

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

³

⁴ **Contents**

⁵ 1 Data Collection	¹
⁶ 1.1 Search strategy	1
⁷ 1.2 Production of phylogenetic tree	2
⁸ 2 Statistical Analyses	²
⁹ 2.1 MCMCglmm parameters	2
¹⁰ 2.2 Without Phylogeny	4
¹¹ 2.2.1 Model 1: Fission vs Cell Number	4
¹² 2.2.2 Model 2: Fission vs Cell Types	5
¹³ 2.2.3 Model 3: Germline vs Cell Numbers	5
¹⁴ 2.2.4 Model 4: Germline vs Cell Types	5
¹⁵ 2.3 Phylogenetically Informed Models	5
¹⁶ 2.3.1 Model 5: Fission vs Cell Number	14
¹⁷ 2.3.2 Model 6: Fission vs Cell Types	14
¹⁸ 2.3.3 Model 7: Germline vs Cell Number	14
¹⁹ 2.3.4 Model 8: Germline vs Cell Types	14
²⁰ 3 References	¹⁶

²¹ **1 Data Collection**

²² **1.1 Search strategy**

²³ Searches were conducted broadly for literature focussed on reproductive mode and germline development
²⁴ across the tree of life. This included chapters reviews and chapters within textbooks.

²⁵ As Fisher paper was used for estimates of individual complexity (which in turn used Bell paper as a foundation), narrower searches were conducted for each species/genus from Bell paper on Web of Knowledge and
²⁶ on Google Scholar.

²⁷ (ALL = (reproduct* OR sex* OR asex* OR vegetat* OR fissi* OR clonal* OR regenerat* OR
²⁸ rhizo* OR germ-line* OR germline* OR germ line* OR bud* OR fragment* OR parthenogen* OR
²⁹ stolon*)) AND (ALL = TAXON)

31 We recorded whether sexual, parthenogenetic/clonal and agametic have been observed as binary values. We
32 did not attempt to capture the relative frequency of different reproductive strategies, as these data do not
33 exist for the majority of species.

34 Research into reproduction and development is heterogeneous across organisms, and this is necessarily re-
35 reflected in the required searching effort for different groups: the reproductive biology and development of
36 model organisms such as *C. elegans*, *M. musculus*, *D. melanogaster*, *A. thaliana* are well known, but in
37 many other groups reproduction may never have been observed. We therefore conducted searches for each
38 species, but if no literature discussing reproductive strategy was observed, then we conducted an additional
39 search at the genus level. This assumes that genera will tend to be relatively similar in their reproductive
40 strategies: this is not always the case. The planarian *S. mediterranea*, for example, has strictly sexual and
41 strictly asexual strains within even the same species. However, we are focussed on patterns through longer
42 spans of evolution than between individual genera.

43 Caveats:

- 44 • Sexual organisms contain more cells because they have gonads that asexual organisms lack...
- 45 • hard to fit algae with gametophyte/sporophyte stages: they have life cycles where some stages can
46 fragment, where some stages can reproduce by spores, parthenogenesis or by sex. If an algae can stay
47 in one loop and reproduce exclusively through parthenogenesis/fragmentation, then counts– similar to
48 organisms that can fission, but don't always.

50 1.2 Production of phylogenetic tree

51 We used a phylogenetic tree to control for non-independence of species based on shared evolutionary history.
52 The tree was constructed within R using the latest evolutionary classifications found on the Tree of Life,
53 AlgaeBase.org, and the World Register of Marine species. The relationships among species were reconstructed
54 by ordering the taxa from Kingdom through to species, and grouping according to these names.

55 As a comparison, we also constructed a tree using the 'R Tree of Life Project'. These two trees were largely
56 congruent: some larger groups had switched places, but within these groups relationships were predominantly
57 the same. As the Rtol tree dropped X data points from the tree, we used the tree based on the taxa
58 names. Multichotomies within the tree were randomly resolved, before branch lengths were generated as
59 described by ([grafen1989?](#)). Branches smaller than 10^{-25} were deleted, and the dichotomies here collapsed
60 to multichotomies. Figure () shows a cophylogeny based on each tree.

61 2 Statistical Analyses

62 2.1 MCMCglmm parameters

63 All analyses were conducted in R (R Core Team 2021) using the package MCMCglmm ([MCMCglmm?](#)),
64 while documents were produced using [RMarkdown]. All data and code are accessible at [github..](#)

65 Model parameters were optimised using the first model described in our results, for which we ran a total of
66 38 MCMCglmm chains of varying lengths (500000 - 10000000 iterations), with varying warm-ups (100000
67 - 1000000, and with thinning of either 100 or 1000 fold, see Figure S1. All subsequent models were then
68 fit using the combination of these parameters where the autocorrelation of successive sampled mean and
69 variance were minimal: 8×10^6 iterations, a warm-up of 10^6 iterations and thinning by a factor of 100. In
70 all fitted models, the autocorrelation was well below the suggested tolerable maximum of 0.1 ([hadfield?](#)).
71 For each model, 6 chains were run which were visually inspected for chain convergence. Convergence was
72 also supported by the Gelman-Rubin ([Gelman-Rubin?](#)) convergence diagnostic, which approximated 1
73 (<1.05) in all cases–these are reported in the summary of each model below.

