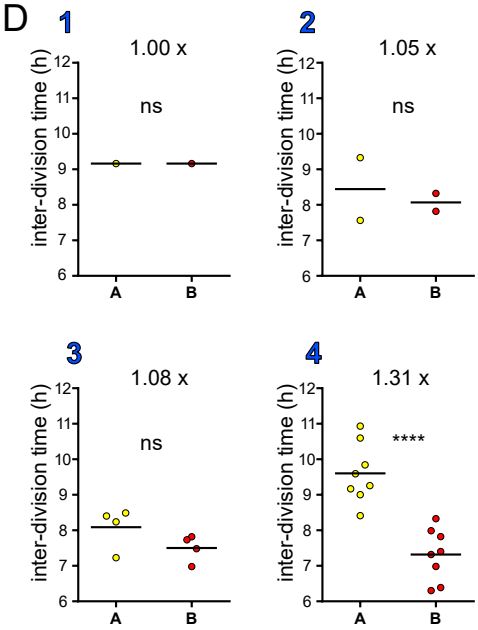

1
1.00 x
ns
2
1.05 x
ns
3
1.08 x
ns
4
1.31 x
****
inter-division time (h)
A
B

Fig. S26

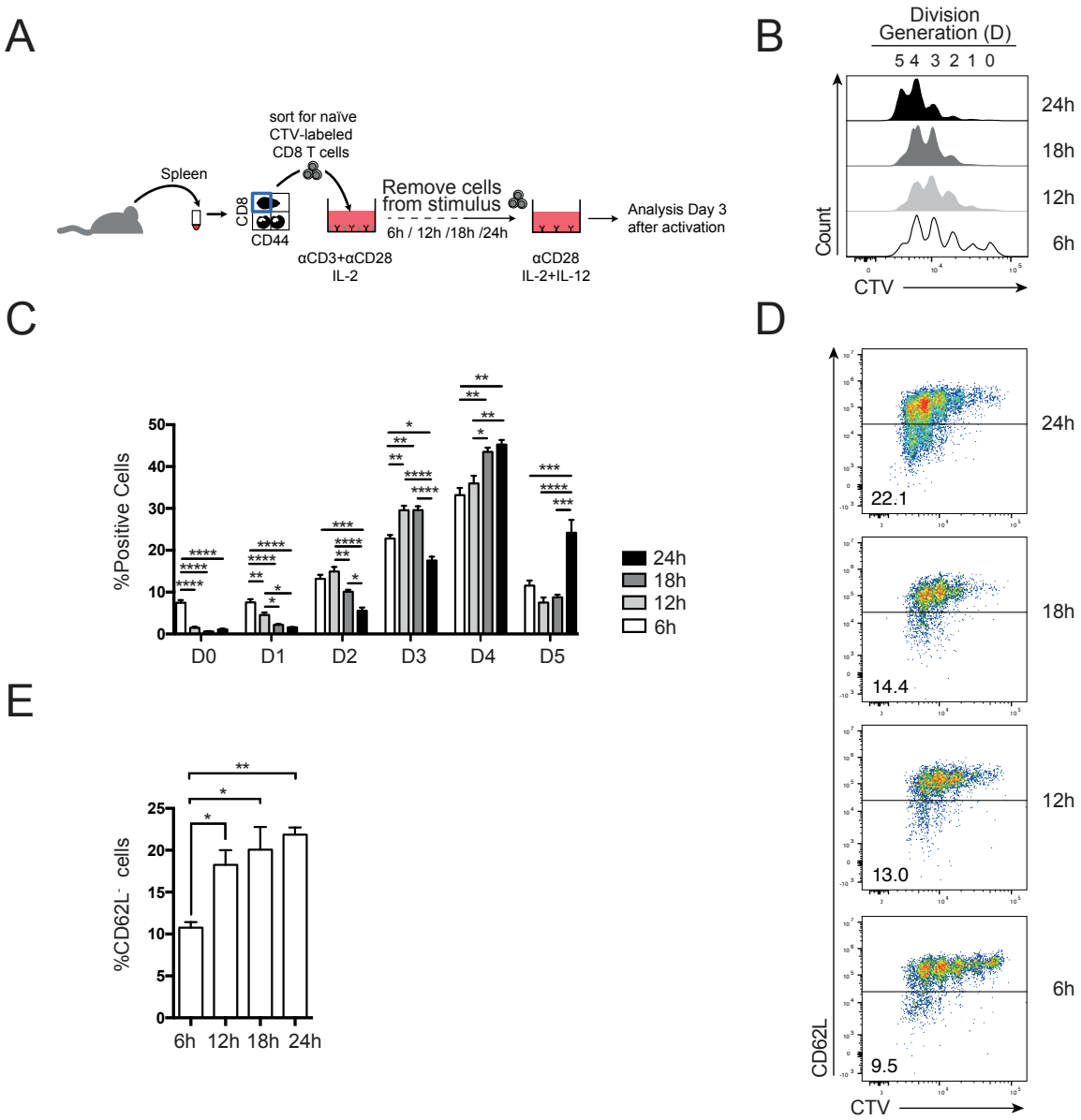

A
sort for naïve
CTV-labeled
CD8 T cells
Spleen
CD8
CD44
Remove cells
from stimulus
6h / 12h /18h /24h
Analysis Day 3
after activation
αCD3+αCD28
IL-2
αCD28
IL-2+IL-12
B
Division
Generation (D)
