

DIFFpop Vignettes

Jeremy Ferlic, Jiantao Shi, Thomas O. McDonald, and Franziska Michor

Vignette 1: Branching Process

Background

To begin our example, let us consider a subtree of the hematopoietic system shown in Figure 1.

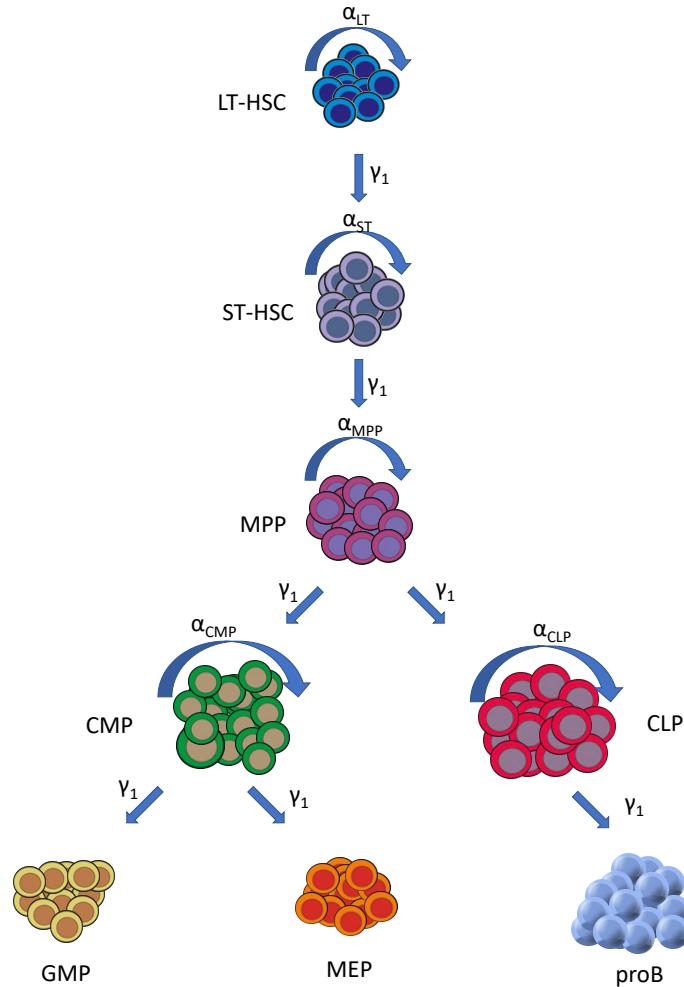


Figure 1: Hematopoietic system modeled using GrowingPops. Abbreviations: long-term hematopoietic stem cell (LT-HSC), short-term hematopoietic stem cell (ST-HSC), multi-potent progenitor (MPP), common myeloid progenitor (CMP), common lymphoid progenitor (CLP), granulocyte-macrophage progenitor (GMP), megakaryocyte-erythroid progenitor (MEP). Events between populations consist of mitosis (α) and mitosis-independent differentiation (γ_1).

This subsystem has been studied in mice by researchers who have developed a mouse model that introduces a fluorescent tag into certain cell population (Busch, 2015). Once activated, this tag, integrated into the genome of the cell, will continue to be present in the progeny of the initially labelled cell. The uptake and washout of the label can then be observed as it descends through the differentiation hierarchy. From this information and assuming that the system has reached a steady state with no population fluxes, Busch et al. used an ordinary differential equation model to estimate the sizes of the compartments and rates of transitions between compartments. Let us use those estimates as a starting point and see how perturbations in various parameters affect the system. In this model, we will only be using a subset of the event types available in DIFFpop. As an example, the events for the short-term hematopoietic stem cells (ST) are shown below in Figure 2.

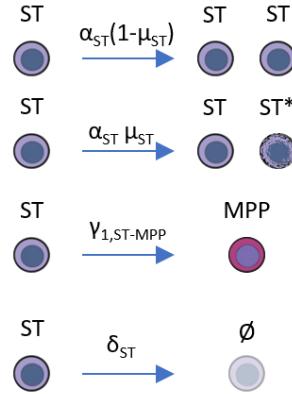


Figure 2: The cellular events for the following simulations include mitosis with and without mutation, mitosis-independent differentiation, and cell death, each occurring according to the rates specified by lowercase Greek letters. Abbreviation: Short-term hematopoietic stem cell (ST).

Using DIFFpop in R

In R, we first specify the populations of the tree using the appropriate DIFFpop functions. To do this, we determine which of the three basic DIFFpop classes is appropriate, give the population a name, initial population size, and initial population barcoding efficiency, what proportion of cells are given a unique barcode during simulation initialization. GrowingPop should be used for systems with homogenous populations that may grow or decline. FixedPop should be used for homogenous populations that maintain constant sizes. DiffTriangle should be used for constant-sized populations with discrete levels of maturation. Here we initially uniquely barcode only LT-HSC cells. Also note that population sizes are given relative to the LT-HSC population size, allowing us to effectively scale our simulations by changing the number of LT-HSCs. The R code to perform the necessary population specification follows:

```

library(diffpop)

nLT = 500

tree1 = DiffTree()
GrowingPop(tree1, "LT", nLT, 1.0)
GrowingPop(tree1, "ST", 2.9*nLT, 0.0)
GrowingPop(tree1, "MPP", 9*nLT, 0.0)

GrowingPop(tree1, "CLP", 13*nLT, 0.0)
GrowingPop(tree1, "CMP", 39*nLT, 0.0)
  
```

```

GrowingPop(tree1, "GMP", as.integer(0.24*39*nLT), 0.0)
GrowingPop(tree1, "MEP", as.integer(0.39*39*nLT), 0.0)

GrowingPop(tree1, "proB", as.integer(108*13*nLT), 0.0)

```

We can then use the addEdge function to specify links between the populations. We use the point estimates from (Busch, 2015), who parameterize a net proliferation rate, which is the death rate subtracted from the self-renewal/mitosis rate. Here, we set this value as the mitotic rate and take the cell death parameter for each population to be zero; however, we could add any positive value to both the alpha (mitosis) and delta (death) event rate to remain within the same parameterization. The R code used to specify these transitions is shown below:

```

addEdge(tree1, "LT", "LT", "alpha", 0.009)
addEdge(tree1, "ST", "ST", "alpha", 0.042)
addEdge(tree1, "MPP", "MPP", "alpha", 4)
addEdge(tree1, "CLP", "CLP", "alpha", 3.00)
addEdge(tree1, "CMP", "CMP", "alpha", 4)

addEdge(tree1, "LT", "ST", "gamma1", 0.009)
addEdge(tree1, "ST", "MPP", "gamma1", 0.045)
addEdge(tree1, "MPP", "CLP", "gamma1", 0.022)
addEdge(tree1, "MPP", "CMP", "gamma1", 3.992)
addEdge(tree1, "CLP", "proB", "gamma1", 2.000)
addEdge(tree1, "CMP", "GMP", "gamma1", 2)
addEdge(tree1, "CMP", "MEP", "gamma1", 3)

addEdge(tree1, "CLP", "CLP", "delta", 1.015)
addEdge(tree1, "GMP", "GMP", "delta", 2*39/(0.24*39))
addEdge(tree1, "MEP", "MEP", "delta", 3*39/(0.39*39))
addEdge(tree1, "proB", "proB", "delta", 2*13/(108*13))

```

To initiate a simulation of the specified tree, the last steps are to first specify a population as the root of the tree, the population that is furthest upstream, and then start the simulation using the simulateTree function with accompanying simulation parameters. Note, for input and output directories, if an absolute path is not specified, DIFFpop will build the specified directory structure from the R working directory. Since we are simulating using GrowingPops, we will set the *fixed* parameter to FALSE.

```

setRoot(tree1, "LT")
simulateTree(tree = tree1,
            fixed = FALSE,
            time = 700,
            indir = "example/",
            outdir = "example/")

```

Results

One output for the simulation is the size of each population after each unit of simulation time. Let us investigate how deviations in one parameter from steady-state parameter set influences the system in terms of population sizes. We varied the self-renewal (mitotic) rate for the LT-HSC population, (α_{LT}). Changing this rate at the top of our differentiation hierarchy has the potential to affect all downstream populations. For each realized value of α_{LT} we have performed 100 simulations and plotted individual simulation trajectories as the fainter trajectory cloud, as well as the mean trajectory as a bolded trajectory for each population over the various α_{LT} . As expected, the steady state α_{LT} value of 0.009 has produced stable population sizes. Increasing this parameter, we observe an increase in the LT population size over time, with the analogous decline in population size for decreased parameter values. We also observe this same trend carry throughout the populations of the hierarchy. Although the differentiation rate to the downstream ST-HSC population remained the same, an increase in LT population size results in an increase net number of cells progressing through the trees. The trajectories are shown in Figure 3.