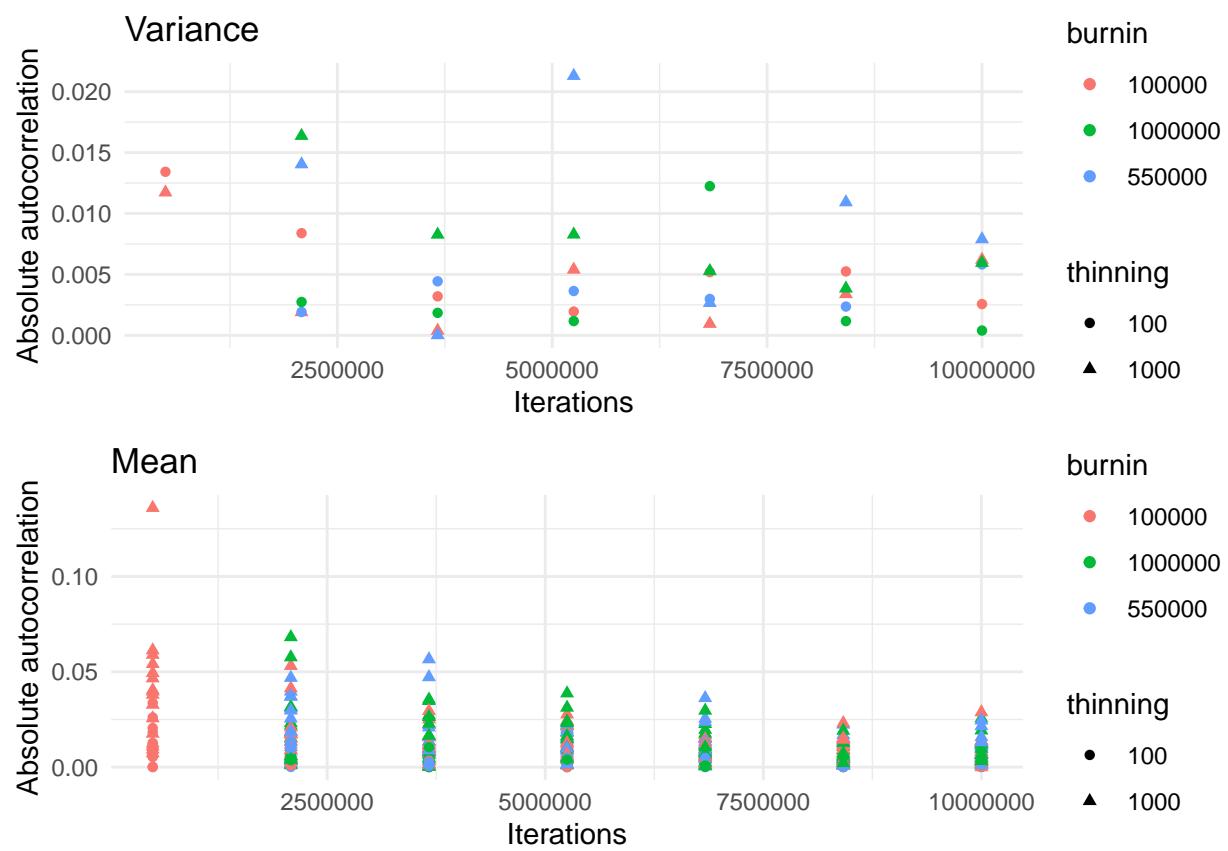


Figure 1: Autocorrelation of successively sampled mean and variance values from posterior distribution

⁷⁴ **2.2 Without Phylogeny**

⁷⁵ **2.2.1 Model 1: Fission vs Cell Number**

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

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

⁷⁸ Should we subset to only those organisms that have sterile cells for the germline models?

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

⁸⁰ **2.3 Phylogenetically Informed Models**

⁸¹ The datapoints are not independent: they have shared evolutionary history of varying degrees. Should we
⁸² exclude some of the

