

<sup>1</sup> *Working Notes:* Single-celled bottlenecks, germlines and the  
<sup>2</sup> evolution of complex multi-cellularity

<sup>3</sup>

<sup>4</sup> **Contents**

<sup>5</sup>	<b>1 Data Collection</b>	<sup>1</sup>
<sup>6</sup>	<b>2 Statistical Analyses</b>	<sup>1</sup>
<sup>7</sup>	2.1 MCMCglmm parameters . . . . .	<sup>1</sup>
<sup>8</sup>	2.2 Without Phylogeny . . . . .	<sup>2</sup>
<sup>9</sup>	2.2.1 <b>Model 1:</b> Fission vs Cell Number . . . . .	<sup>2</sup>
<sup>10</sup>	2.2.2 <b>Model 2:</b> Fission vs Cell Types . . . . .	<sup>3</sup>
<sup>11</sup>	2.2.3 <b>Model 3:</b> Germline vs Cell Numbers . . . . .	<sup>3</sup>
<sup>12</sup>	2.2.4 <b>Model 4:</b> Germline vs Cell Types . . . . .	<sup>3</sup>
<sup>13</sup>	2.3 Phylogenetically Informed Models . . . . .	<sup>3</sup>
<sup>14</sup>	2.3.1 <b>Model 5:</b> Fission vs Cell Number . . . . .	<sup>3</sup>
<sup>15</sup>	2.3.2 <b>Model 6:</b> Fission vs Cell Types . . . . .	<sup>3</sup>
<sup>16</sup>	2.3.3 <b>Model 7:</b> Germline vs Cell Number . . . . .	<sup>3</sup>
<sup>17</sup>	2.3.4 <b>Model 8:</b> Germline vs Cell Types . . . . .	<sup>11</sup>
<sup>18</sup>	<b>3 References</b>	<sup>12</sup>

<sup>19</sup> **1 Data Collection**

- <sup>20</sup> Data were collected by...  
<sup>21</sup> Phylogeny was constructed by

<sup>22</sup> **2 Statistical Analyses**

<sup>23</sup> **2.1 MCMCglmm parameters**

- <sup>24</sup> All analyses were conducted in R (R Core Team 2021) using the package MCMCglmm ([MCMCglmm?](#)),  
<sup>25</sup> while documents were produced using [RMarkdown]. All data and code are accessible at [github..](#)  
<sup>26</sup> Model parameters were optimised using the first model described in our results, for which we ran a total of  
<sup>27</sup> 38 MCMCglmm chains of varying lengths (500000 - 10000000 iterations), with varying warm-ups (100000  
<sup>28</sup> - 1000000, and with thinning of either 100 or 1000 fold, see Figure S1. All subsequent models were then

29 fit using the combination of these parameters where the autocorrelation of successive sampled mean and  
 30 variance were minimal:  $8 \times 10^6$  iterations, a warm-up of  $10^6$  iterations and thinning by a factor of 100. In  
 31 all fitted models, the autocorrelation was well below the suggested tolerable maximum of 0.1 (**hadfield?**).  
 32 For each model, 6 chains were run which were visually inspected for chain convergence. Convergence was  
 33 also supported by the Gelman-Rubin (**Gelman-Rubin?**) convergence diagnostic, which approximated 1  
 34 ( $<1.05$ ) in all cases—these are reported in the summary of each model below.

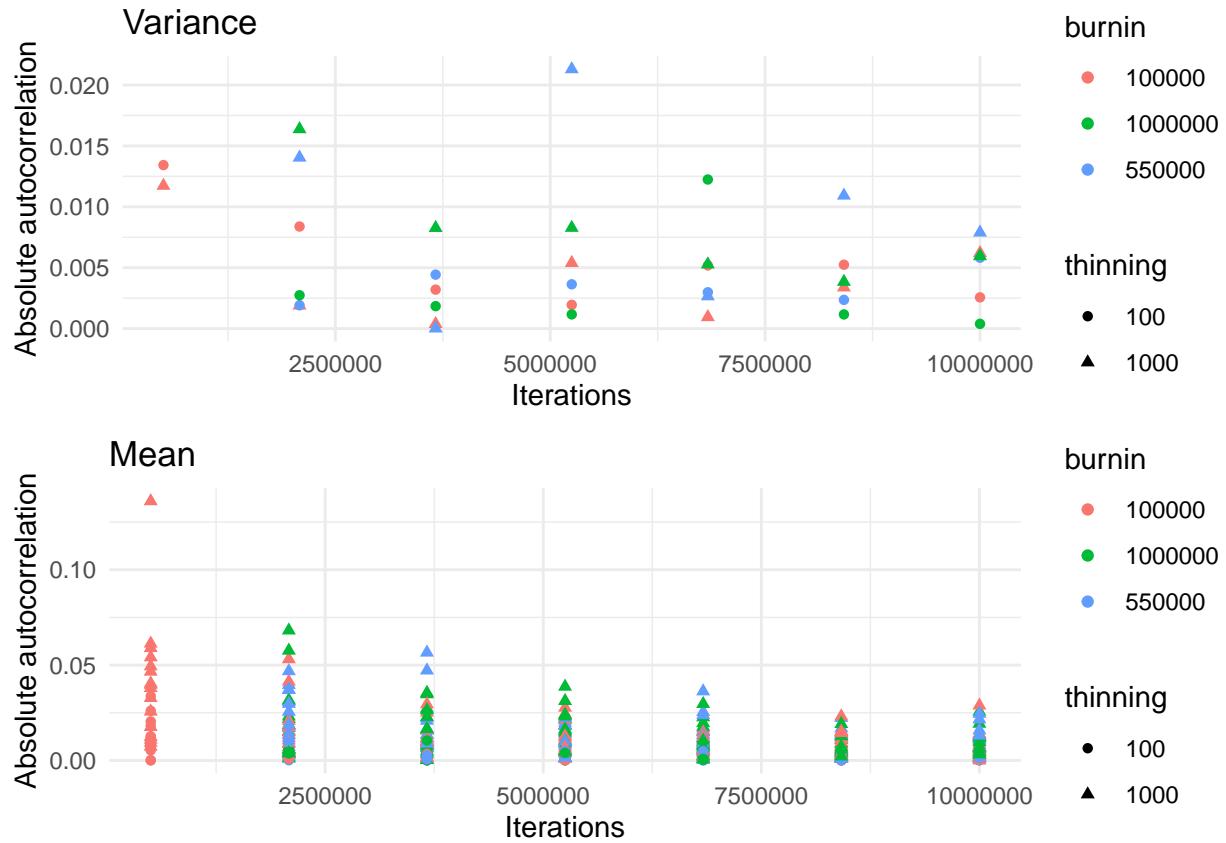


Figure 1: Autocorrelation of successively sampled mean and variance values from posterior distribution