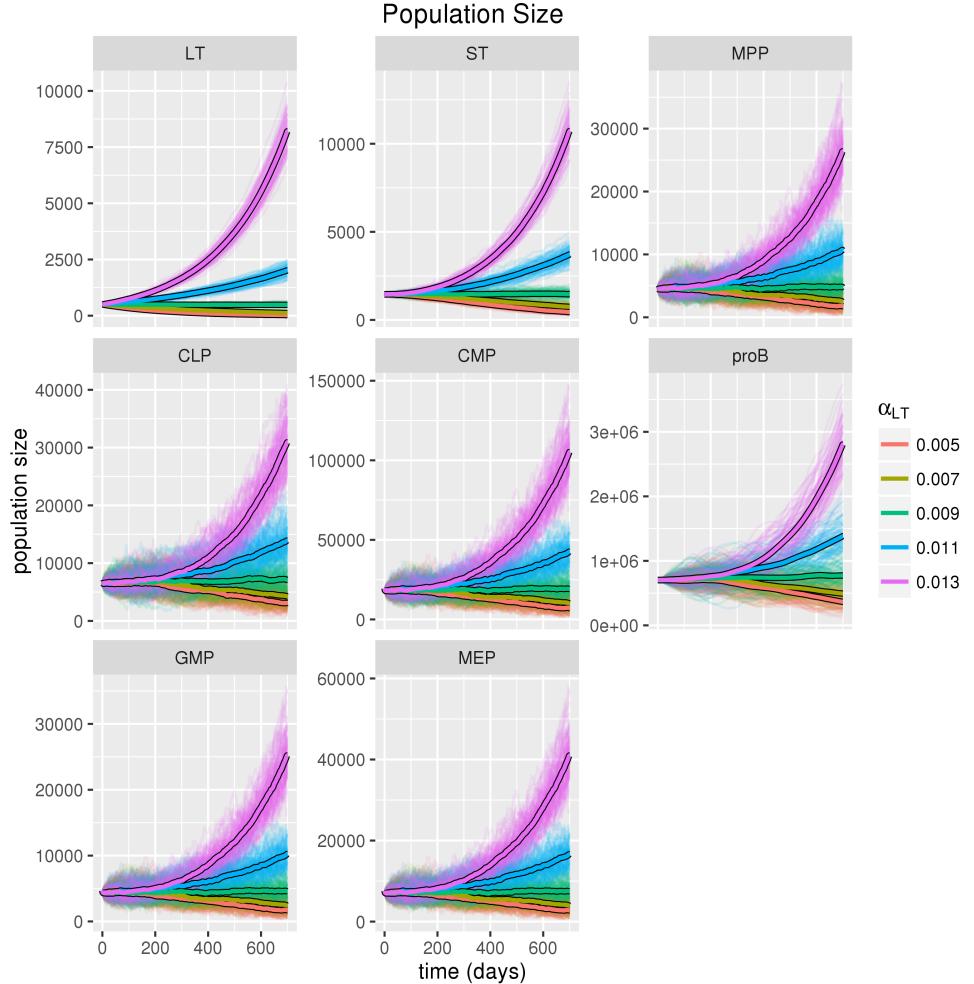


Figure 3: Trajectories of population size over time are shown for varying α_{LT} rates: red ($\alpha_{LT} = 0.005$), orange ($\alpha_{LT} = 0.007$), green ($\alpha_{LT} = 0.009$), blue ($\alpha_{LT} = 0.011$), purple ($\alpha_{LT} = 0.013$). 100 simulation trajectories for each α_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

In addition to varying the α_{LT} rate, we also varied the differentiation rate from the LT-HSC population to the ST-HSC population, γ_{LT} . For γ_{LT} rates that exceed the stable γ_{LT} rate of 0.009, we see a decline in the LT-HSC population, as self-renewal is not able to balance the decline in the population due to differentiation downstream. A different effect occurs in the downstream populations for these higher γ_{LT} rates, where initially receiving additional cells from the LT-HSC population due to the high differentiation rate causes an increase in population sizes relative to the stable state. As the LT-HSC population declines in size, however, all downstream populations also begin to decline, as they do not receive sufficient input from the LT-HSC population to offset their own differentiation downstream. For γ_{LT} rates lower than the stable γ_{LT} rate of 0.009, we notice the opposite effect. These trajectories are shown in Figure 4.

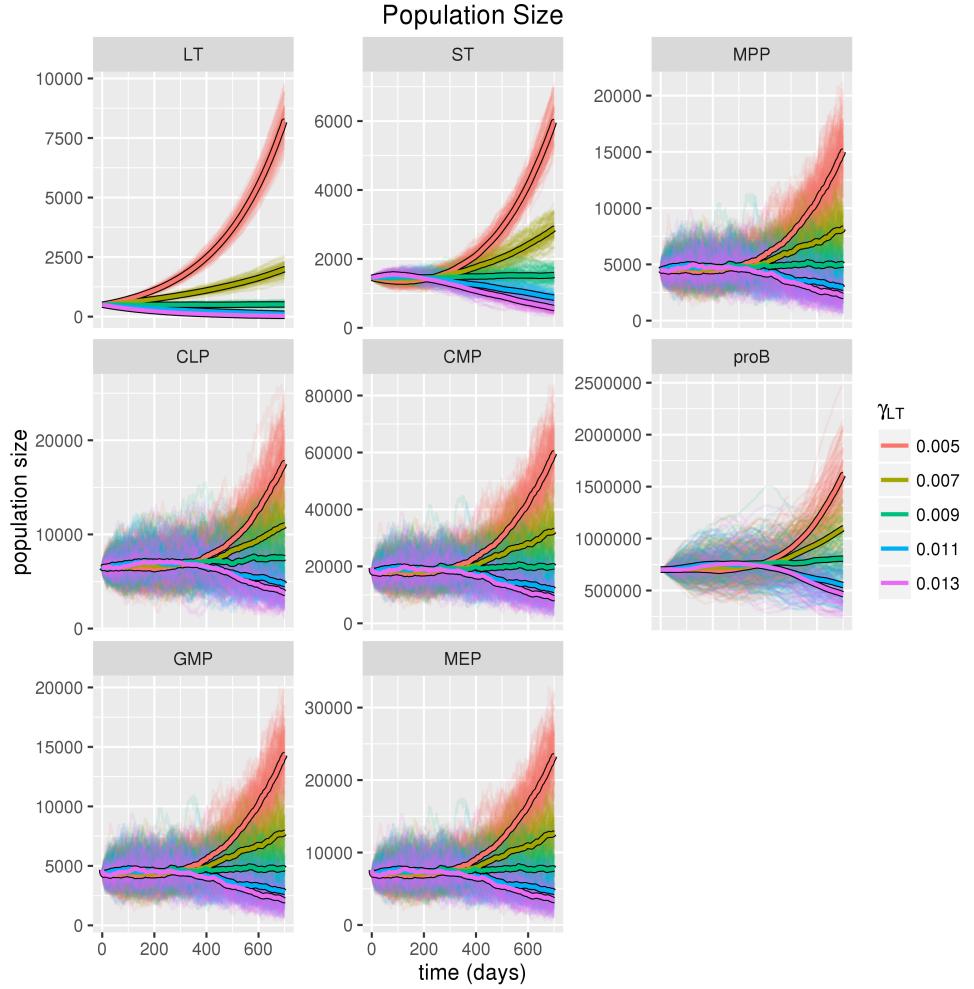


Figure 4: Trajectories of population size over time are shown for varying γ_{LT} rates: red ($\gamma_{LT} = 0.005$), orange ($\gamma_{LT} = 0.007$), green ($\gamma_{LT} = 0.009$), blue ($\gamma_{LT} = 0.011$), purple ($\gamma_{LT} = 0.013$). 100 simulation trajectories for each γ_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

DIFFpop also tracks the fraction of cells in a population that contain a barcode or label. Let us investigate the dynamics of label uptake for various α_{LT} and γ_{LT} rates if we initially label only the LT-HSC population. In general, we see a quicker uptake in label with increasing α_{LT} rate throughout the downstream populations. The reasoning is similar to that in the population size results, whereby as the LT-HSC population grows due to an increased mitotic rate relative to the differentiation rate, a net increase in the number of labelled cells moves through the differentiation hierarchy. Similar to the results shown for population sizes, we see that higher γ_{LT} rates result in an initially higher uptake of label throughout the hierarchy, however; this rate is not sustainable due to the overall decline of the LT-HSC population and label uptake declines relative to the lower γ_{LT} rates. The label fraction plots over various α_{LT} and γ_{LT} rates are show in Figures 5 and 6 respectively.

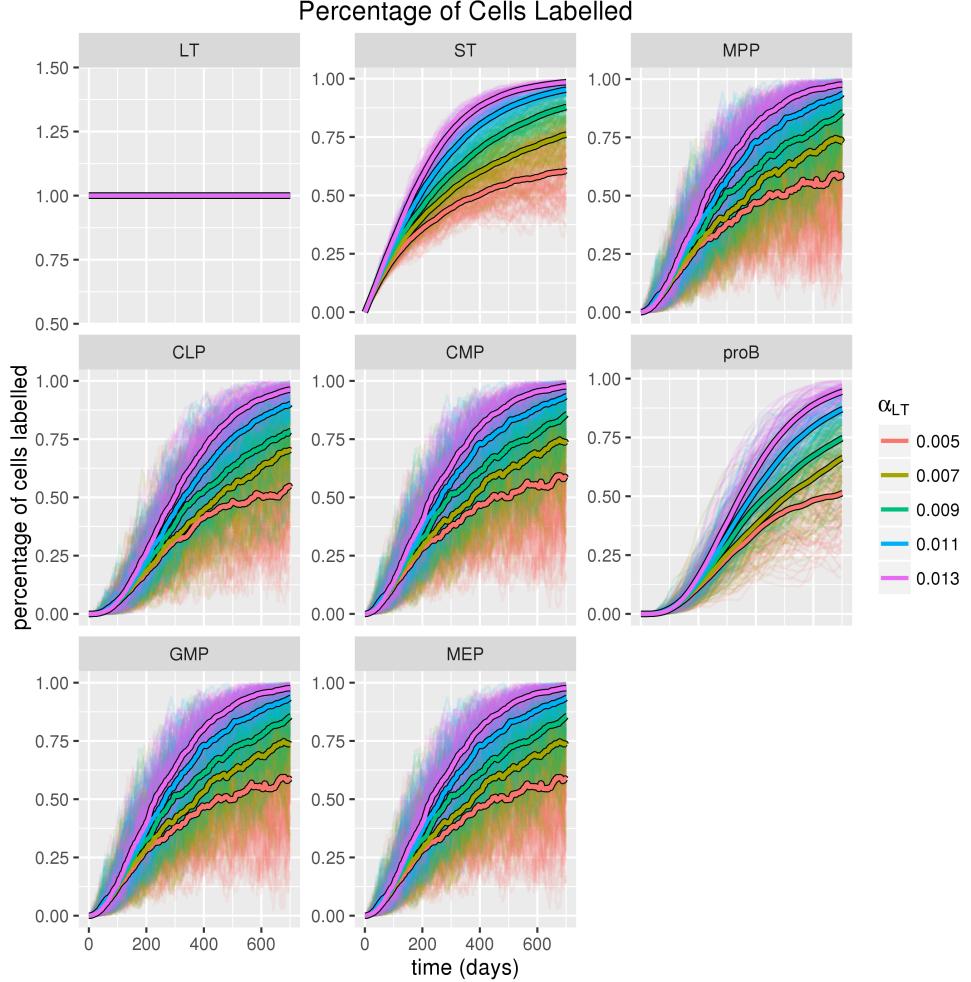


Figure 5: Trajectories of the fraction of cells in each population that express a label over time are shown for varying α_{LT} rates: red ($\alpha_{LT} = 0.005$), orange ($\alpha_{LT} = 0.007$), green ($\alpha_{LT} = 0.009$), blue ($\alpha_{LT} = 0.011$), purple ($\alpha_{LT} = 0.013$). 100 simulation trajectories for each α_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