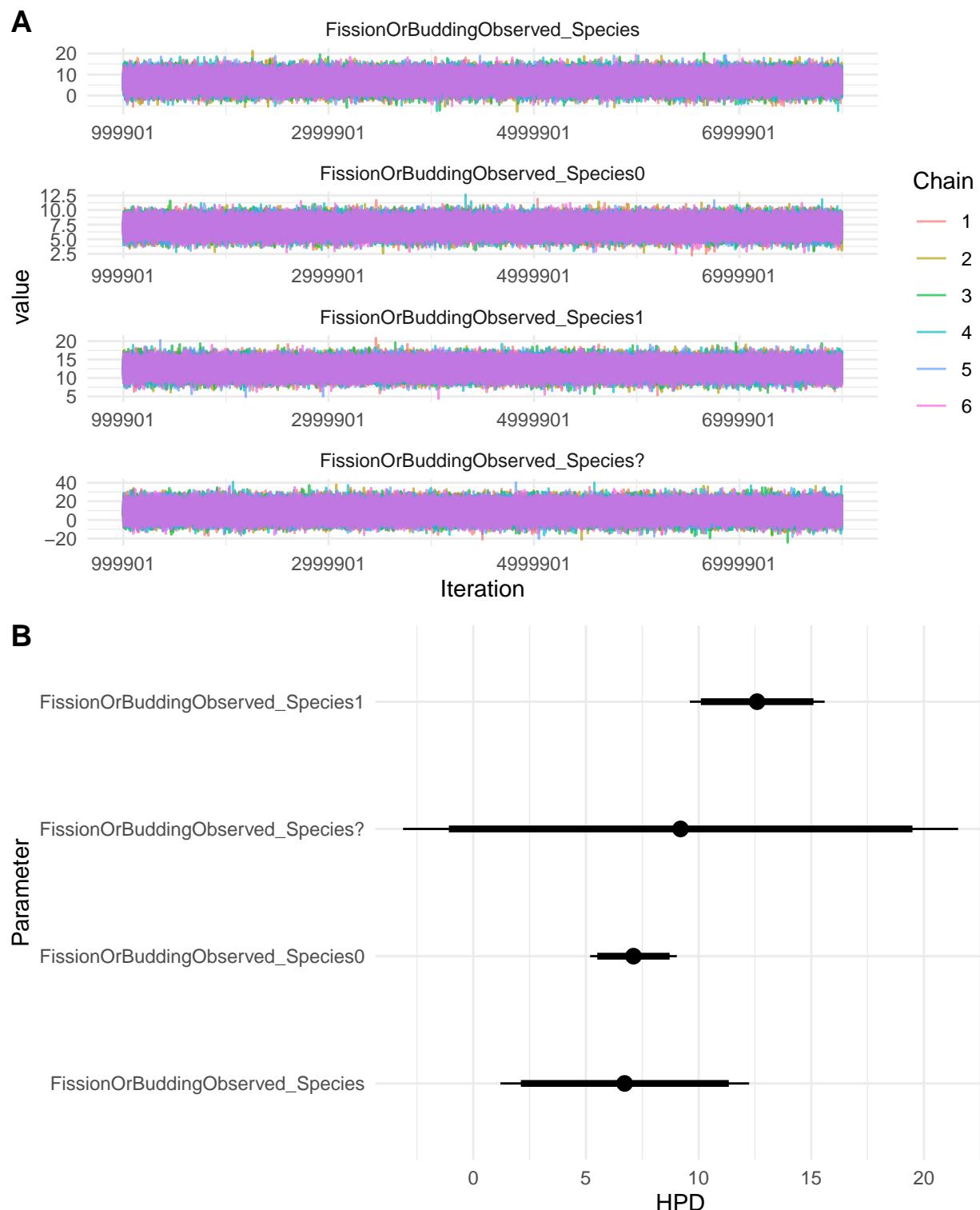


Figure 2: **Model 1: Cell Numbers vs Fission** *A* Traceplots for the estimated means f , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

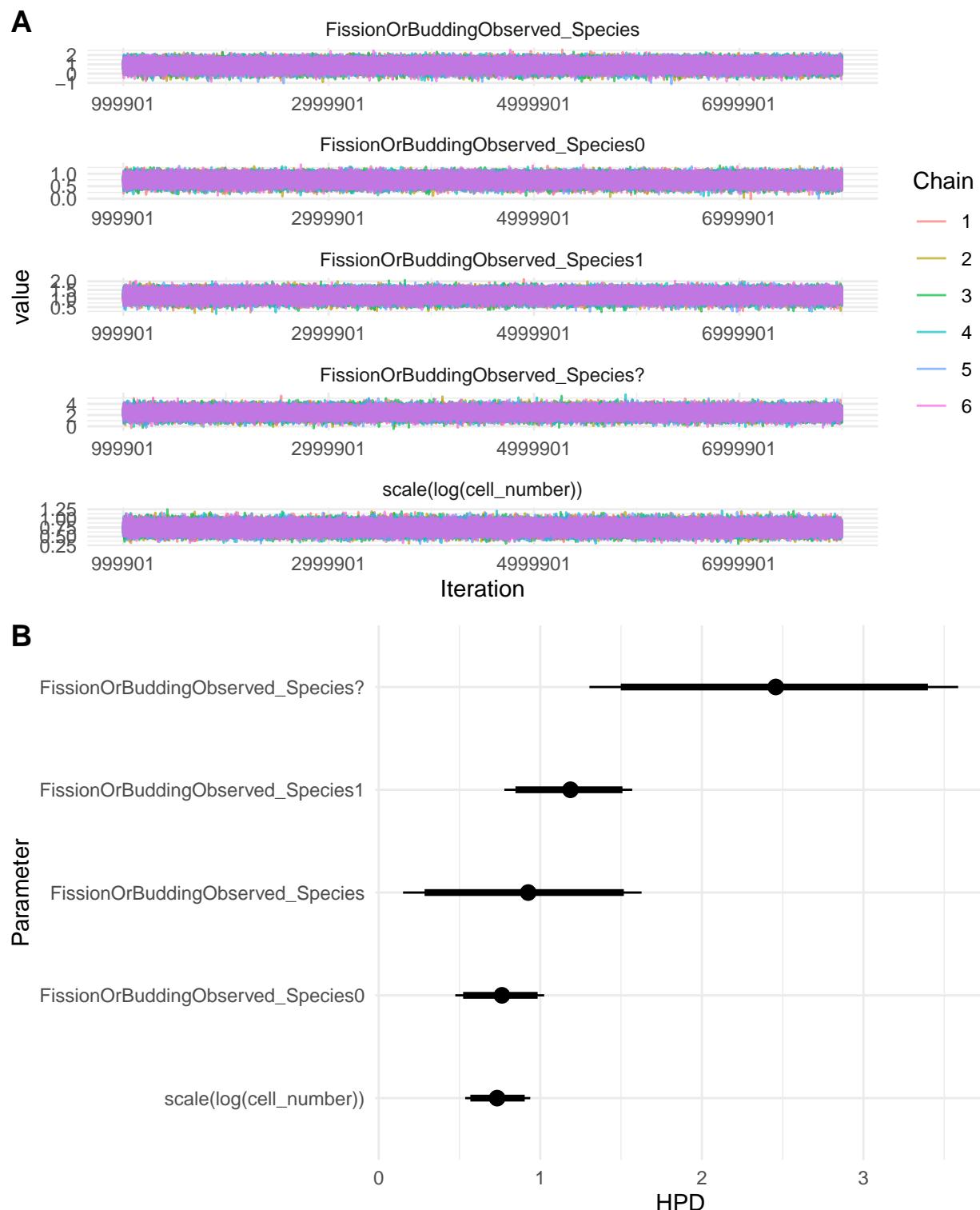


Figure 3: **Model 2: Cell Types vs Fission** *A* Traceplots for the estimated means f , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