## 35 2.2 Without Phylogeny

### 36 2.2.1 Model 1: Fission vs Cell Number

37 **2.2.2 Model 2: Fission vs Cell Types**

38 **2.2.3 Model 3: Germline vs Cell Numbers**

39 Should we subset to only those organisms that have sterile cells for the germline models?

40 **2.2.4 Model 4: Germline vs Cell Types**

41 **2.3 Phylogenetically Informed Models**

42 The datapoints are not independent: they have shared evolutionary history of varying degrees. Should we  
43 exclude some of the

44 **2.3.1 Model 5: Fission vs Cell Number**

45 **2.3.2 Model 6: Fission vs Cell Types**

```
46 ##      Iteration      Chain          Parameter
47 ##  Min.   : 1   Min.   :1.0   FissionOrBuddingObserved_Species : 420000
48 ##  1st Qu.:17501  1st Qu.:2.0   FissionOrBuddingObserved_Species?: 420000
49 ##  Median :35000  Median :3.5   FissionOrBuddingObserved_Species0: 420000
50 ##  Mean   :35000  Mean   :3.5   FissionOrBuddingObserved_Species1: 420000
51 ##  3rd Qu.:52500  3rd Qu.:5.0   scale(log(cell_number))       : 420000
52 ##  Max.   :70000  Max.   :6.0   species.Anabaena             : 420000
53 ##                               (Other)                      :52500000
54 ##      value
55 ##  Min.   :-5.3960
56 ##  1st Qu.:-0.6164
57 ##  Median :-0.1694
58 ##  Mean   :-0.1334
59 ##  3rd Qu.: 0.2997
60 ##  Max.   : 5.7347
61 ##
```

62 **2.3.3 Model 7: Germline vs Cell Number**

```
63 ##      Iteration      Chain          Parameter
64 ##  Min.   : 1   Min.   :1.0   germline_timing_simpleadult    : 420000
65 ##  1st Qu.:17501  1st Qu.:2.0   germline_timing_simpleearly   : 420000
66 ##  Median :35000  Median :3.5   germline_timing_simpleno_germline : 420000
67 ##  Mean   :35000  Mean   :3.5   species.Anabaena           : 420000
68 ##  3rd Qu.:52500  3rd Qu.:5.0   species.Astrephomene_gubernaculifera: 420000
69 ##  Max.   :70000  Max.   :6.0   species.Astrephomene_perforata  : 420000
70 ##                               (Other)                      :51660000
71 ##      value
72 ##  Min.   :-30.886
73 ##  1st Qu.:-7.096
74 ##  Median :-3.102
75 ##  Mean   :-2.121
76 ##  3rd Qu.: 2.123
77 ##  Max.   : 33.382
78 ##
```

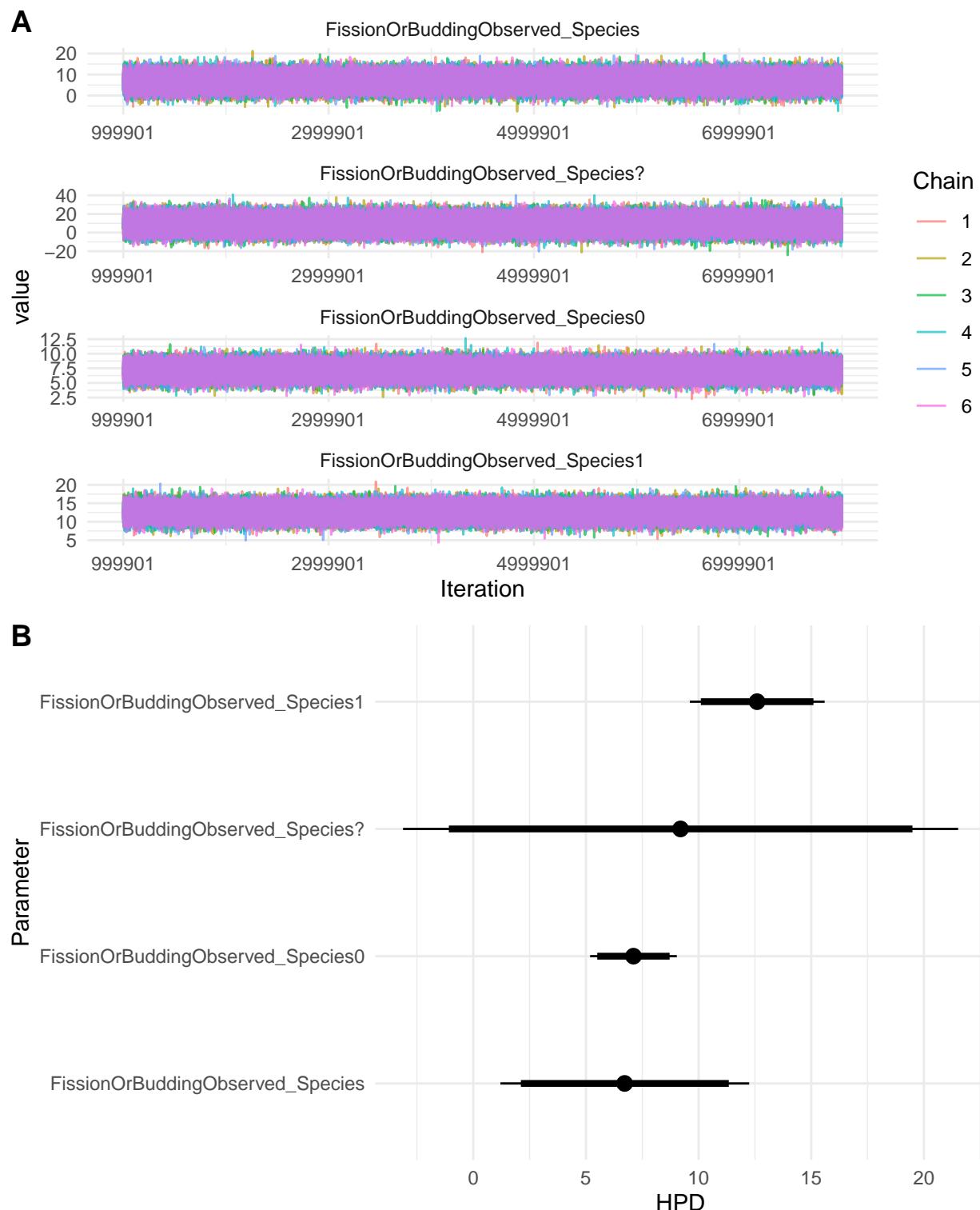


Figure 2: **Model 1: Cell Numbers vs Fission** *A* Traceplots for the estimated means  $f$ , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