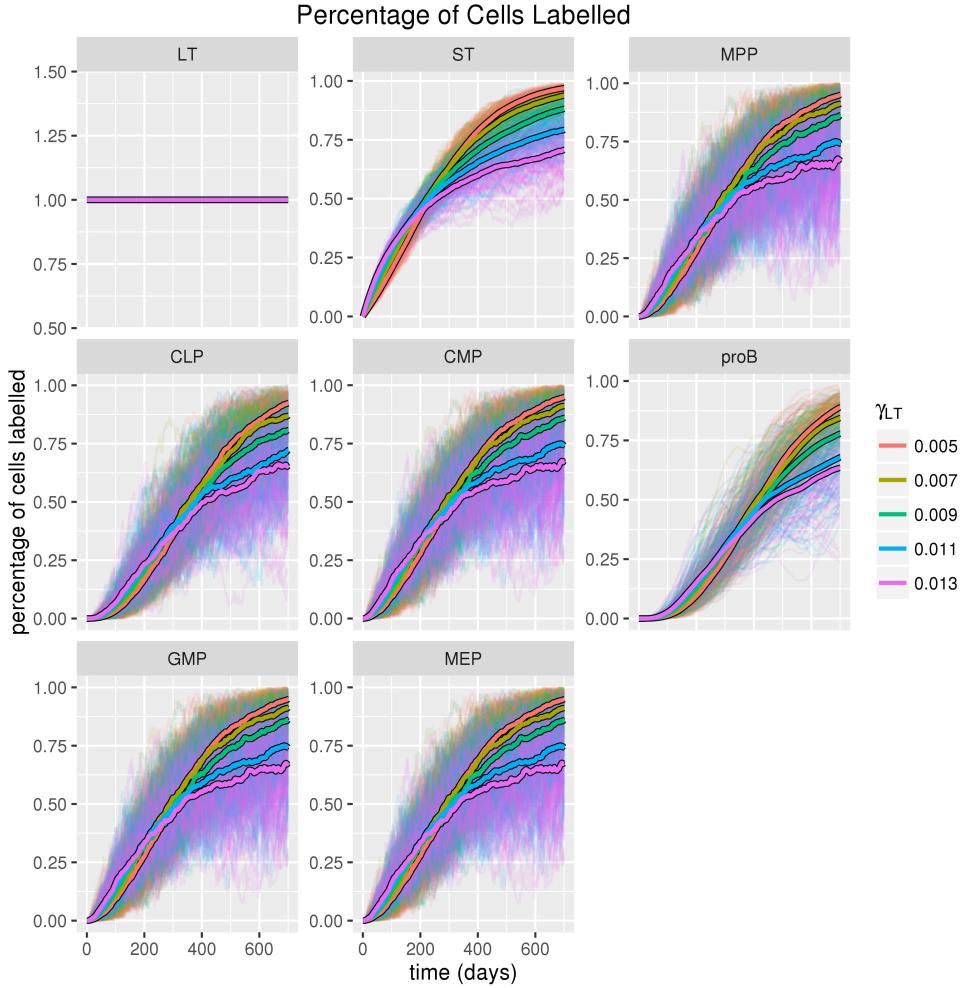


Figure 6: Trajectories of the fraction of cells in each population that express a label over time are shown for varying γ_{LT} rates: red ($\gamma_{LT} = 0.005$), orange ($\gamma_{LT} = 0.007$), green ($\gamma_{LT} = 0.009$), blue ($\gamma_{LT} = 0.011$), purple ($\gamma_{LT} = 0.013$). 100 simulation trajectories for each γ_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

We next investigated changes in diversity of the cell populations over time. To this end, we used Shannon's Equitability for various α_{LT} and γ_{LT} rates. Shannon's Equitability is based on Shannon's Diversity Index, which is defined as $SDI = \sum_j p_j \log p_j$, where p_j is the proportion of cells belonging to clone j , where here, a clone is defined by a certain barcode. Then, Shannon's Equitability is Shannon's Diversity Index divided by its maximum, scaling the range of the equitability to $[0, 1]$. Here, we have once again barcoded all LT-HSC cells uniquely to start and not barcoded any downstream populations. We see similar patterns here as we do in label uptake, with the addition of the effect of declining diversity in the $LT - HSC$ population. Note that this decline is not due to selection, as the introduction of a barcode has no fitness effects on the cells. The Shannon's Equitability over various α_{LT} and γ_{LT} rates are shown in Figures 7 and 8 respectively.

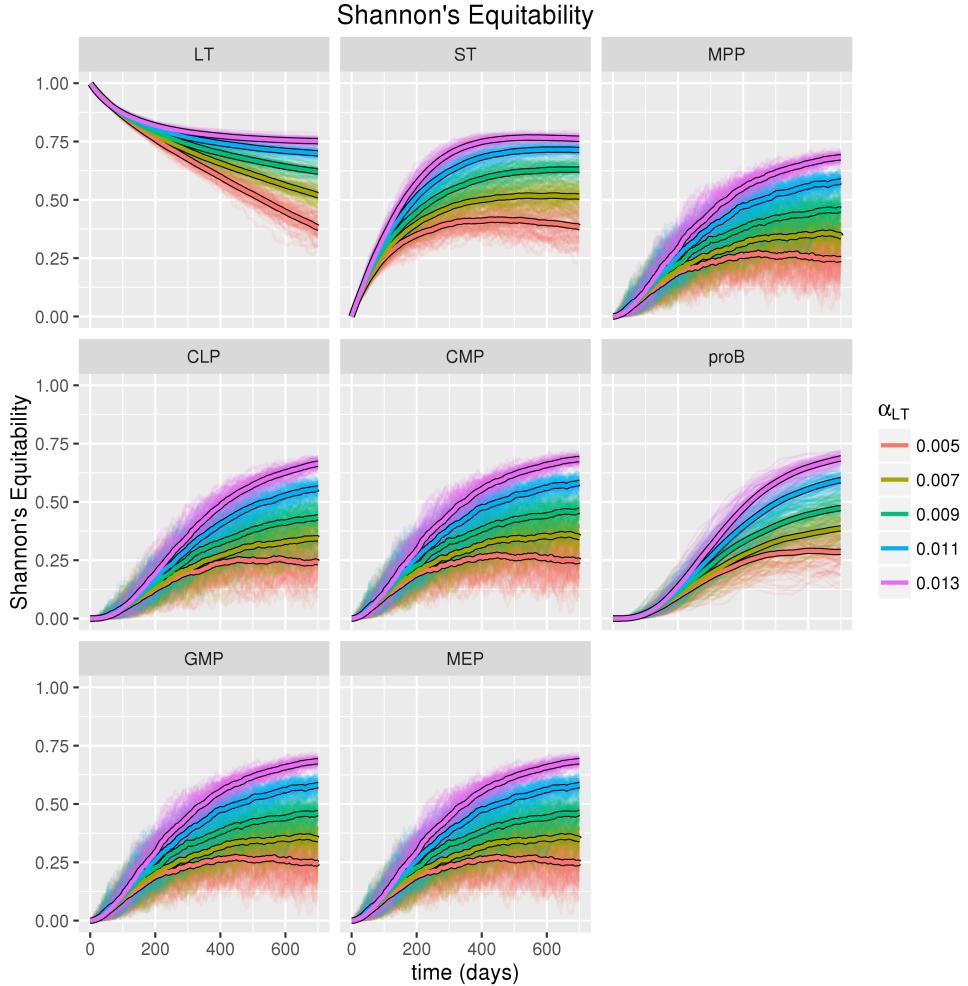


Figure 7: Trajectories of Shannon's Equitability in each population that express a label over time are shown for varying α_{LT} rates: red ($\alpha_{LT} = 0.005$), orange ($\alpha_{LT} = 0.007$), green ($\alpha_{LT} = 0.009$), blue ($\alpha_{LT} = 0.011$), purple ($\alpha_{LT} = 0.013$). 100 simulation trajectories for each α_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