5 4 3 2 1 0
24h
18h
12h
6h
Count
CTV
C
%Positive Cells
D0
D1
D2
D3
D4
D5
24h
18h
12h
6h
D
22.1
14.4
13.0
9.5
CD62L
CTV
E
%CD62L⁻ cells
6h 12h 18h 24h

Fig. S27

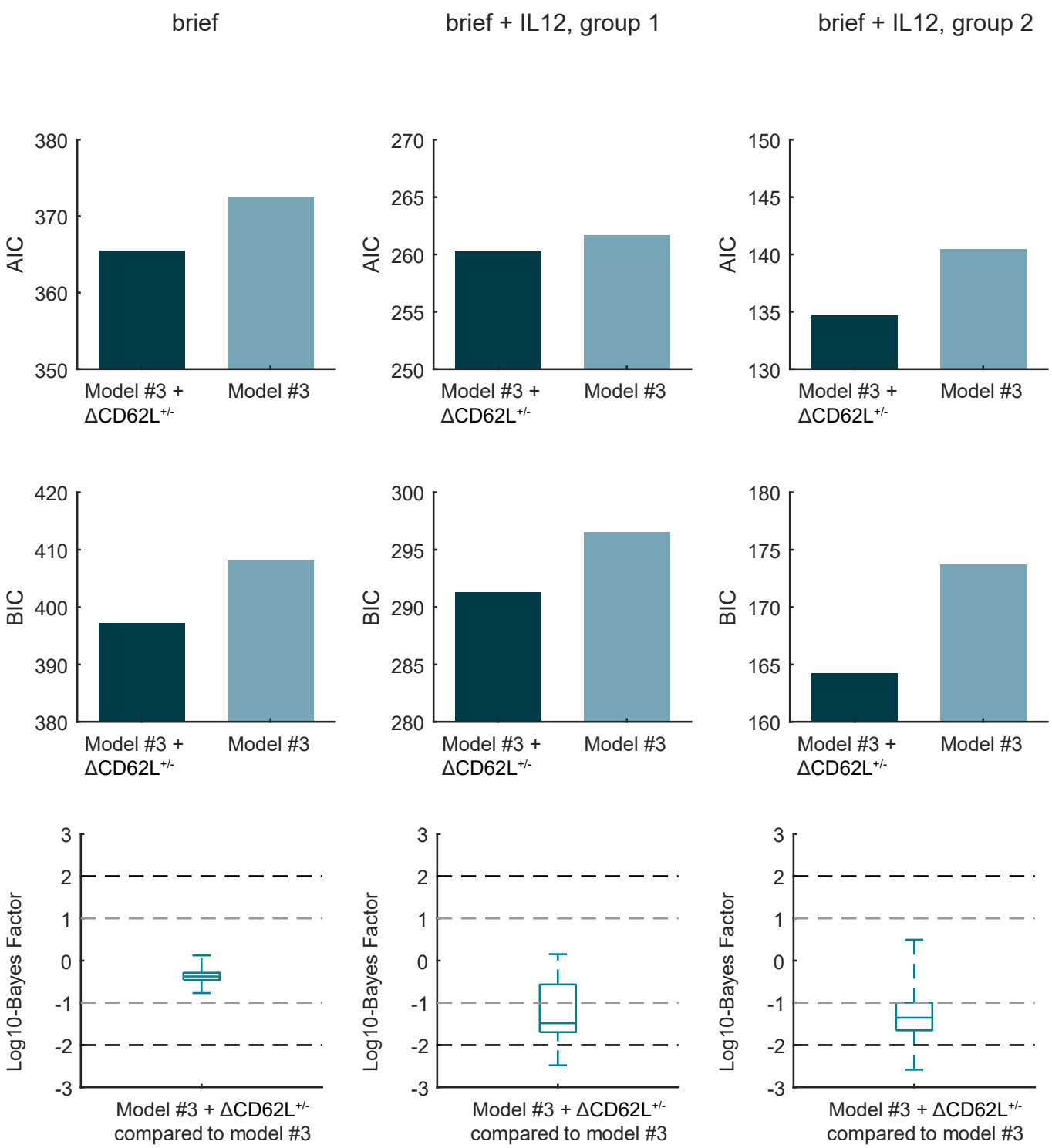

brief
brief + IL12, group 1
brief + IL12, group 2
AIC
BIC
Log10-Bayes Factor
Model #3 + ΔCD62L+/-
Model #3
Model #3 + ΔCD62L+/- compared to model #3

Fig. S28