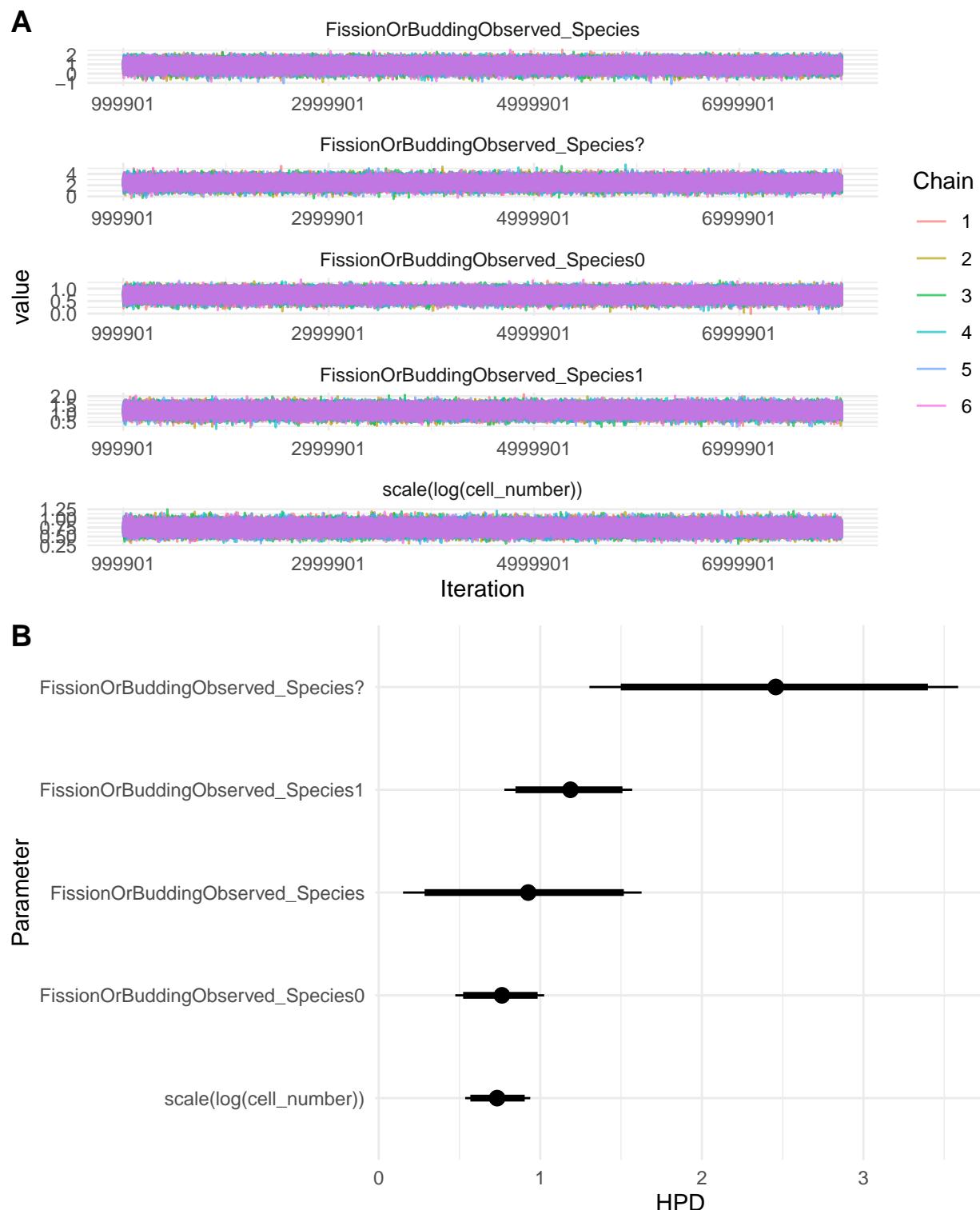


Figure 3: **Model 2: Cell Types vs Fission** *A* Traceplots for the estimated means  $f$ , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