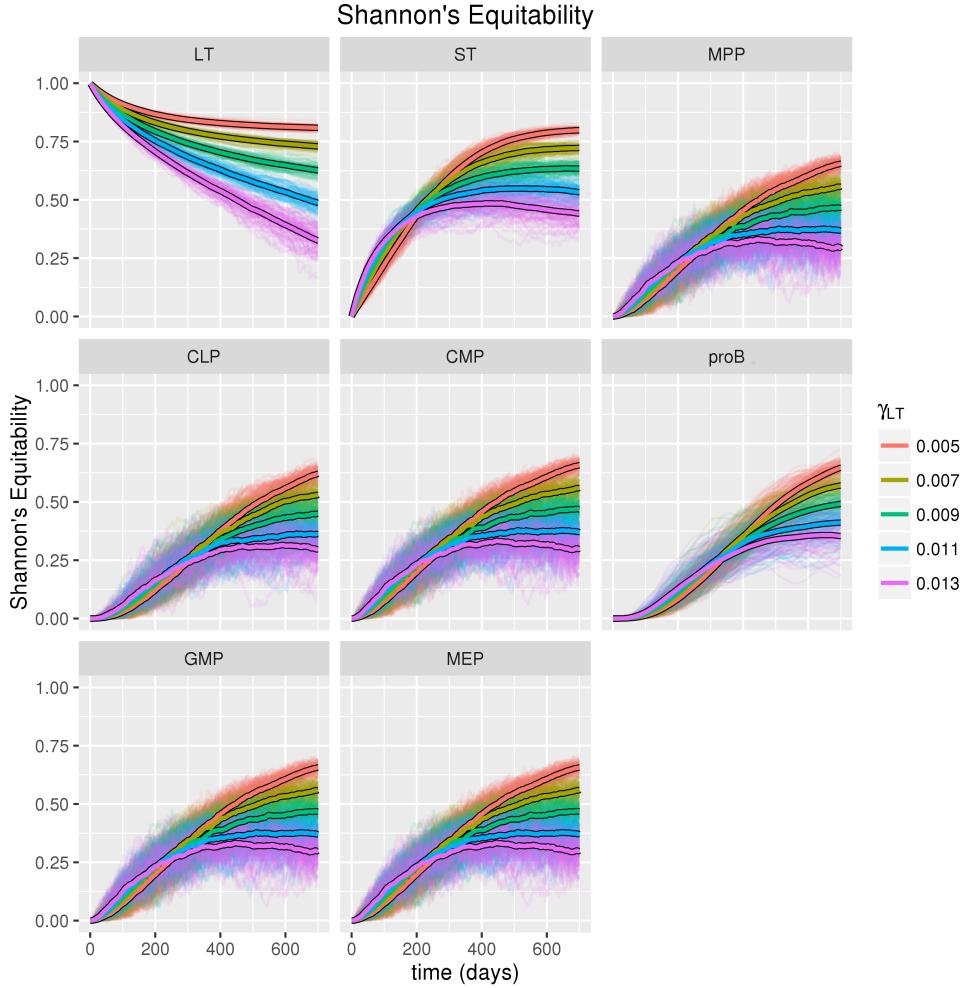


Figure 8: Trajectories of Shannon’s Equitability in each population that express a label over time are shown for varying γ_{LT} rates: red ($\gamma_{LT} = 0.005$), orange ($\gamma_{LT} = 0.007$), green ($\gamma_{LT} = 0.009$), blue ($\gamma_{LT} = 0.011$), purple ($\gamma_{LT} = 0.013$). 100 simulation trajectories for each γ_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

DIFFpop is also capable of simulating how changes in fitness introduced by mutation can influence the dynamics of the system. Towards understanding clonal dynamics within a population, the user can specify how often to output a full census of the hierarchy. Here, we show the clonal dynamics from a single simulation where each population has a mutation probability of 1×10^{-7} per mitotic event and fitness changes are drawn from a double exponential distribution with equal slope parameter 1. In Figure 9, the size of the colored bars correspond to the number of cells from each clone. Note in the following simulation, we have also increased our population size to size of the hematopoietic system in adult mouse by setting $nLT = 17000$. We can introduce these changes by adding the following lines to our tree specification before calling the simulate function with *census* argument equal to 1.

```
addEdge(tree1, "LT", "LT", "mu", 1e-7)
addEdge(tree1, "ST", "ST", "mu", 1e-7)
addEdge(tree1, "MPP", "MPP", "mu", 1e-7)
addEdge(tree1, "CLP", "CLP", "mu", 1e-7)
addEdge(tree1, "CMP", "CMP", "mu", 1e-7)

setFitnessDistribution(tree = tree1,
                      distribution = "doubleexp",
                      alpha_fitness = 1,
                      beta_fitness = 1,
                      pass_prob = 0,
                      upper_fitness = NA,
                      lower_fitness = 0)
```

Clonal Dynamics over Time

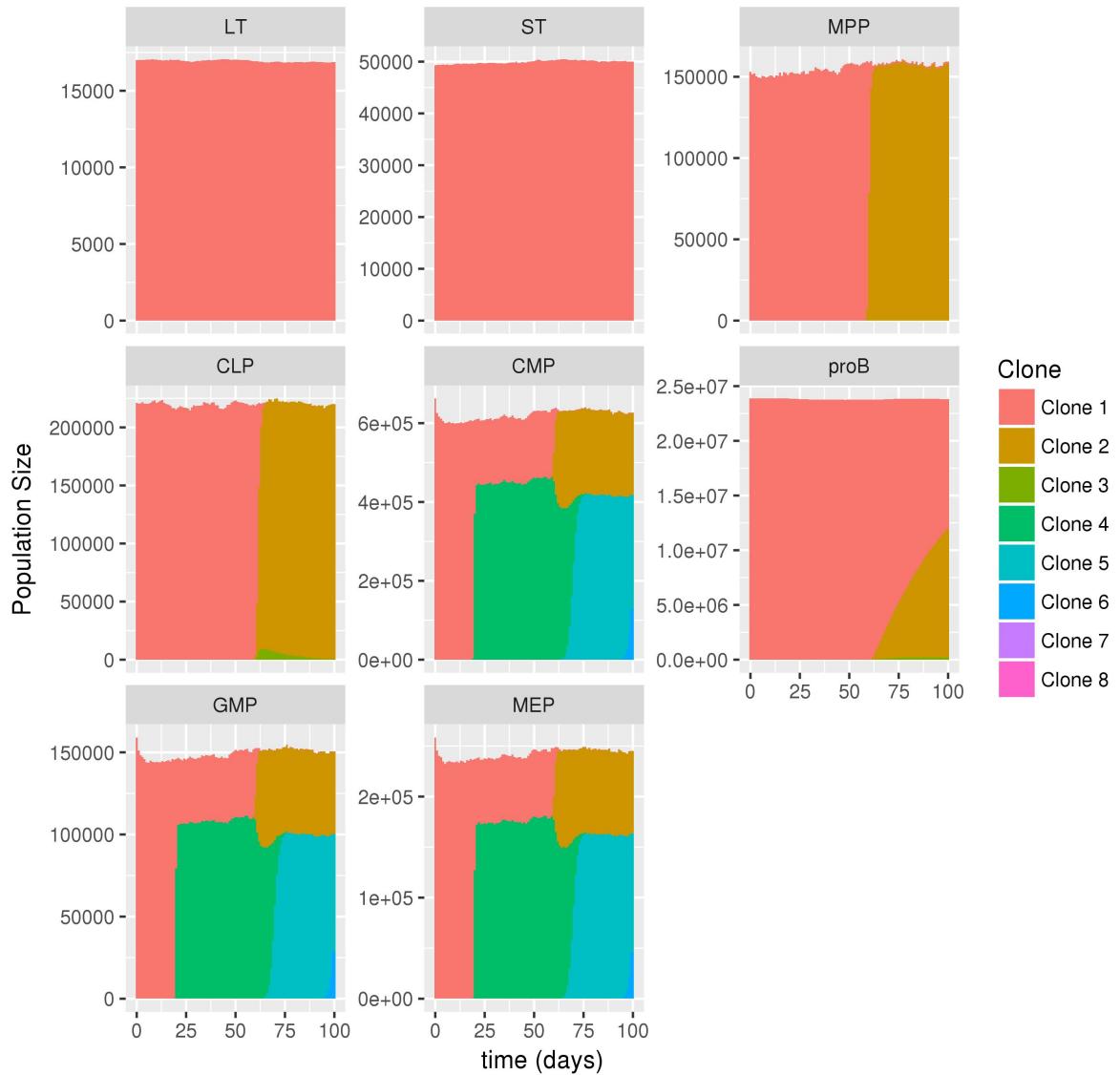


Figure 9: The above plot displays the size of each clone as a different color over time. New clones are initiated by mutation during a mitosis event, with a change in fitness drawn from a double exponential distribution with equal positive and negative slope parameter of 1. Simulations were performed for 100 days at the system size in adult mouse with 17,000 LT-HSCs.

Vignette 2: Multi-type Moran Process

We can also model the system assuming equilibrium has been reached and no population within the hierarchy is experiencing any significant change in size. Towards that end, let us consider adapting our model to use FixedPop and DiffTriangle structures, which guarantee that at every time point, the size of each population remains stable. This model is shown in Figure 10.