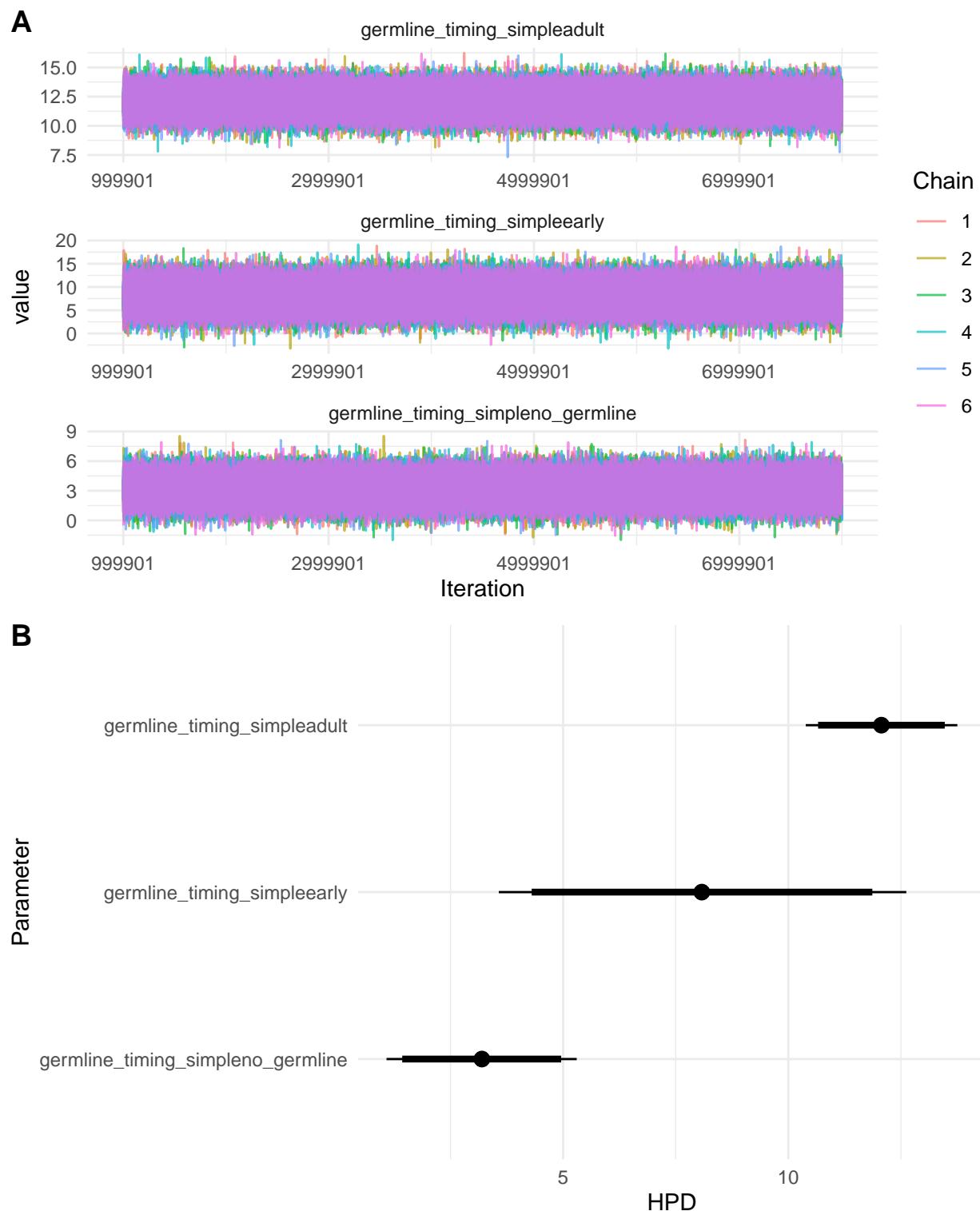


Figure 4: **Model 3: Cell number vs Germline A** Traceplots for the estimated means f , B Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