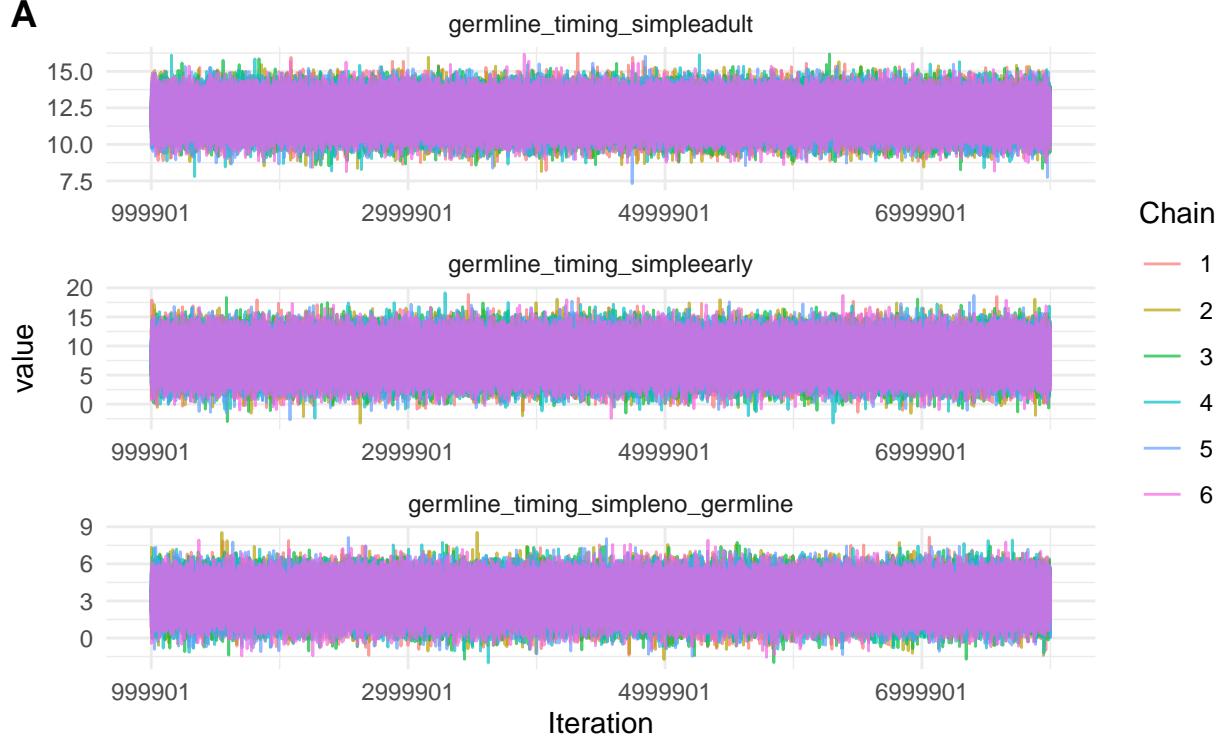
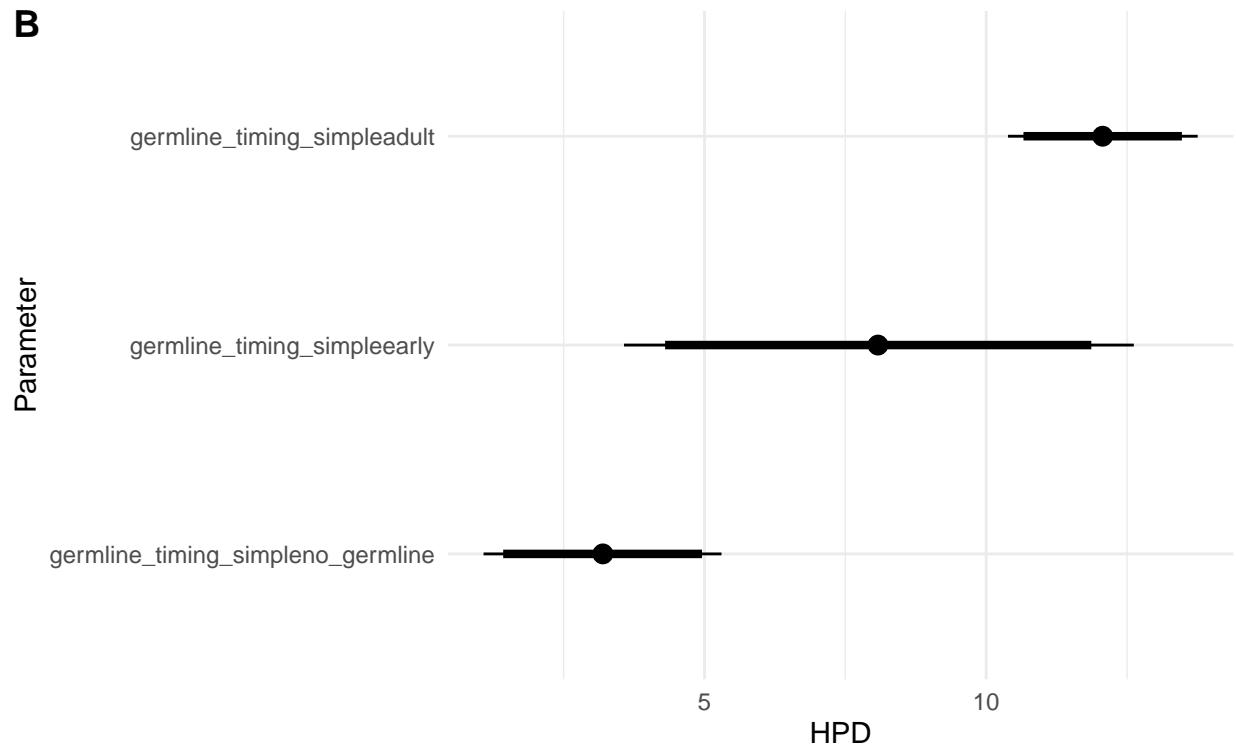
**A****B**

Figure 4: **Model 3: Cell number vs Germline A** Traceplots for the estimated means  $f$ ,  $B$  Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

