GeneBench-Pro Case Study: Wright–Fisher Selection under Ancient-DNA Error

GeneBench-Pro

June 26, 2026

1 Overview

Ancient-DNA allele-frequency time series are useful for studying recent selection, but the observed read counts are not direct measurements of the favored allele when allele orientation, post-mortem damage, sparse sampling, and changing effective population size all enter the same analysis. This case study asks an analyst to estimate the positive haploid selection coefficient for the more strongly selected of two loci from four compressed tables: two per-locus read-count time series, a variant-orientation table, and a time-varying effective-population-size schedule. The released answer has two fields, `selected_locus` and `s`; the realized target values and matching rules are summarized in the answer-field table below.

The correct analysis is a three-stage cascade. Stage 1 uses the outgroup allele to polarize the derived state, converting ALT read counts to derived read counts only when the outgroup equals REF; here outgroup equals ALT, so the derived allele is REF. Stage 2 recognizes that the REF/ALT pair is C/T in ancient samples and maps the reported average error 0.16, together with the prompt's approximately 1% instrument-error floor, into directional rates $e_{\text{derived}\to\text{ancestral}} = 0.31$ and $e_{\text{ancestral}\to\text{derived}} = 0.01$. Stage 3 fits a hidden Markov model in which unobserved haploid allele counts evolve under Wright–Fisher selection and drift through the supplied N_g schedule, then compares the maximum-likelihood $\hat{s}$ values for loci A and B. These steps are decision-relevant because treating ALT as derived, treating the average error as symmetric, or replacing the likelihood with a simple trend all produce numerically plausible answers outside the public tolerance.

2 Released Prompt and Files

Prompt

```
You are given allele-frequency time series data from two haploid loci sampled over
    multiple generations.
One locus is under stronger positive selection than the other. Estimate the selection
    coefficient s for the more strongly selected locus, where s > 0 means the derived
    allele is favored.
Assume instrument-driven sequencing error is ~1%. The seq_error column is the average of
    the two directional allele-miscall rates for that locus and sample.
The selected_locus value must be "A" or "B".
These data came from a real experiment; you will be graded not just on numerical
    correctness but the quality of analytical reasoning you exhibit; do not attempt to
    take any shortcuts.
Return your final answer as exactly one JSON object.
Do not wrap the JSON in markdown.
Do not add prose before or after the JSON.
Do not omit any keys shown in the example.
Return the JSON object in your final answer:
{
"answer": {
"selected_locus": "<string>",
"s": <float>
},
"reasoning": "<description of method and QC>"
}
```

Released data files

File	Format	Contents
`locus_A_timeseries.tsv.gz`	.tsv.gz	Locus A observations by sampled generation, with ALT read counts, total read depth, average sequencing-error field, and ancient sample year.
`locus_B_timeseries.tsv.gz`	.tsv.gz	Same schema for locus B, allowing the analyst to compare the two candidate selected loci under the same demographic schedule.
`variant_info.tsv.gz`	.tsv.gz	Per-locus REF/ALT alleles and outgroup allele used to polarize ancestral versus derived state before fitting the time-series model.
`Ne_schedule.tsv.gz`	.tsv.gz	Piecewise haploid effective-population-size schedule, with generation intervals and N_g values used in the Wright–Fisher transition model.

3 Answer Fields and Tolerances

The answer file must contain

$$\begin{aligned}\texttt{selected\_locus} &= \arg\max\{\hat{s}_A, \hat{s}_B\},\\ \texttt{s} &= \max\{\hat{s}_A, \hat{s}_B\}.\end{aligned}$$

The reference likelihood estimates are $\hat{s}_A = 0.101255$ and $\hat{s}_B = 0.024774$.

Answer field	Ground truth	Tolerance / matching rule	Interpretation
`selected_locus`	`A`	Exact case-sensitive string match	Locus with the larger fitted positive haploid selection coefficient.
`s`	0.101256	Absolute error ≤ 0.020; valid range $[-1, 1]$	Selection coefficient for the selected locus; accepted interval $[0.081256, 0.121256]$.

4 Structure Diagram

Figure 1 is the report roadmap: it shows the solver-visible evidence, the three decision points that change the answer, and the graded output that the later walkthrough derives numerically.