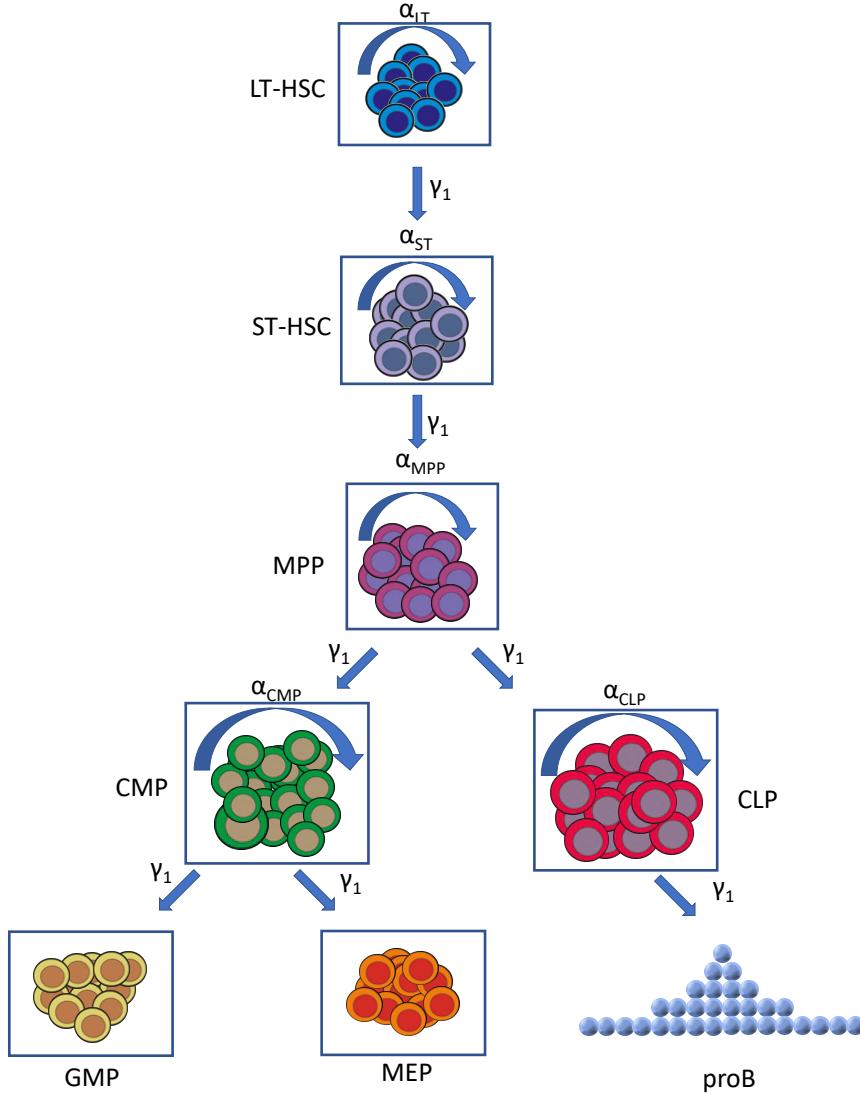


Figure 10: Hematopoietic System model using FixedPops (boxes) and DiffTriangle (pro-B population).

Abbreviations: long-term hematopoietic stem cell (LT-HSC), short-term hematopoietic stem cell (ST-HSC), multi-potent progenitor (MPP), common myeloid progenitor (CMP), common lymphoid progenitor (CLP), granulocyte-macrophage progenitor (GMP), megakaryocyte-erythroid progenitor (MEP).

Events between populations consist of mitosis (α) and mitosis-independent differentiation (γ_1).

In R using the appropriate DIFFpop functions, we first specify the populations of the tree. To do this, we determine which of the three basic DIFFpop class is appropriate, give the population a name, initial

population size, and initial population barcoding efficiency (what proportion of cells are given a unique barcode simulation initialization). Here, we use FixedPops and DiffTriangles. Again, we only uniquely barcode the LT-HSC population.

```
library(diffpop)

nLT = 500

tree2 = DiffTree()
FixedPop(tree2, "LT", nLT, 1.0)
FixedPop(tree2, "ST", 2.9*nLT, 0.0)
FixedPop(tree2, "MPP", 9*nLT, 0.0)

FixedPop(tree2, "CLP", 13*nLT, 0.0)
FixedPop(tree2, "CMP", 39*nLT, 0.0)

FixedPop(tree2, "GMP", as.integer(0.24*39*nLT), 0.0)
FixedPop(tree2, "MEP", as.integer(0.39*39*nLT), 0.0)

DiffTriangle(tree2, "proB", height = 6, first_level = 2*13*nLT)
```

We can then use the addEdge function to specify links between the populations. We will be using the point estimates from (Busch, 2015) as our event rates. As before, a parameterization for net proliferation is used, which is the self-renewal/mitosis rate minus the cell death rate. Here, we set this value as our mitotic rate; however, we could add any positive value to both the alpha and delta (death) event rate to remain within the correct parameterization.

```
addEdge(tree2, "LT", "LT", "alpha", 0.009)
addEdge(tree2, "ST", "ST", "alpha", 0.042)
addEdge(tree2, "MPP", "MPP", "alpha", 4)
addEdge(tree2, "CLP", "CLP", "alpha", 3.00)
addEdge(tree2, "CMP", "CMP", "alpha", 4)

addEdge(tree2, "LT", "ST", "gamma1", 0.009)
addEdge(tree2, "ST", "MPP", "gamma1", 0.045)
addEdge(tree2, "MPP", "CLP", "gamma1", 0.022)
addEdge(tree2, "MPP", "CMP", "gamma1", 3.992)
addEdge(tree2, "CLP", "proB", "gamma1", 2.000)
addEdge(tree2, "CMP", "GMP", "gamma1", 2)
addEdge(tree2, "CMP", "MEP", "gamma1", 3)
```

To initiate a simulation of the specified tree, the last steps are to first specify which population is the root of the tree (the population that is furthest upstream), write our tree input files to a specified location, and then start the simulation using the simulateTree function, this time with the fixed parameter set to TRUE.

```
setRoot(tree2, "LT")
simulateTree(tree = tree2,
            fixed = TRUE,
            time = 700,
            indir = "example/",
            outdir = "example/")
```

Let us begin by looking at the population sizes over time for various α_{LT} and γ_{LT} rates. Because we are using structures that maintain a constant population size, we see no fluctuation in the population sizes over time. These population size over time plots for various α_{LT} and γ_{LT} rates are shown in Figures 11 and 12 respectively.

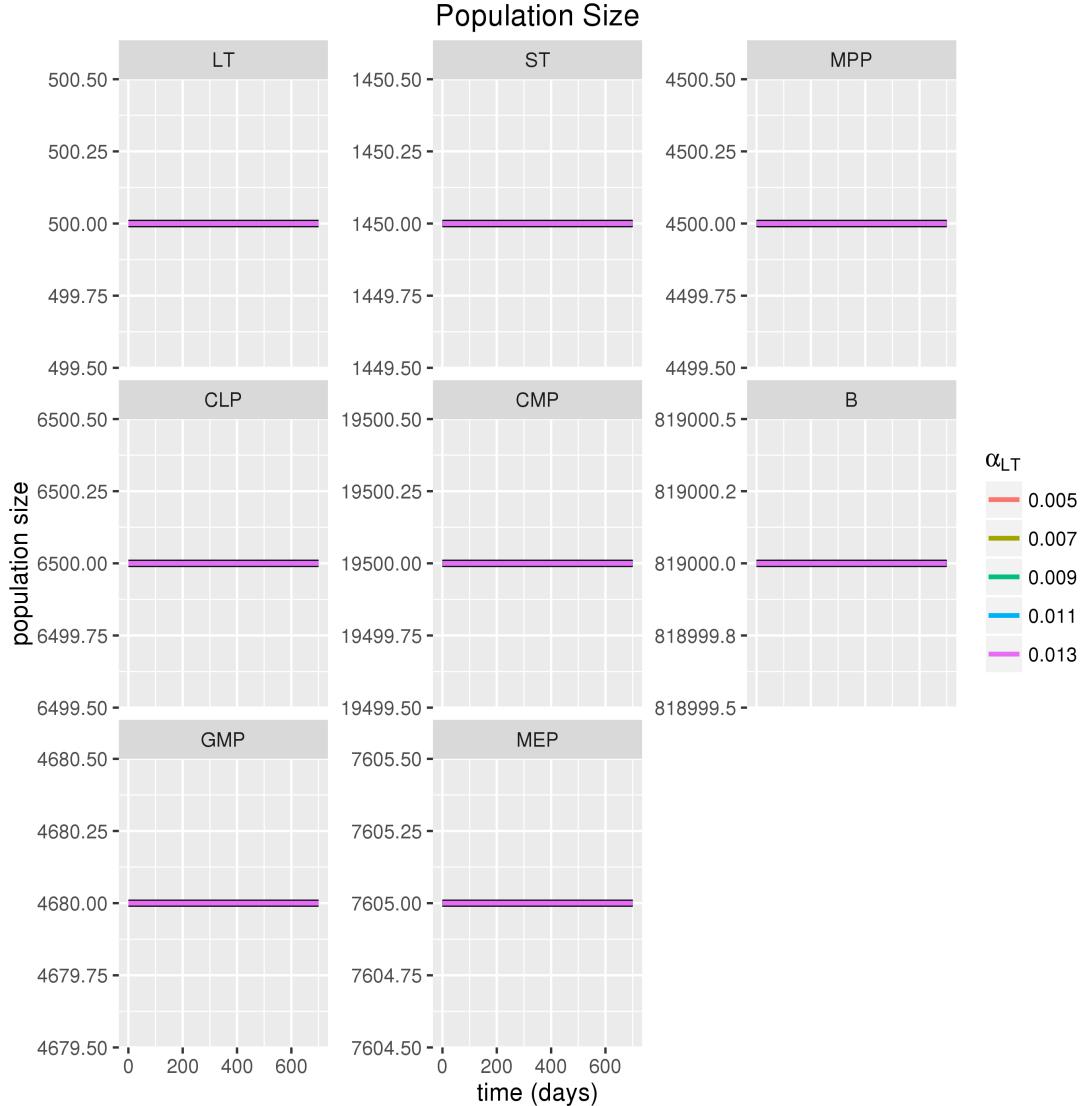


Figure 11: Trajectories of population size over time are shown for varying α_{LT} rates: red ($\alpha_{LT} = 0.005$), orange ($\alpha_{LT} = 0.007$), green ($\alpha_{LT} = 0.009$), blue ($\alpha_{LT} = 0.011$), purple ($\alpha_{LT} = 0.013$). 100 simulation trajectories for each α_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

Note: all trajectories plotted are flat and hence only the last trajectory plotted (purple) is visible.