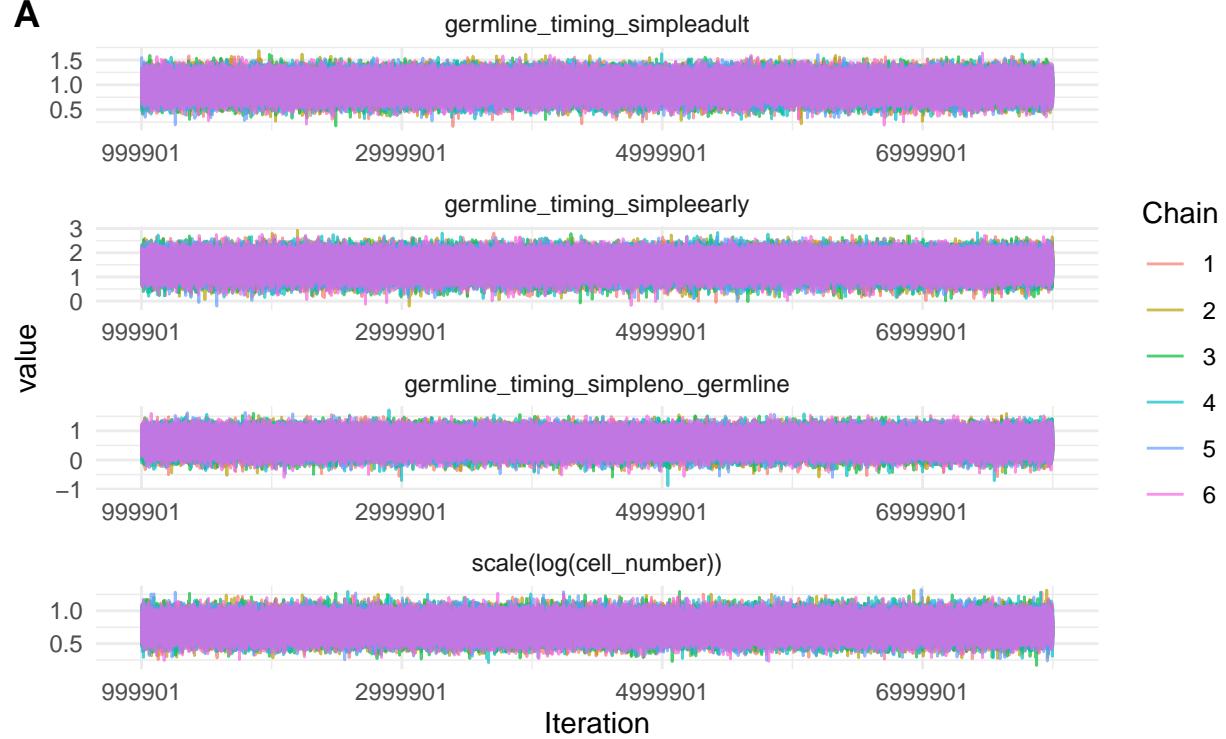
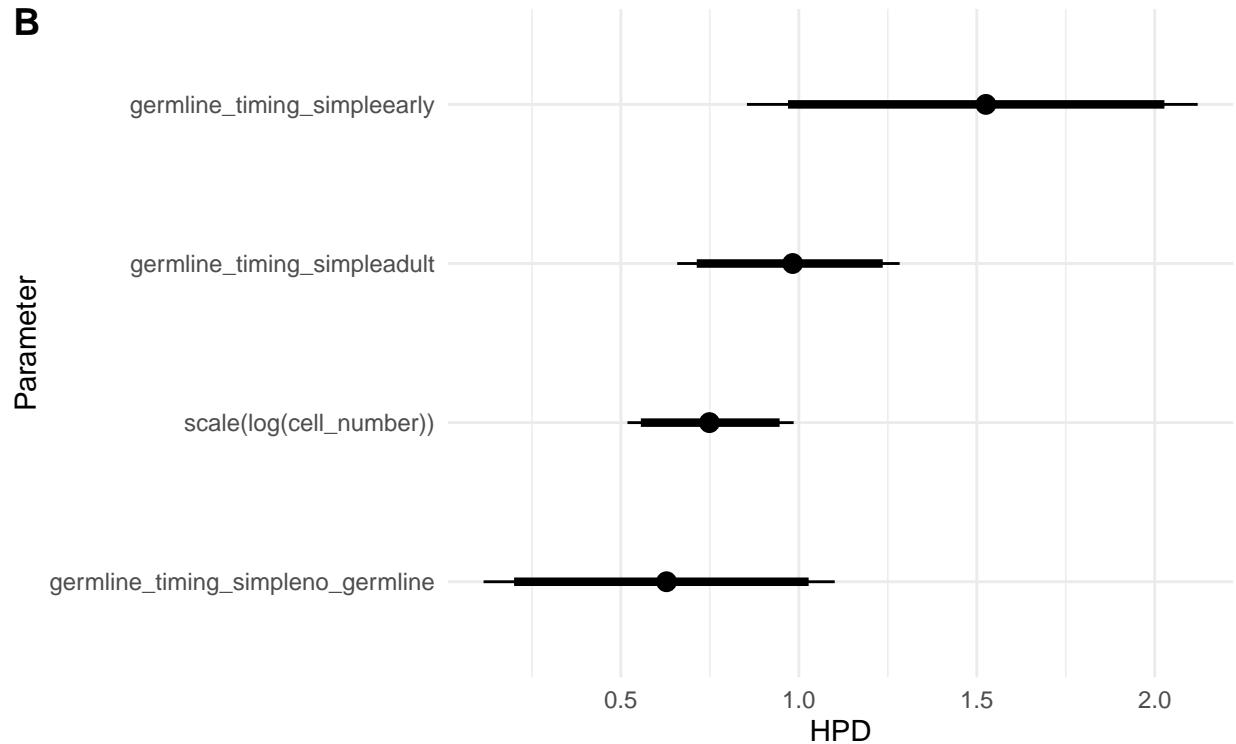
**A****B**

Figure 5: **Model 4: Cell Types vs Germline** A Traceplots for the estimated means f, B Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