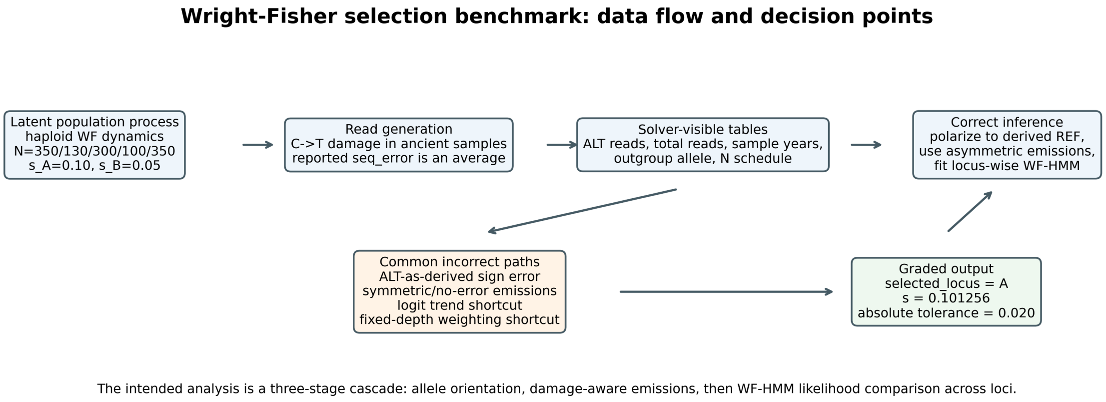

Figure 1: Data flow and decision points. Arrows show the left-to-right analysis flow. Blue boxes are the intended data-generating and solver-visible stages, the orange box lists plausible but incorrect shortcut paths, and the green box is the graded output. WF denotes Wright–Fisher dynamics and WF–HMM denotes the Wright–Fisher hidden Markov model used for the locus-wise likelihood comparison. REF and ALT are the reference and alternate alleles in the released variant table; here the correct orientation polarizes the derived allele to REF.

5 Variables and Assumptions

- $\ell \in \{A, B\}$: locus index. Locus A is generated with stronger selection than locus B.

- $g \in \{0, \ldots, 60\}$: generation index.
- $N_g \in \{350, 130, 300, 100\}$: haploid effective population size, fixed by the schedule in Equation (1).
- $p_{\ell,g} \in [0, 1]$: latent derived-allele frequency at locus ℓ and generation g.
- $X_{\ell,g} \in \{0, \ldots, N_g\}$: latent derived-allele count after drift. Conditional distribution is binomial under the Wright–Fisher transition.
- s_ℓ: haploid selection coefficient for the derived allele. The simulator uses $s_A = 0.10$ and $s_B = 0.05$.
- t_j: sampled generations $(6, 12, 18, 24, 30, 36, 42, 48, 54, 60)$.
- R_j: fixed read depth at sampled generation t_j, equal to $(40, 45, 55, 70, 90, 0, 130, 170, 210, 250)$.
- Y_j: sample year, spanning -4500 to -2500 across the ten sampled generations.
- $K^{\mathrm{ALT}}_{\ell,j}$: observed ALT read count in the visible time-series table.
- $K_{\ell,j}$: derived read count after orientation. Here $K_{\ell,j} = R_j - K^{\mathrm{ALT}}_{\ell,j}$ because derived equals REF.
- e_{da}: probability that a true derived base is called ancestral; for this C/T case study it is 0.31.
- e_{ad}: probability that a true ancestral base is called derived; it is fixed at the instrument-error floor 0.01.
- $\bar{e}$: reported average error in the visible tables; here $\bar{e} = (0.31 + 0.01)/2 = 0.16$.

The scientific assumptions are haploid selection, Wright–Fisher drift, binomial read sampling, and ancient-DNA damage that primarily elevates C→T misincorporation in old samples. Published ancient-DNA papers support the directionality and practical importance of post-mortem cytosine deamination [4,5]. The numerical split into 0.31 and 0.01 is reconstructed from released evidence: the prompt supplies the approximately 1% instrument-error floor and states that `seq_error` is an average of two directional rates, the time-series files report average error 0.16, and the sample ages plus C/T orientation identify the high-error direction.

6 Data-Generating Process

The population-size schedule is fixed:

$$N_g = \begin{cases} 350, & 0 \le g \le 18, \\ 130, & 19 \le g \le 26, \\ 300, & 27 \le g \le 37, \\ 100, & 38 \le g \le 47, \\ 350, & 48 \le g \le 60. \end{cases} \tag{1}$$

This schedule creates a high-coverage late epoch but also two bottleneck intervals where genetic drift can partially reverse a deterministic upward trajectory.