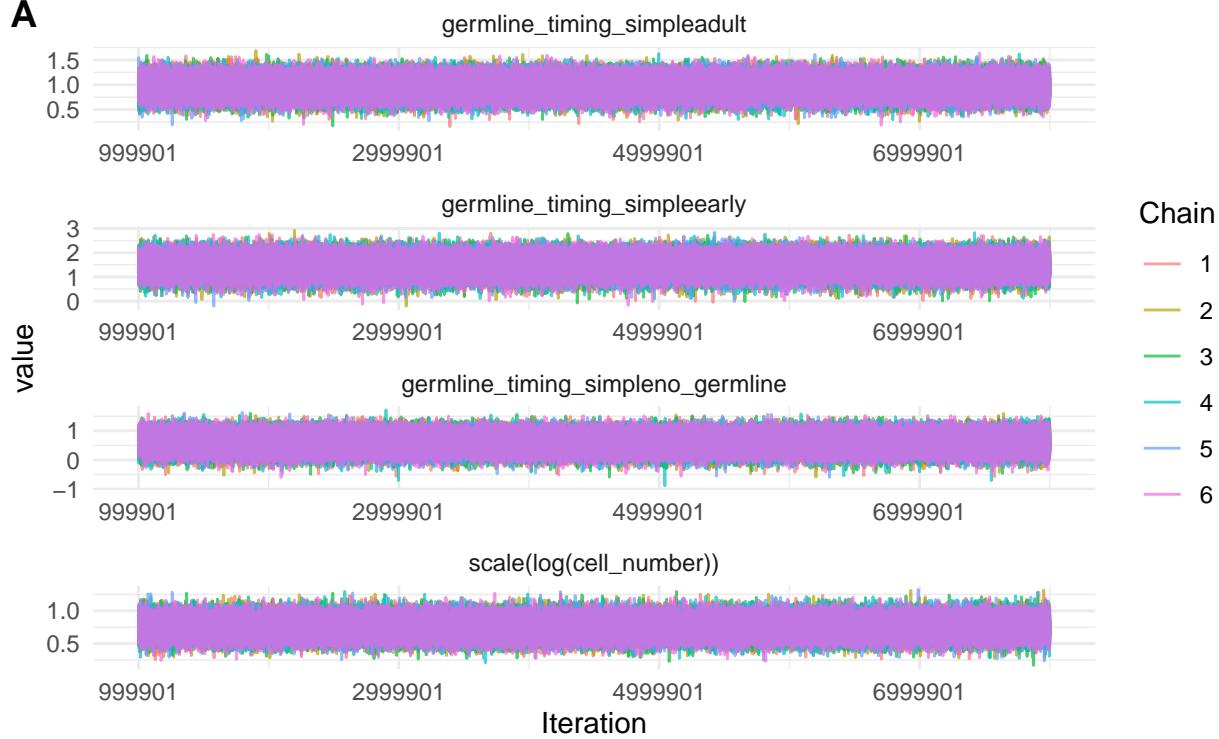
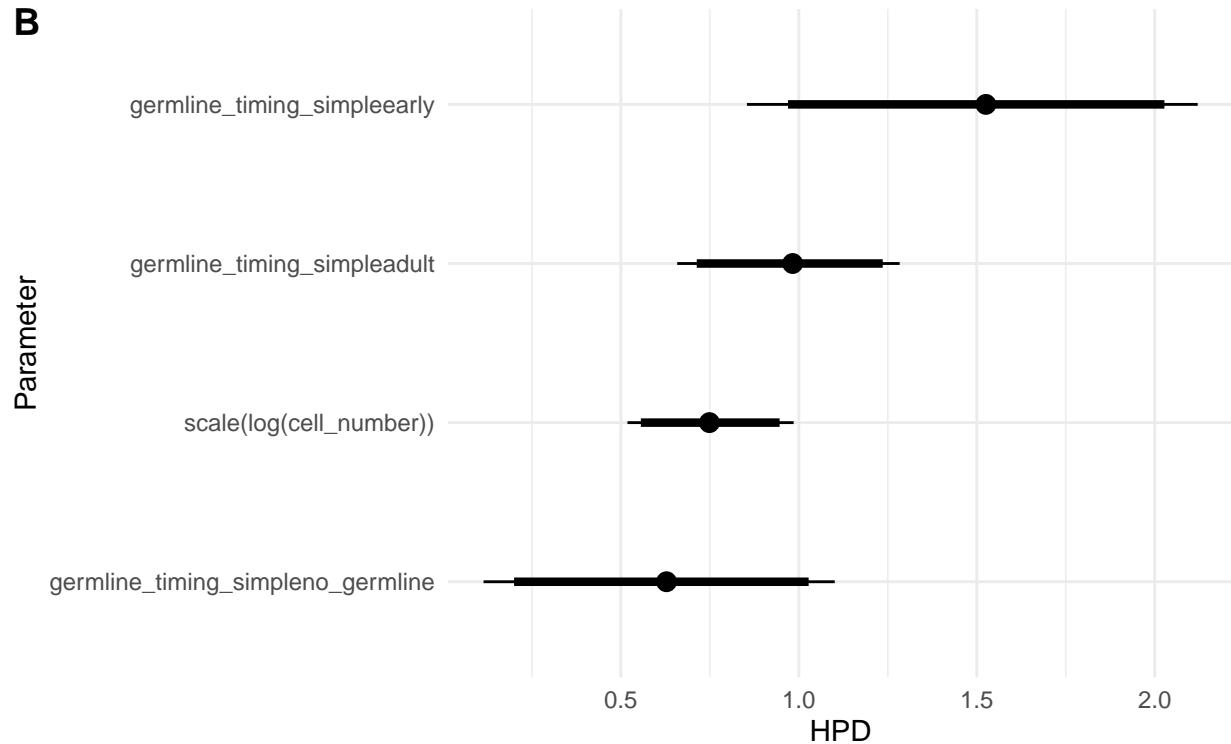
A**B**

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.