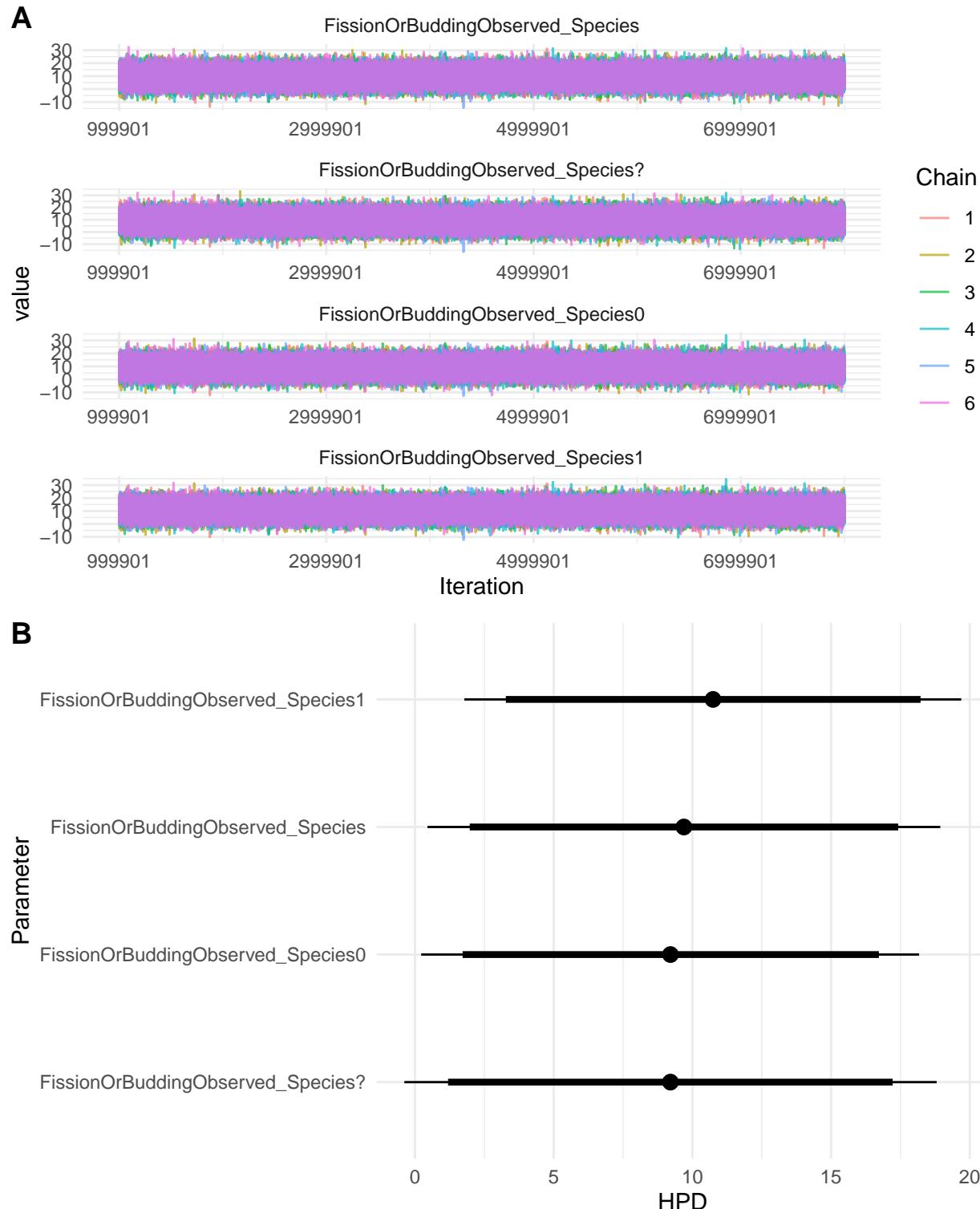


Figure 6: **Model 5: Cell Number vs Fission with germline** *A* Traceplots for the estimated means  $f$ , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

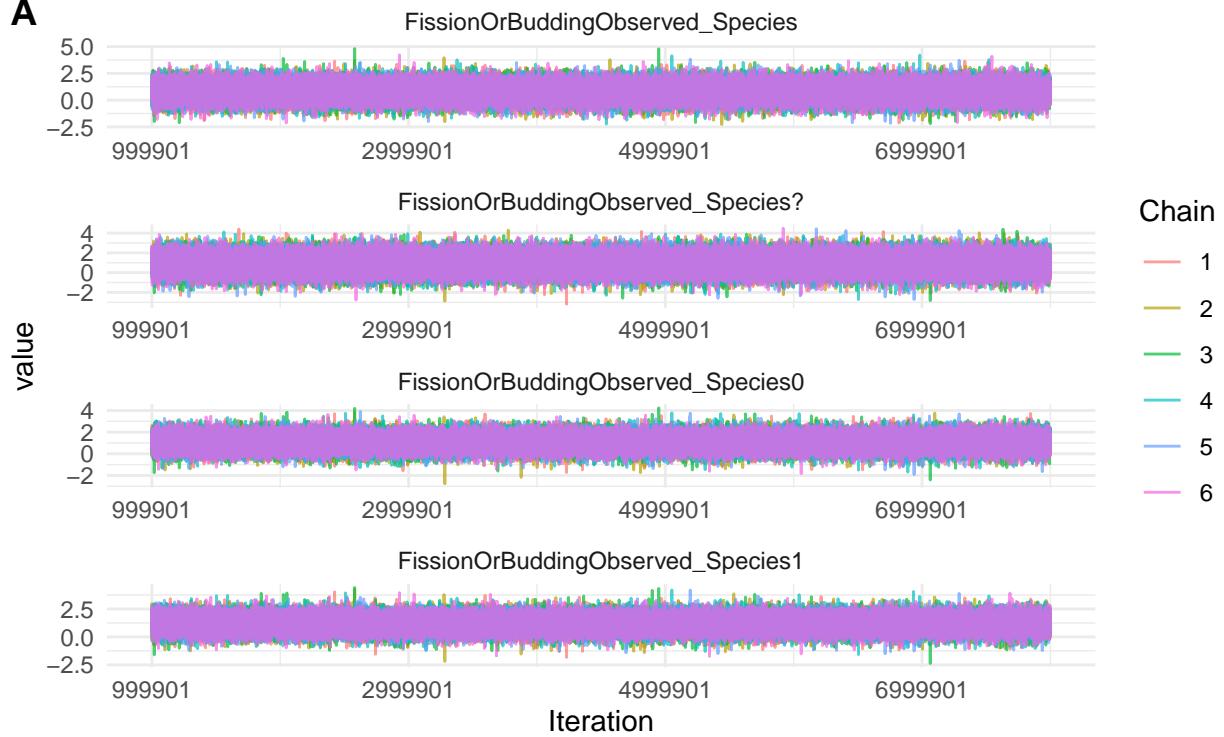
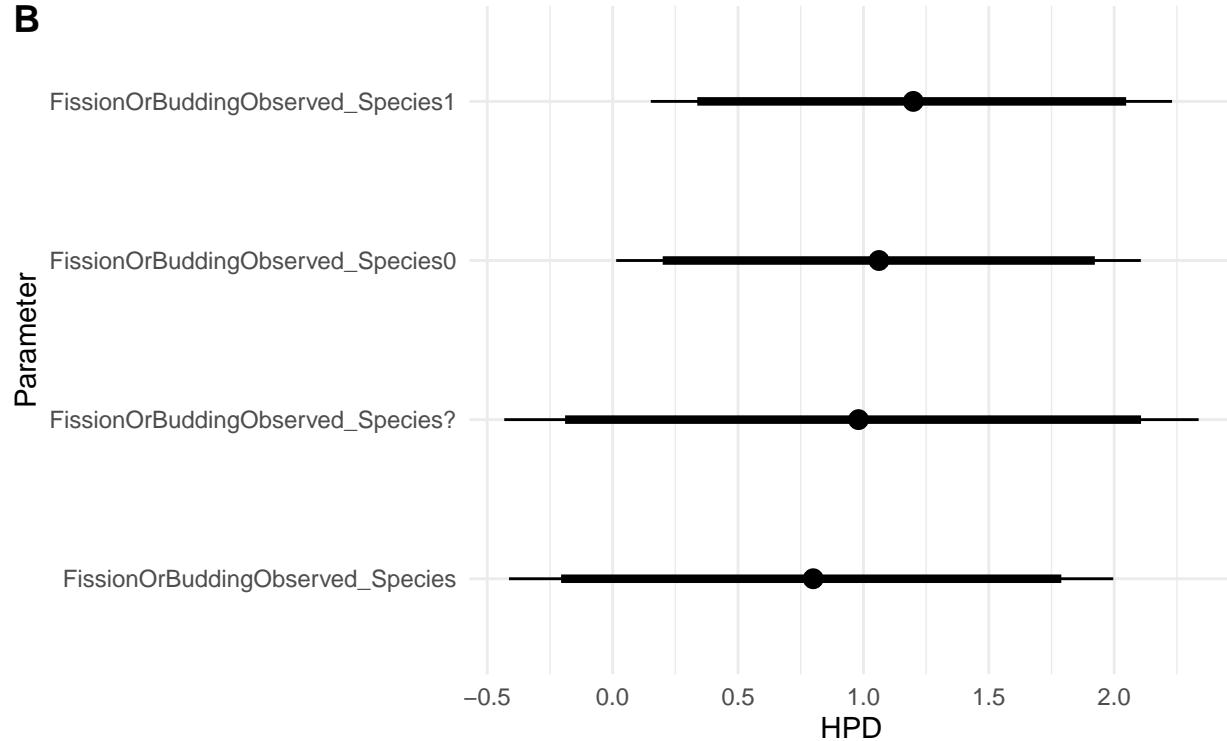
**A****B**