Initial derived frequencies are fixed at

$$p_{A,0} = 0.15, \qquad p_{B,0} = 0.35. \tag{2}$$

Locus B starts higher, so a naive comparison of raw frequencies can make the weaker locus look competitive.

Haploid selection maps the current derived frequency to the post-selection frequency

$$\tilde{p}_{\ell,g} = \frac{p_{\ell,g}(1+s_\ell)}{1+p_{\ell,g}s_\ell}, \qquad s_A = 0.10, \quad s_B = 0.05. \tag{3}$$

The derived allele has fitness $1 + s_\ell$ and the ancestral allele has fitness 1.

Drift samples the next generation's count:

$$X_{\ell,g+1} \mid p_{\ell,g}, s_\ell \sim \text{Binomial}(N_{g+1}, \tilde{p}_{\ell,g})\,, \tag{1}$$

$$p_{\ell,g+1} = X_{\ell,g+1}/N_{g+1}. \tag{4}$$

This is the Wright–Fisher part of the generator and is also the transition model used by the correct estimator.

The visible variant table gives REF=C, ALT=T, and outgroup=T for both loci. The outgroup allele is treated as ancestral, so

$$\text{derived allele} = \begin{cases} \text{ALT}, & \text{outgroup} = \text{REF}, \\ \text{REF}, & \text{outgroup} = \text{ALT}. \end{cases} \tag{5}$$

For these loci the second case applies, so derived reads are REF reads.

Ancient-DNA damage is encoded as asymmetric read error. With baseline instrument error 0.01 and C→T damage boost 0.30,

$$e_{\text{REF}\to\text{ALT}} = 0.01 + 0.30 = 0.31, \tag{2}$$

$$e_{\text{ALT}\to\text{REF}} = 0.01, \tag{3}$$

$$\bar{e} = (0.31 + 0.01)/2 = 0.16. \tag{6}$$

Because derived=REF and ancestral=ALT, $e_{\text{REF}\to\text{ALT}}$ is the derived-to-ancestral error rate. The visible `seq_error` column contains only $\bar{e}$.

At sampled generation t_j, read depth is fixed by

$$R_j \in (40, 45, 55, 70, 90, 0, 130, 170, 210, 250). \tag{7}$$

The zero-depth observation at generation 36 is intentionally missing and should contribute no emission likelihood.

The probability of observing an ALT read is

$$\pi^{\text{ALT}}_{\ell,j} = p_{\ell,t_j}\, e_{\text{REF}\to\text{ALT}} + (1 - p_{\ell,t_j})(1 - e_{\text{ALT}\to\text{REF}}), \tag{8}$$

and observed ALT counts are sampled as

$$K^{\text{ALT}}_{\ell,j} \mid p_{\ell,t_j}, R_j \sim \text{Binomial}(R_j, \pi^{\text{ALT}}_{\ell,j}). \tag{9}$$

The analyst must invert orientation before fitting, so

$$K_{\ell,j} = R_j - K^{\text{ALT}}_{\ell,j}. \tag{10}$$

7 Analyst Walkthrough

Stage 0: inspect the raw trajectories before modeling

The first visible pattern is that both loci have apparent movement over time, but the raw ALT frequencies are misleading. Locus A starts at 36/40 ALT reads (ALT frequency 0.900) and ends at 82/250 ALT reads (ALT frequency 0.328). If ALT were incorrectly treated as derived, that trajectory would look like declining selection. After orientation, the same rows are 4/40 derived reads (0.100) and 168/250 derived reads (0.672). Locus B also moves, from 10/40 derived reads (0.250) to 105/250 (0.420), but with weaker signal.

```
info = read_tsv("variant_info.tsv.gz")
derived_is_alt = info.outgroup_allele == info.ref
derived_reads = alt_reads if derived_is_alt else total_reads - alt_reads
derived_freq = derived_reads / total_reads
```

Figure 2 is used at this first decision point: it makes clear why the visible ALT trajectory cannot be read as the selected derived-allele trajectory without outgroup polarization.