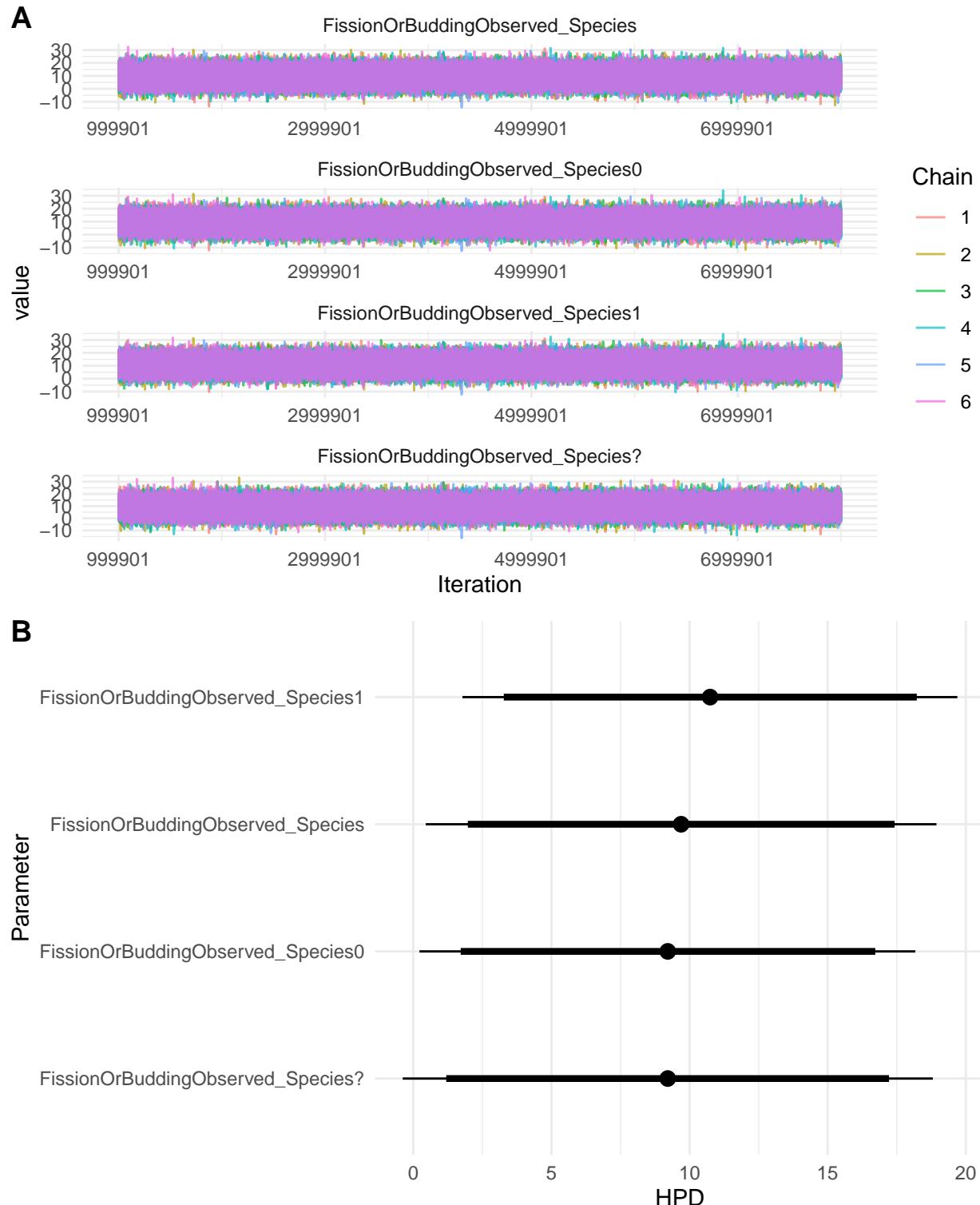


Figure 6: **Model 5: Cell Number vs Fission with germline** *A* Traceplots for the estimated means f , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