Figure 7: **Model 6: Cell Number vs Fission with germline** *A* Traceplots for the estimated means  $f$ , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

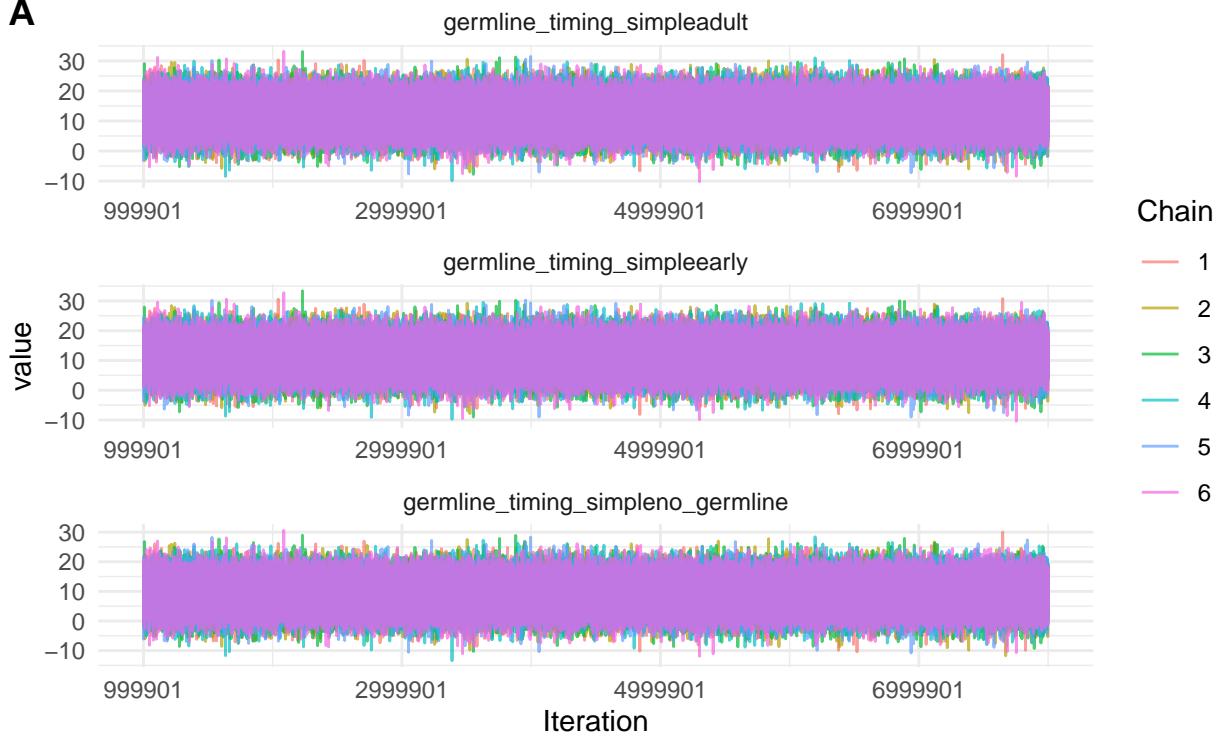
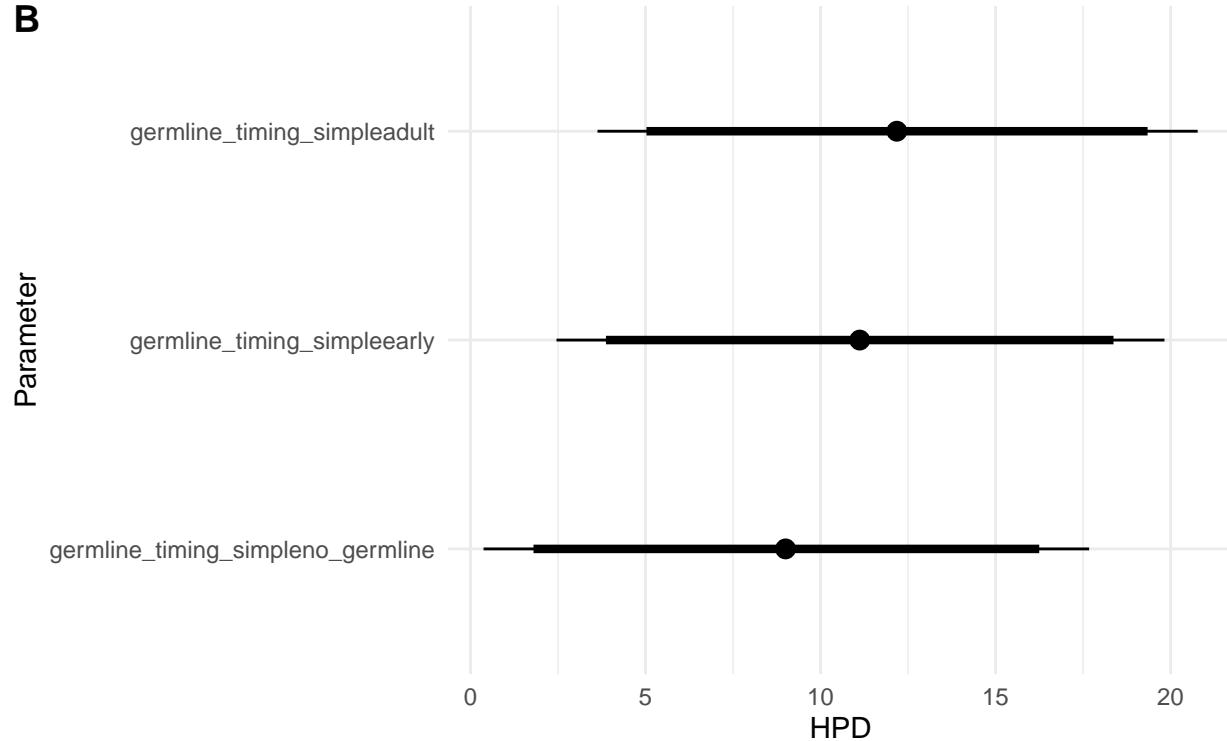
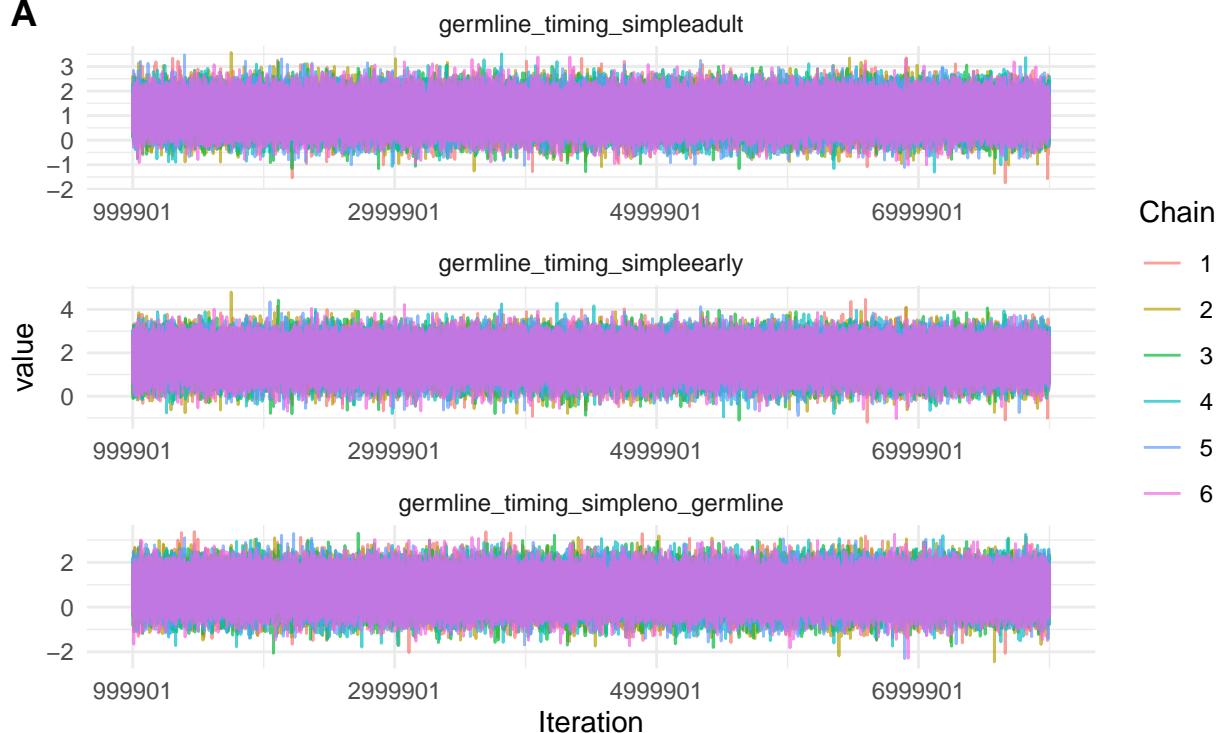
**A****B**