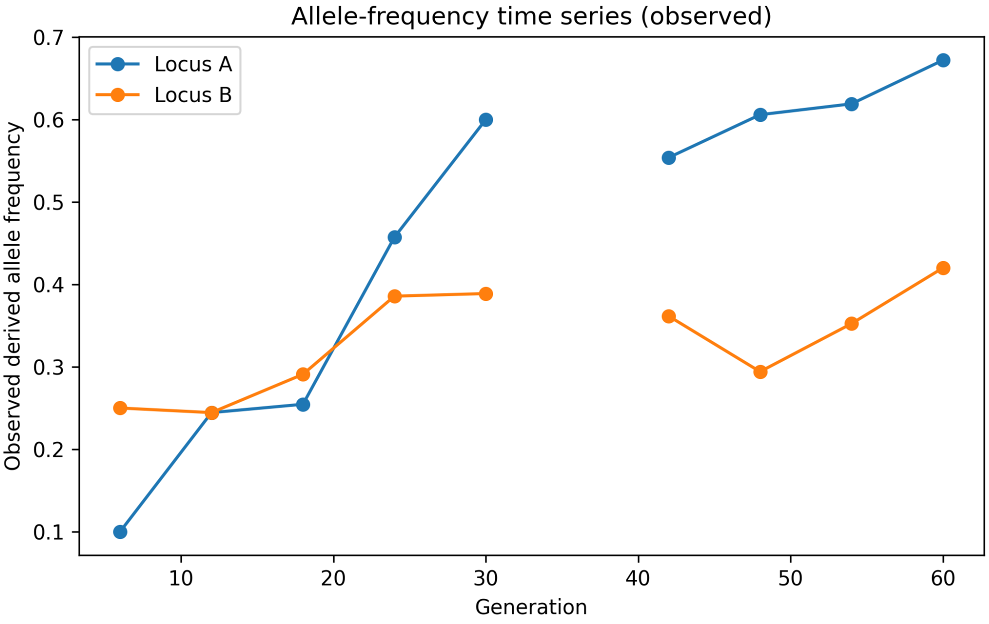

Figure 2: Observed derived-frequency trajectories after outgroup orientation. Blue circles and line show locus A; orange circles and line show locus B. The y-axis is the observed derived-allele frequency after converting released ALT read counts to derived REF read counts because the outgroup allele equals ALT. Circles mark sampled generations with nonzero read depth, and the lines connect adjacent nonmissing observations; the zero-depth sample at generation 36 contributes no point.

Stage 1: use the outgroup to resolve allele orientation

The orientation decision is data-discoverable from `variant_info.tsv.gz`: both rows have REF=C, ALT=T, and outgroup=T. Outgroup-based polarization is a standard way to distinguish ancestral from derived alleles when the outgroup state is reliable [6]. Here the outgroup matches ALT, so ALT is ancestral and REF is derived. The incorrect ALT-as-derived shortcut gives $\hat{s}_A = -0.046168$, $\hat{s}_B = -0.010266$, picks locus B, and fails exact matching on `selected_locus`. The corrected orientation changes the endpoint for locus A from apparent ALT frequency 0.328 to derived frequency 0.672.

```
if outgroup_allele == alt:
    derived_reads = total_reads - alt_reads
else:
    derived_reads = alt_reads
```

Stage 2: recognize that the reported error is an average, not a symmetric rate

The visible samples are ancient: sample years run from -4500 to -2500. This should be a strong diagnostic clue, not a decorative metadata column. The prompt says instrument-driven sequencing error is about 1% and states that `seq_error` is the average of two directional allele-miscall rates, while the visible `seq_error` column is 0.16 at every time point. A sixteen-percent average is far too large to interpret as ordinary instrument noise alone. The natural explanation is that the column summarizes a damage process associated with old DNA molecules rather than a modern sequencing-quality parameter.

The allele identity makes the direction of that damage consequential. Both loci are C/T variants. In ancient-DNA data, post-mortem cytosine deamination is a canonical source of excess C→T observations [4,5]. After Stage 1, C is the derived allele and T is the ancestral allele. Therefore, the age plus C/T context says that damaged derived molecules will often be read as ancestral molecules. This is not symmetric noise around the true allele frequency. It specifically pushes observed derived-read counts downward, flattening the apparent rise of the selected allele and attenuating $\hat{s}$ if the analyst treats the error as two-way exchangeable noise.