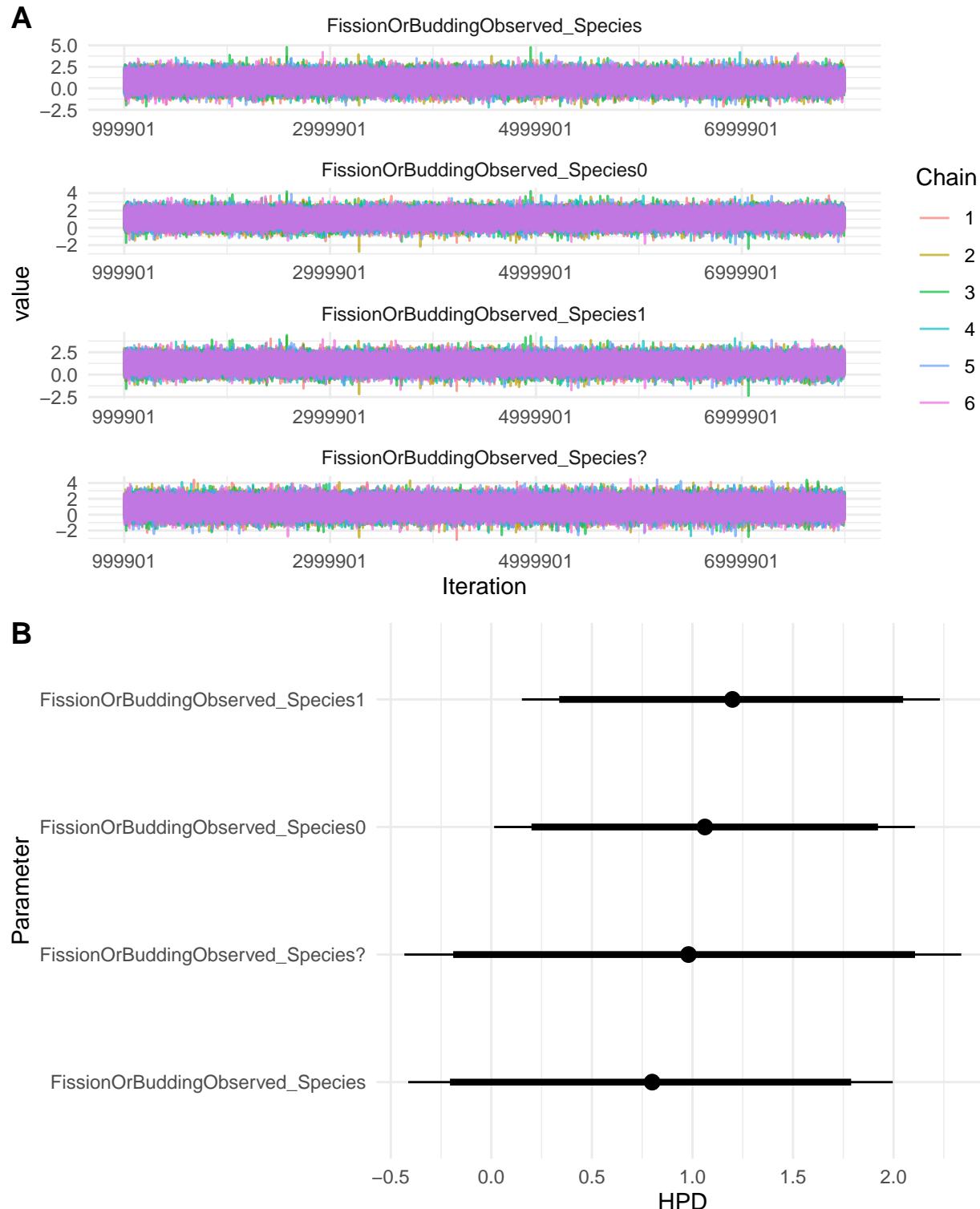


Figure 7: **Model 6: Cell Number vs Fission with germline** *A* Traceplots for the estimated means f , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

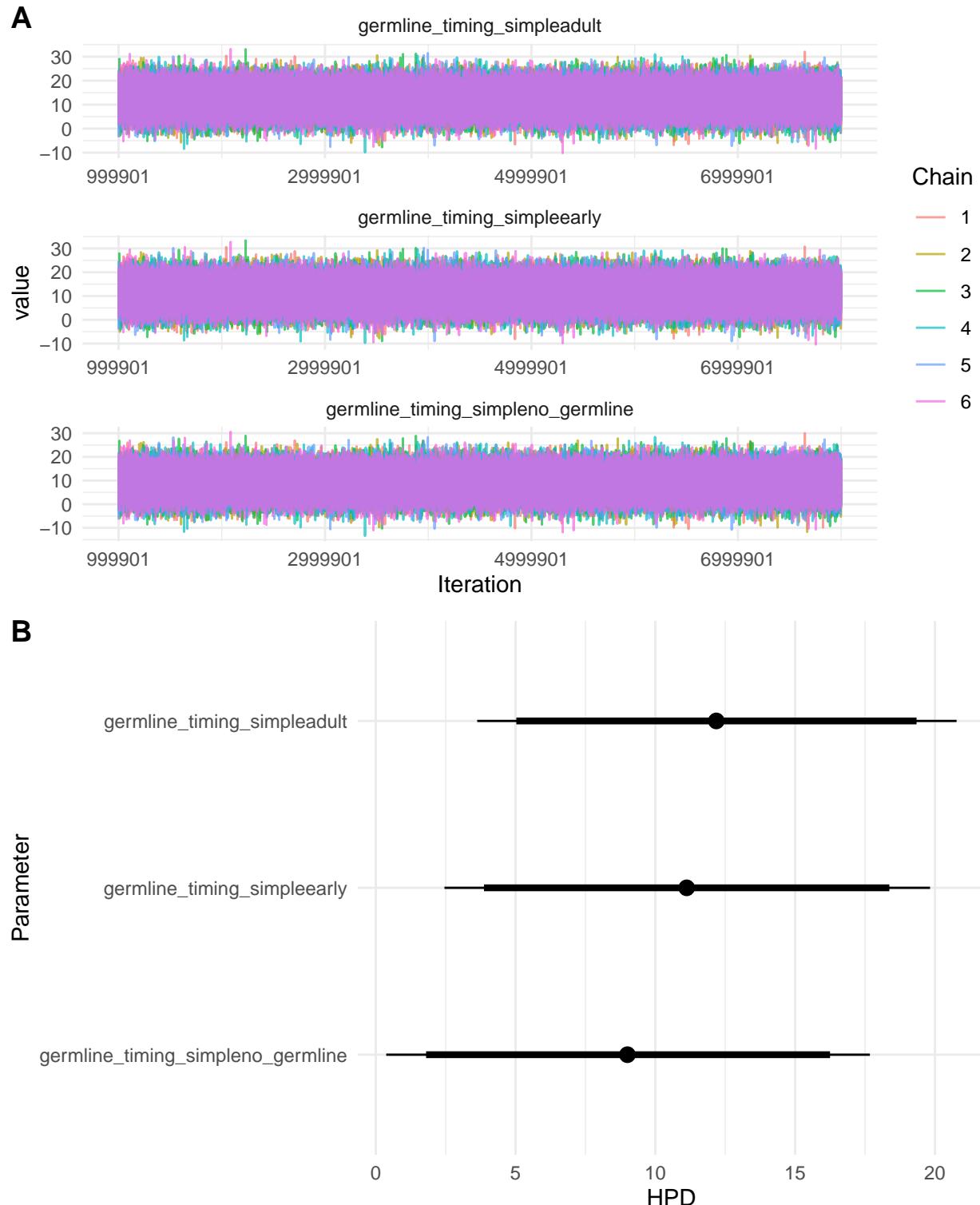


Figure 8: **Model 7: Cell Number vs Fission with germline** *A* Traceplots for the estimated means f , *B* Estimates for means from posterior distribution, dots represent median, thick and thin lines indicate 90% and 95% of highest posterior density regions, respectively.