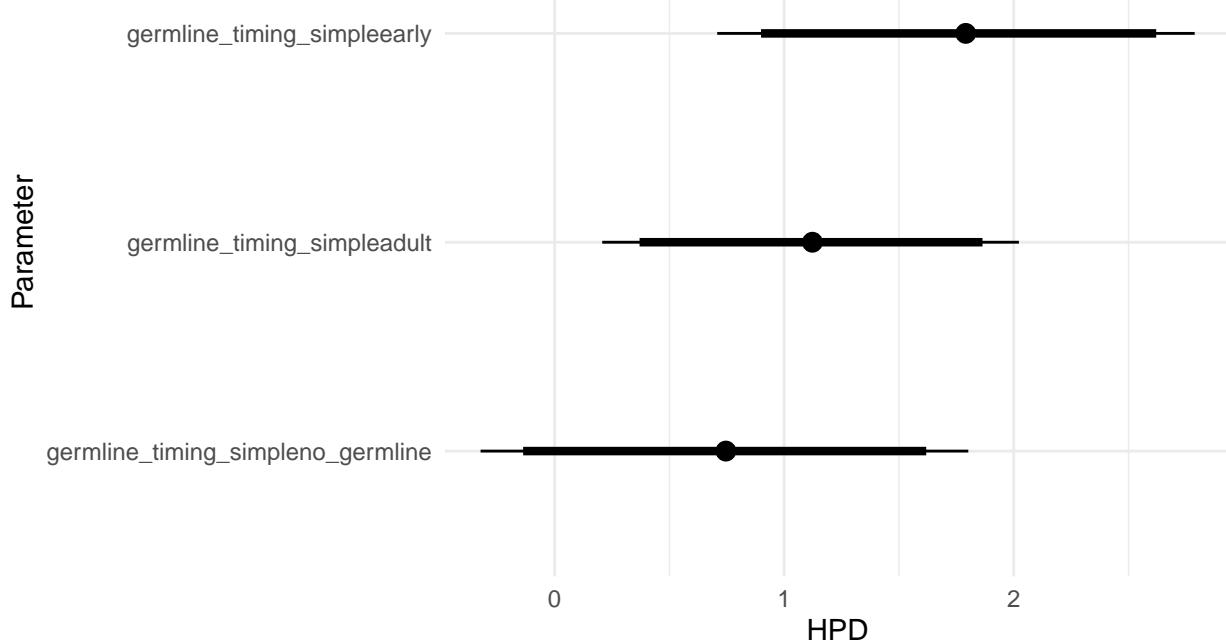
Figure 8: **Model 7: Cell Number vs Fission with germline** *A* Traceplots for the estimated means  $f$ , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

<sup>79</sup> **2.3.4 Model 8: Germline vs Cell Types**

**A**



**B**



<sup>80</sup>

<sup>81</sup> ## Open Questions

<sup>82</sup> Phylogenetic correlation between germline and fission– multivariate model? How to test this?

<sup>83</sup> **3 References**

<sup>84</sup> R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria:  
<sup>85</sup> R Foundation for Statistical Computing. <https://www.R-project.org/>.