The point is easiest to see by contrasting the two interpretations. A symmetric-error model says a true derived read is miscalled ancestral with probability 0.16 and a true ancestral read is miscalled derived with the same probability 0.16. That ignores both the prompt's 1% instrument-error clue and the C→T age-damage clue. The damage-aware interpretation instead treats the reported value as an average of a high C→T direction and a low reverse direction. Because derived=REF=C and ancestral=ALT=T in this case study, that means a high derived-to-ancestral rate and a low ancestral-to-derived rate.

Treating the visible 0.16 as a two-way symmetric error gives a plausible hidden-state analysis but an incorrect estimate: $\hat{s}_A = 0.063559$, $\hat{s}_B = 0.028676$, final $\hat{s} = 0.063559$, outside the 0.020 tolerance. The analyst can instead estimate the directional split from the visible information. The prompt gives the low-direction instrument-error floor, approximately 0.01; the files give the average error, $\bar{e} = 0.16$; and the ancient C/T context gives the high-error direction. Since the reported value is the average of the two directional rates, the resulting reconstruction is

$$e_{\text{da}} = 2\bar{e} - 0.01 = 0.31, \qquad e_{\text{ad}} = 0.01.$$

Here e_{da} is derived-to-ancestral because the derived allele is C and the ancestral allele is T.

```
avg_error = mean(seq_error)       # 0.16 in both time-series files
base_error = 0.01                 # prompt-level instrument floor
e_da = 2 * avg_error - base_error
e_ad = base_error
```

Figure 3 is the compact diagnostic for Stage 2 rather than a separate QC screen: the coverage panel shows which samples can contribute emissions, and the damage panel shows the directional error split that changes the fitted selection coefficient.

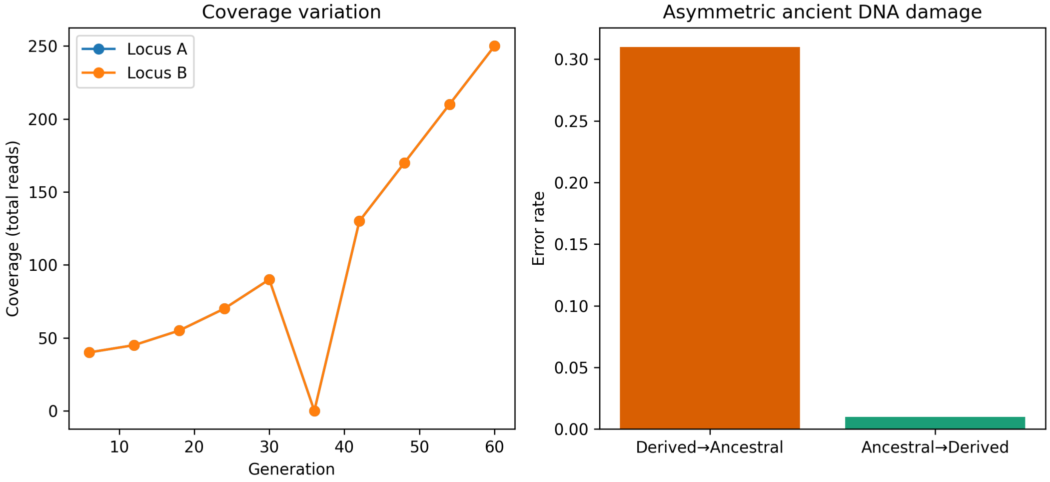

Figure 3: Coverage and damage diagnostics. In the left panel, blue and orange encode loci A and B as in the trajectory figure; the curves overlap because both loci use the same read-depth schedule, including the zero-depth observation at generation 36. In the right panel, the orange bar is the derived-to-ancestral C→T damage rate $e_{\mathrm{da}} = 0.31$ and the green bar is the reverse ancestral-to-derived instrument-error floor $e_{\mathrm{ad}} = 0.01$. These rates decompose the visible average error as $(0.31 + 0.01)/2 = 0.16$ after orientation makes derived=REF=C and ancestral=ALT=T.