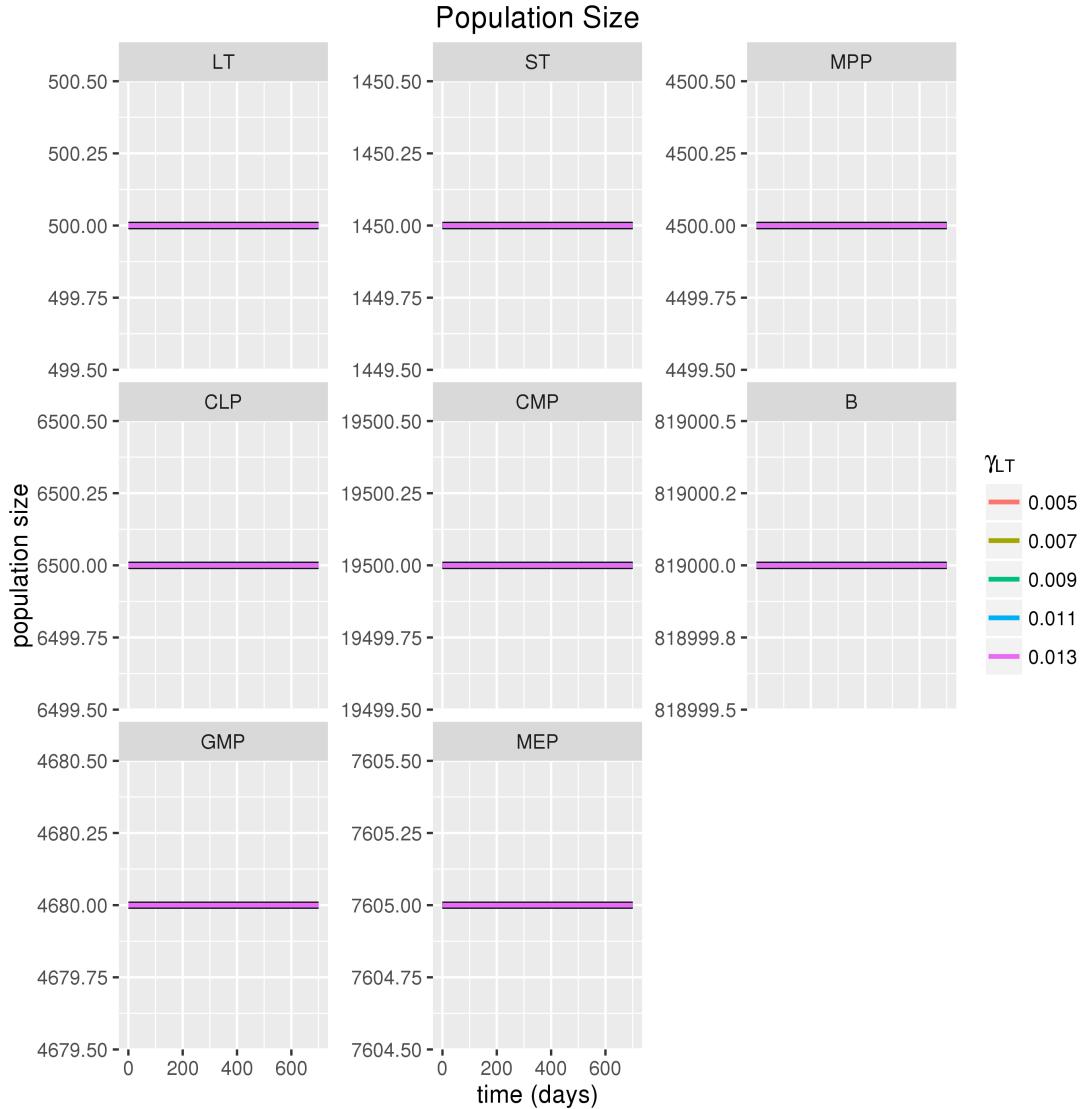


Figure 12: Trajectories of population size over time are shown for varying γ_{LT} rates: red ($\gamma_{LT} = 0.005$), orange ($\gamma_{LT} = 0.007$), green ($\gamma_{LT} = 0.009$), blue ($\gamma_{LT} = 0.011$), purple ($\gamma_{LT} = 0.013$). 100 simulation trajectories for each γ_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

Note: all trajectories plotted are flat and hence only the last trajectory plotted (purple) is visible.

In Figure 13, looking at the fraction of cells that have a barcode for various α_{LT} and γ_{LT} rates, we observe how our simulation procedure for fixed populations differs from our growing populations. Across various α_{LT} rates, we see the same trajectory for the fraction of cells that have a barcode. This is because when we change our α_{LT} rate, the fixed simulation automatically adjusts the net proliferation, adjusting either α_{LT} or δ_{LT} , to maintain a constant population size. Because we are not adjusting the differentiation rate downstream γ_{LT} , we experience the same level of differentiation of barcoded cells for all levels of α_{LT} .

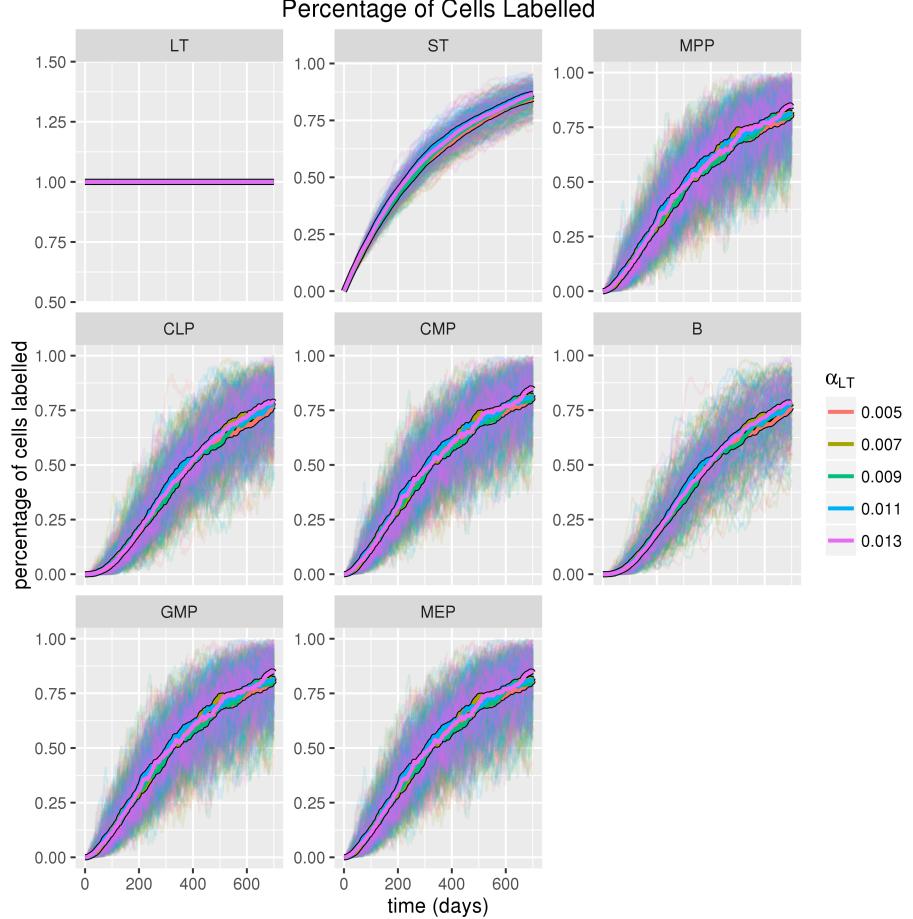


Figure 13: Trajectories of the fraction of cells in each population that express a label over time are shown for varying α_{LT} rates: red ($\alpha_{LT} = 0.005$), orange ($\alpha_{LT} = 0.007$), green ($\alpha_{LT} = 0.009$), blue ($\alpha_{LT} = 0.011$), purple ($\alpha_{LT} = 0.013$). 100 simulation trajectories for each α_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

We also varied the mitosis-independent differentiation rate from LT-HSC to ST-HSC, γ_{LT} . Across various γ_{LT} rates in Figure 14, we observe trajectories that match intuition. The higher the differentiation rate to downstream populations, the more barcoded cells appear in the downstream populations. In this case, because the LT-HSC population size is fixed, we do not observe a decline in the fraction of barcoded cells for higher γ_{LT} rates like we did in the branching process model.

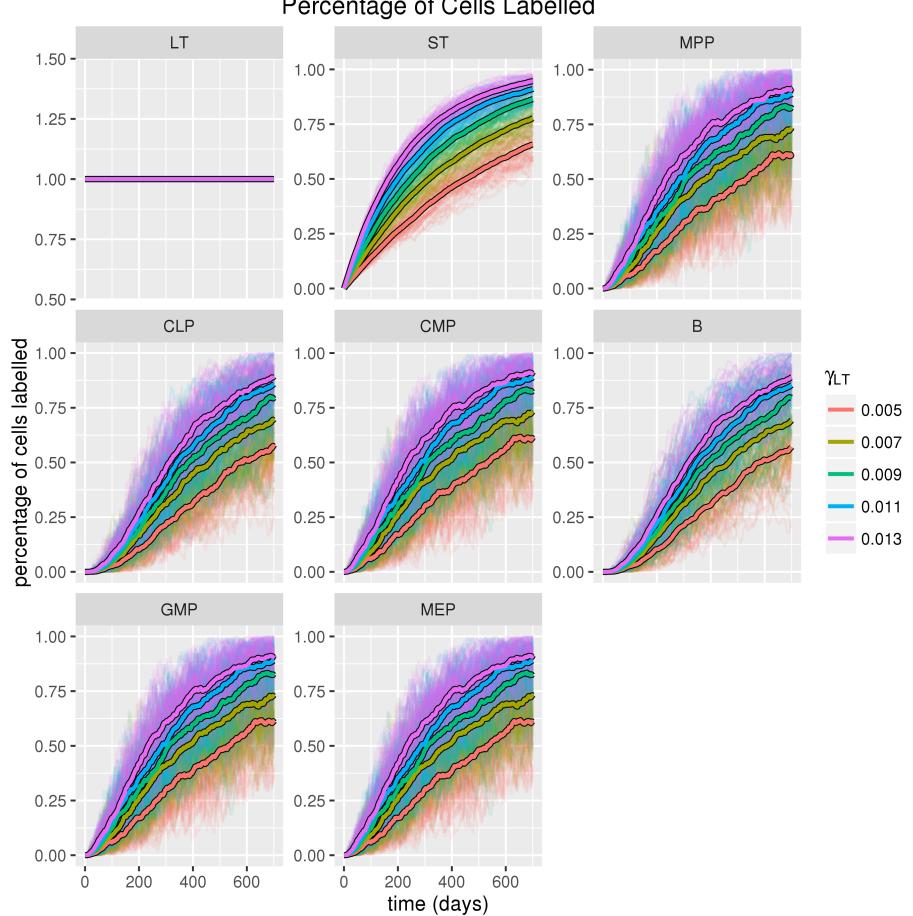


Figure 14: Trajectories of the fraction of cells in each population that express a label over time are shown for varying γ_{LT} rates: red ($\gamma_{LT} = 0.005$), orange ($\gamma_{LT} = 0.007$), green ($\gamma_{LT} = 0.009$), blue ($\gamma_{LT} = 0.011$), purple ($\gamma_{LT} = 0.013$). 100 simulation trajectories for each γ_{LT} rate are plotted as well as a bold mean trajectory. Simulations were run for 700 days, around the average lifespan for a mouse using all other parameters as shown in code excerpts above.