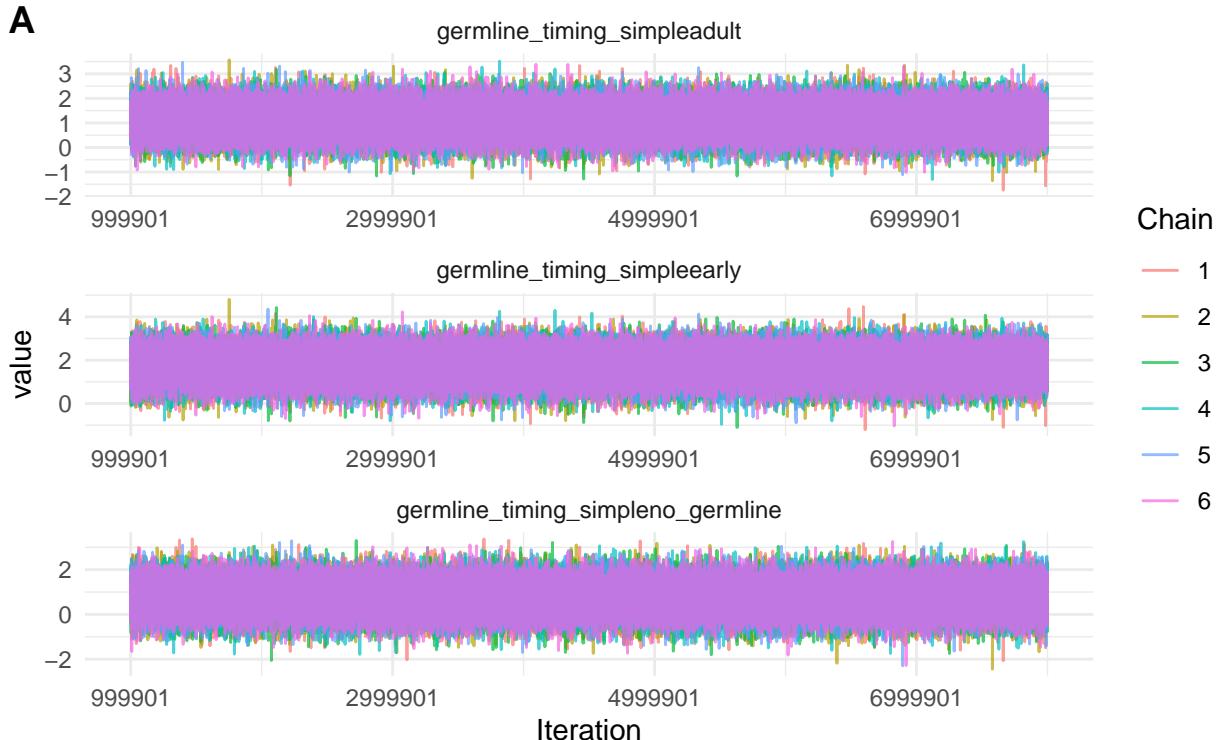
⁸³ **2.3.1 Model 5: Fission vs Cell Number**

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

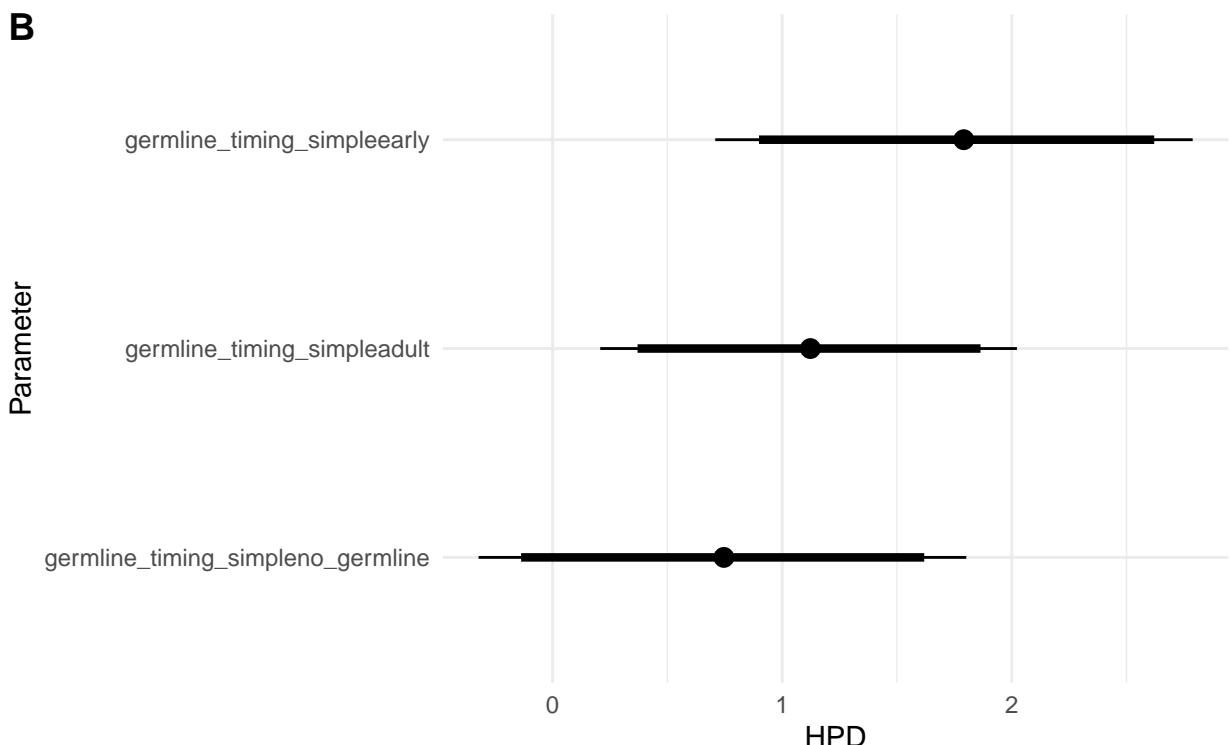
⁸⁵ **2.3.3 Model 7: Germline vs Cell Number**

⁸⁶ **2.3.4 Model 8: Germline vs Cell Types**

A



B



⁸⁷

⁸⁸ ## Open Questions

⁸⁹ Phylogenetic correlation between germline and fission– multivariate model? How to test this?

90 **3 References**

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