Stage 3: fit a Wright–Fisher hidden-state likelihood

A Wright–Fisher model describes allele-frequency evolution by selection followed by random genetic drift. A hidden Markov model is appropriate because the true population count is unobserved at each generation; the analyst sees only noisy read counts at selected generations. The forward algorithm sums over all possible hidden counts rather than committing to a single reconstructed trajectory [1–3]. Simple shortcuts fail: terminal-frequency reporting gives $s = 0.672000$, end-to-start delta gives $s = 0.010593$, linear frequency trend gives $s = 0.009804$, and logit regression gives $s = 0.046168$.

```
for s in grid_or_optimizer:
    alpha = initialize_hidden_count_prior()
    for generation in 6..60:
        alpha = propagate_wf_transition(alpha, N_g, s)
        if reads_observed:
            alpha *= binomial_emission(derived_reads, e_da, e_ad)
```

Figure 4 visualizes the answer-changing likelihood contrast in Stage 3: after orientation, the asymmetric-error model places high support near the released answer, whereas the symmetric-average-error model is lower and centered toward smaller values of s.

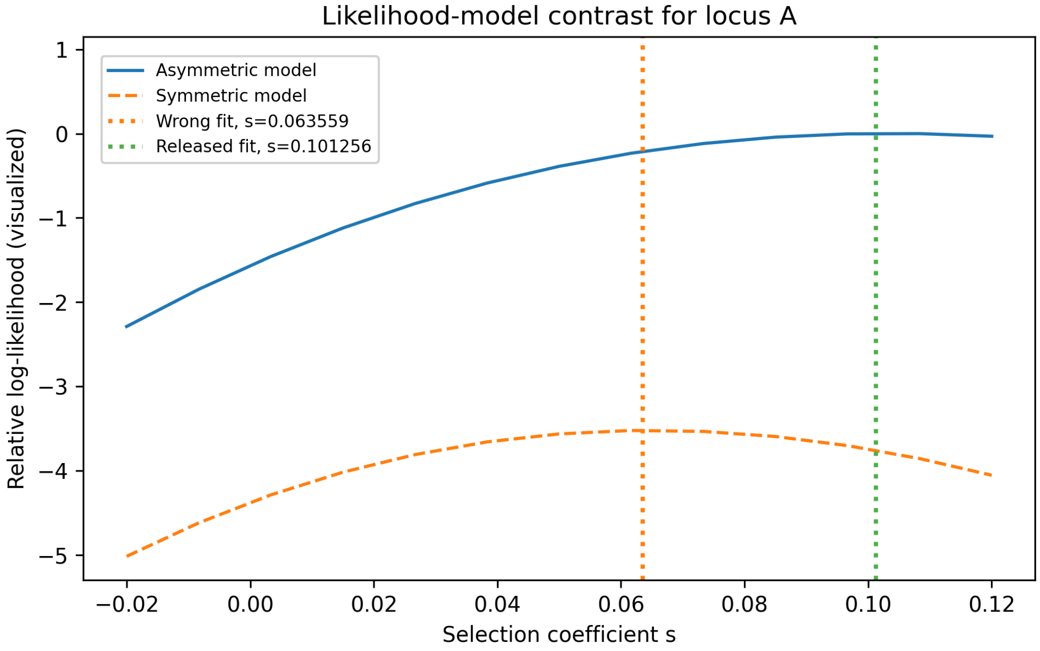

Figure 4: Likelihood-model contrast for locus A. The x-axis is the haploid selection coefficient s, and the y-axis is relative log-likelihood for a downsampled visualization of the likelihood shape. The blue solid curve is the asymmetric ancient-DNA-error model and the orange dashed curve is the symmetric-average-error model; these curves are qualitative diagnostics, not the source of the reported estimate. The orange dotted vertical line marks the wrong symmetric-average fit, $s = 0.063559$, and the green dotted vertical line marks the released full forward-algorithm fit, $s = 0.101256$.

Accounting for the final answer. The reference implementation applies the same orientation and asymmetric-emission rules to both loci. It obtains

$$\hat{s}_A = 0.101255, \qquad \hat{s}_B = 0.024774.$$

The released answer field `selected_locus` is therefore `A`, and the released `s` value is $\max(\hat{s}_A, \hat{s}_B) = 0.101256$ after rounding. A mild reverse-direction baseline sensitivity with $e_{\mathrm{ad}} = 0.02$ gives $\hat{s}_A = 0.096888$ and still passes, while incorrect damage rates 0.25, 0.35, and 0.40 fail in opposite directions.