As with the branching process model, we can introduce mutations into our fixed population model and track the clonal dynamics over time. Here in Figure 15, we show the clonal dynamics from a single simulation where each population has a mutation probability of 1×10^{-7} per mitotic event and fitness changes are drawn from a double exponential distribution with equal slope parameter 1, where the size of the colored bars represent the number of cells from each clone. Notice the that total height of each bar remains constant over time, as expected. Once again, we can make these changes to *tree2* before simulating:

```

addEdge(tree2, "LT", "LT", "mu", 1e-7)
addEdge(tree2, "ST", "ST", "mu", 1e-7)
addEdge(tree2, "MPP", "MPP", "mu", 1e-7)
addEdge(tree2, "CLP", "CLP", "mu", 1e-7)
addEdge(tree2, "CMP", "CMP", "mu", 1e-7)

setFitnessDistribution(tree = tree2,
                      distribution = "doubleexp",
                      alpha_fitness = 1,
                      beta_fitness = 1,
                      pass_prob = 0,
                      upper_fitness = NA,
                      lower_fitness = 0)

```

Clonal Dynamics over Time

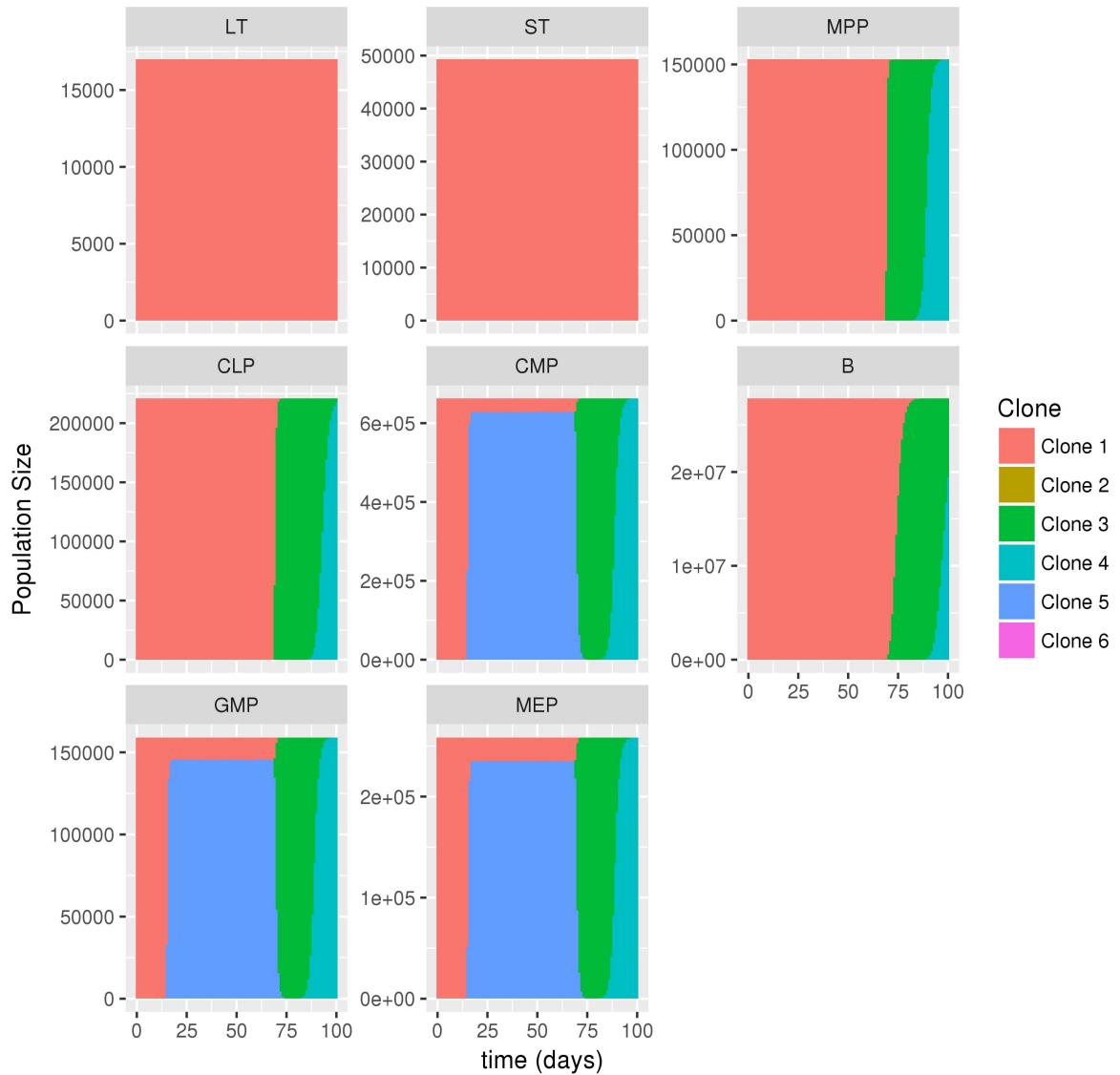


Figure 15: The above plot displays the size of each clone as a different color over time. New clones are initiated by mutation during a mitosis event, with a change in fitness drawn from a double exponential distribution with equal positive and negative slope parameter equal to 1. Simulations were performed for 100 days at the true system size starting with 17,000 LT-HSCs.

As we stated earlier, (Busch, 2015) parameterize a joint net proliferation rate ($\alpha - \delta$) for each population, meaning that if we increase both α and δ by the same amount for a population, we remain within this parameterization. Doing this changes both the amount of mutations that occur and the amount of time it takes for a mutation to fix in the population. Because there are more mitotic divisions in a time unit, the number of mutations increase with increasing α . Also, because there are more mitosis events, there is more opportunity for selection events to favor dominating, high-fitness clones, resulting in less time for a clone to fix in the population. In Figures 16, 17, and 18, we show the clonal dynamics bar plots for three scenarios all of which fit the net proliferation parameter for each population, only we have increased α and δ for each population by 100% (Figure 17) and 1000% (Figure 18).

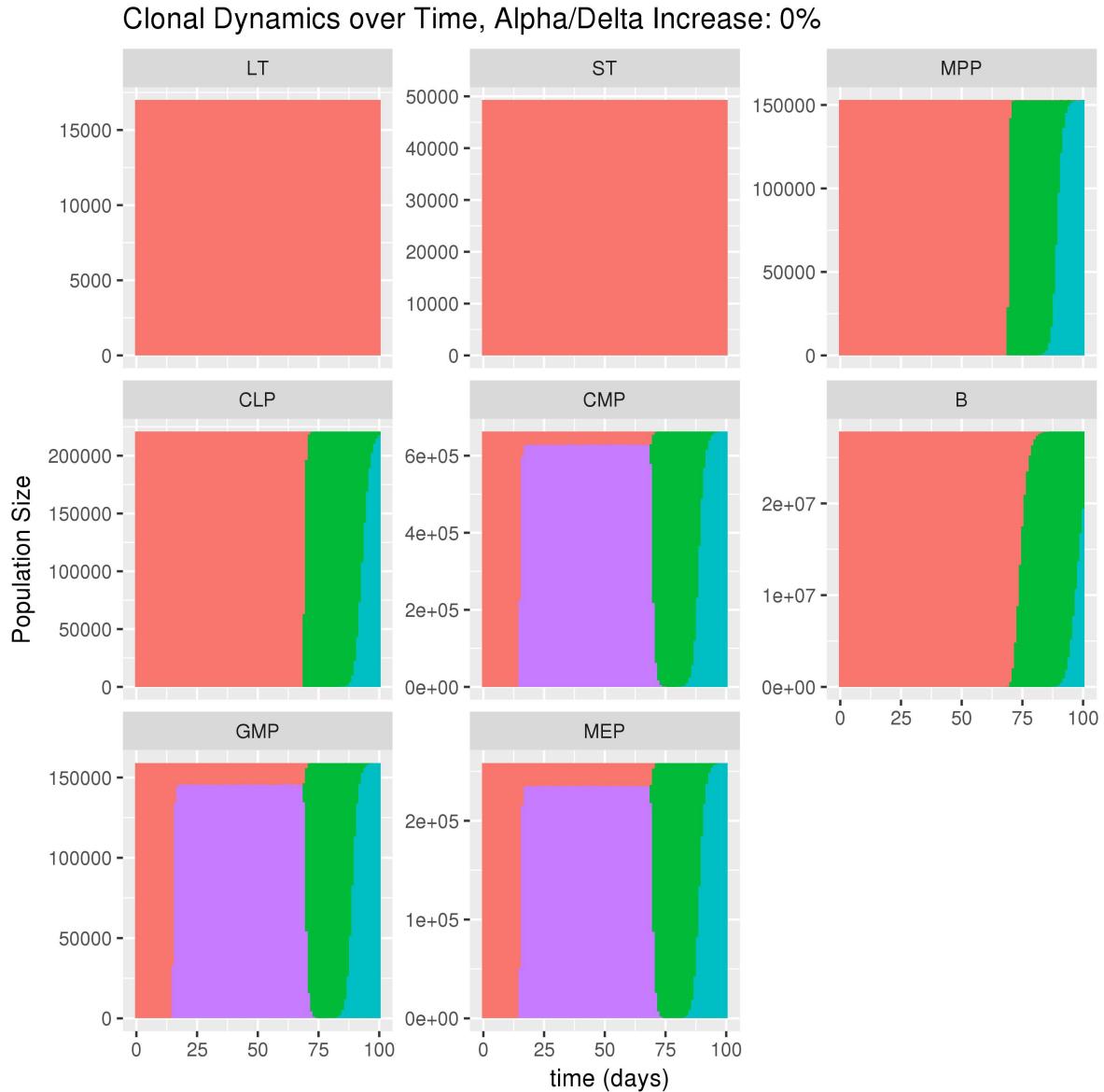


Figure 16: No increase in either α or δ for any population. Clones sizes are plotted using different colors for each clone over time. At this increase level, few clones arise in all populations. Simulations were run at the true system size beginning with 17,000 LT-HSCs for 100 days.

Clonal Dynamics over Time, Alpha/Delta Increase: 100%

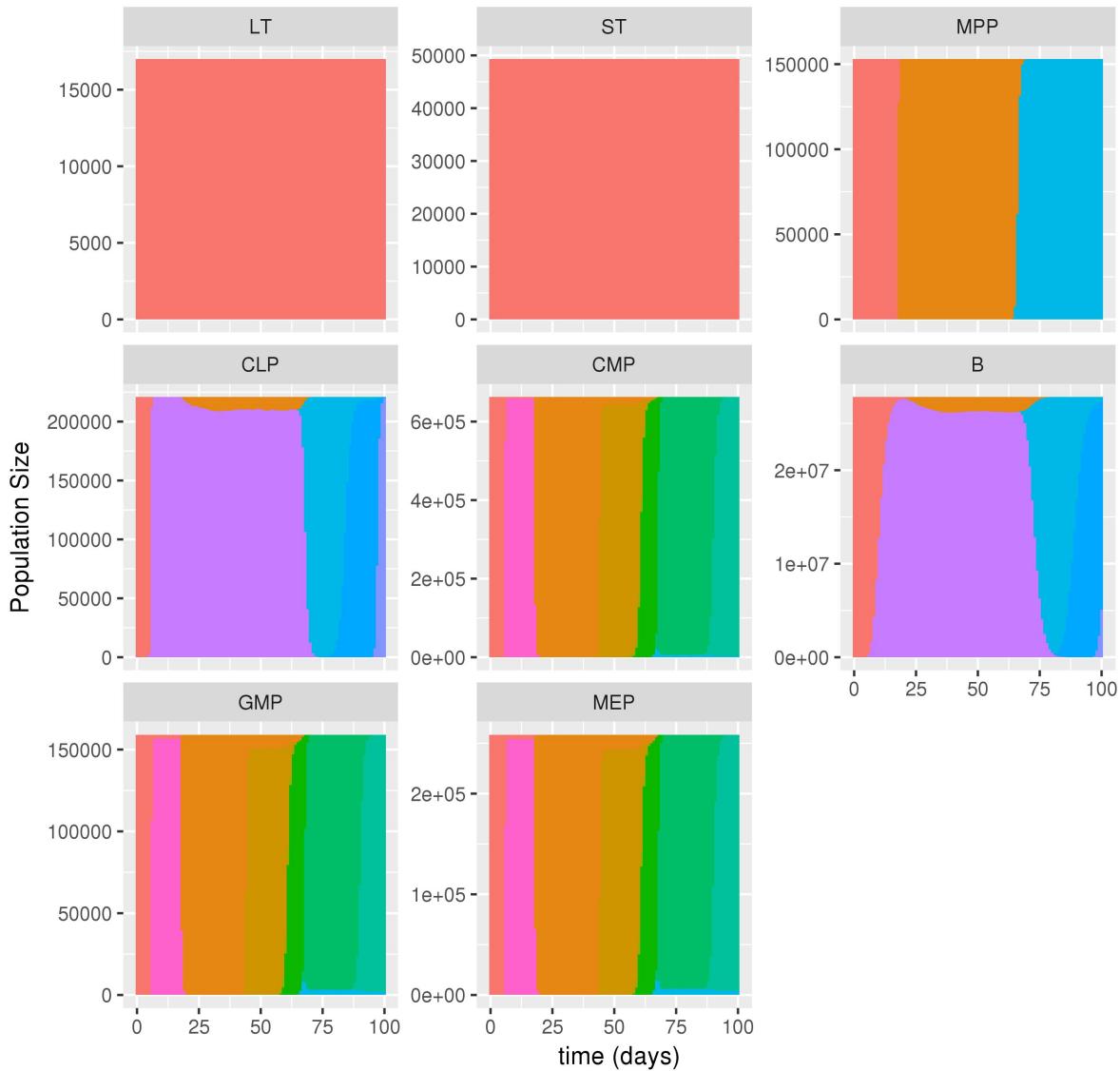


Figure 17: 100% increase in both α and δ for all populations. Clones sizes are plotted using different colors for each clone over time. At this increase level, many clones arise and fix in the quickly proliferating populations (CMP, GMP, MEP). Simulations were run at the true system size beginning with 17,000 LT-HSCs for 100 days.

Clonal Dynamics over Time, Alpha/Delta Increase: 1000%

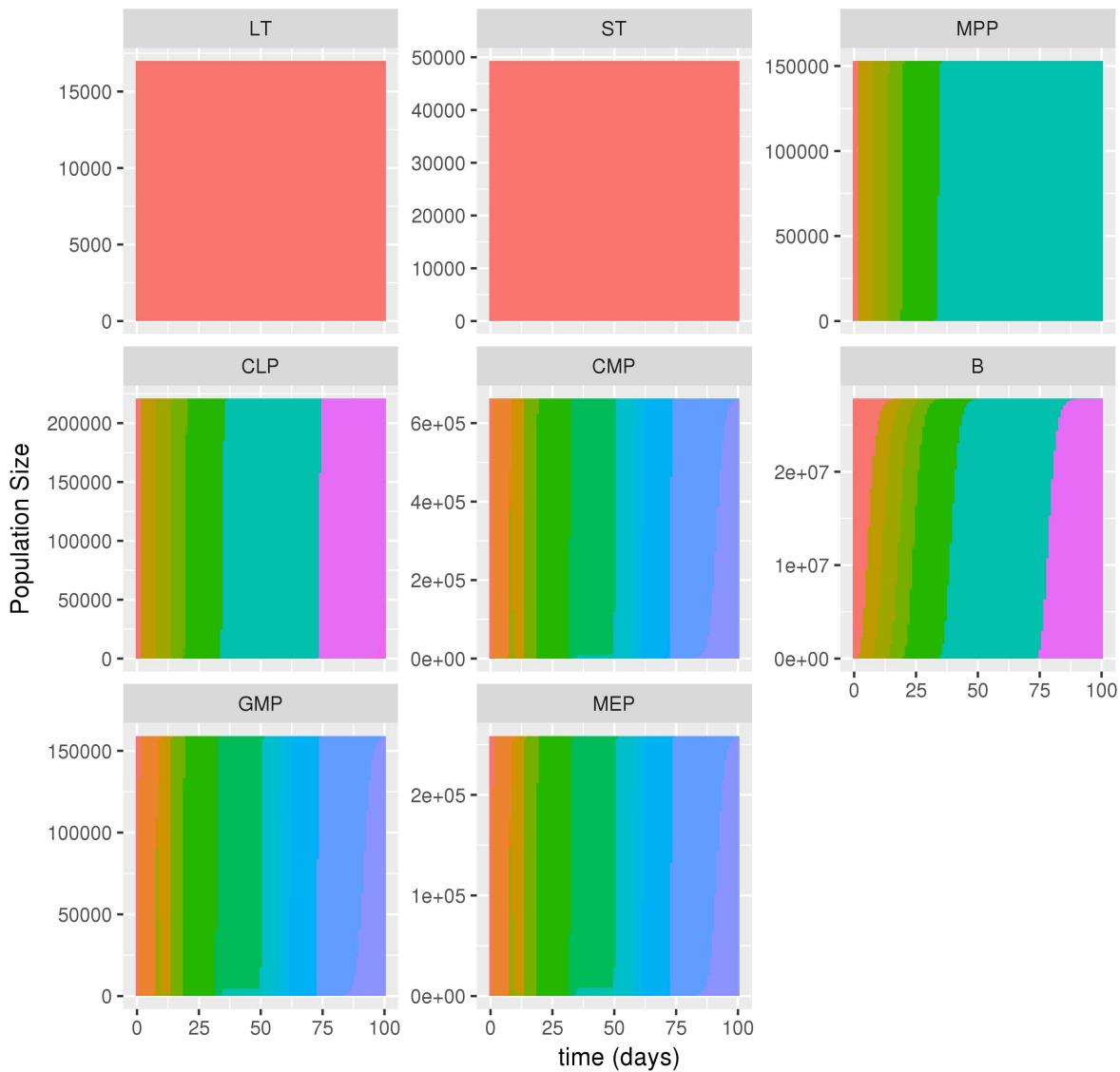


Figure 18: 1000% increase in both α and δ for all populations. Clones sizes are plotted using different colors for each clone over time. At this increase level, many clones arise and fix in the quickly in all downstream populations. Simulations were run at the true system size beginning with 17,000 LT-HSCs for 100 days.

References

Busch, Katrin, et al. “Fundamental properties of unperturbed haematopoiesis from stem cells in vivo.” Nature 518.7540 (2015): 542.