8 Estimand

The target is the finite-sample maximum-likelihood selection coefficient for the locus with the larger fitted positive selection coefficient under the oriented, damage-aware haploid Wright–Fisher model:

$$\theta^* = \max_{\ell \in \{A,B\}} \hat{s}_\ell, \tag{11}$$

$$\ell^* = \arg \max_{\ell \in \{A,B\}} \hat{s}_\ell. \tag{12}$$

This is not the raw simulator parameter by definition; it is the estimator output recoverable from the released data. In this build the MLE for locus A is close to the generating value 0.10, so the generative parameter and finite-sample target are aligned. The estimand differs from a descriptive slope, an endpoint frequency change, or a symmetric-error hidden-state fit.

9 Estimator

Plain-language summary: for each candidate value of s, the estimator asks how likely the observed derived-read counts are if the population evolved generation by generation under haploid selection and drift, then produced reads with asymmetric ancient-DNA error. Because the true population allele count at each generation is not observed, the estimator keeps a probability distribution over all possible hidden counts and updates it forward through time. The value of s with highest likelihood is the maximum-likelihood estimate.

For observed derived count $K_{\ell,j}$ at generation t_j, hidden count x, and population size N_{t_j}, define the latent derived frequency $q = x/N_{t_j}$. The corrected emission probability is

$$\Pr(K_{\ell,j} = k \mid X_{\ell,t_j} = x) = \mathrm{Binomial}(k; R_j, q(1 - e_{\mathrm{da}}) + (1 - q)e_{\mathrm{ad}})\,. \tag{13}$$

The transition from generation g to $g + 1$ is

$$\Pr(X_{\ell,g+1} = y \mid X_{\ell,g} = x, s) = \mathrm{Binomial}\left(y; N_{g+1}, \frac{(x/N_g)(1 + s)}{1 + (x/N_g)s}\right). \tag{14}$$

Let $\alpha_g(x)$ be the forward probability of the data observed up to generation g and hidden count x. At the first observed generation, the implementation uses an uninformative uniform prior over hidden counts, $\alpha_{t_1}(x) \propto 1$ before applying the first emission. This avoids injecting an unobserved founding frequency into a problem whose released files start at generation 6. At non-observed generations the recursion applies only the transition; at observed generations it applies transition and emission:

$$\alpha_{g+1}(y) \propto E_{g+1}(y) \sum_x \alpha_g(x) T_g(x, y; s), \tag{15}$$

where T_g is Equation (14) and E_{g+1} is Equation (13), or 1 when no reads are observed. The implementation uses log-space normalization with `logsumexp` for numerical stability [3,8].

For each locus,

$$L_\ell(s) = \log \Pr(\{K_{\ell,j}\}_j \mid s), \tag{16}$$

$$\hat{s}_\ell = \arg \max_{s \in [-0.5, 0.5]} L_\ell(s). \tag{17}$$

The current implementation optimizes the one-dimensional objective with SciPy's bounded scalar minimizer applied to $-L_\ell(s)$, with `xatol` set to 10^{-6} [7,9].

10 Decision-Point and Ablation Walkthrough

The table below combines the full ablation set with the stage at which each analysis fails. The released answer contract requires exact recovery of the selected locus and selection coefficient error no larger than 0.020.

Decision point	Analysis / ablation	Quantitative output	Pass?	Failure point	Why the approach is wrong
Allele polarization	Wrong polarization (ALT as derived)	locus B; $s_A = -0.046168$, $s_B = -0.010266$, reported -0.010266; error 0.111522	no	Stage 1	ALT is ancestral, so treating ALT as derived reverses the frequency-change sign.
Shortcut trend	Terminal-frequency shortcut	locus A; reported $s = 0.672000$; error 0.570744	no	Stage 3	Terminal frequency is not a Wright–Fisher selection coefficient.
Shortcut trend	End-to-start delta shortcut	locus A; reported $s = 0.010593$; error 0.090663	no	Stage 3	Raw frequency change ignores drift, sampling noise, and changing population size.
Shortcut trend	Linear frequency trend	locus A; reported $s = 0.009804$; error 0.091452	no	Stage 3	Linear slope is descriptive and not a generative likelihood estimate.
Shortcut trend	Logit regression ignoring error	locus A; reported $s = 0.046168$; error 0.055088	no	Stage 3	Treats damaged read counts as true allele frequencies.
Read-error model	HMM with no error correction	locus A; reported $s = 0.043995$; error 0.057261	no	Stage 2	Omits directional C-to-T damage and attenuates the latent derived trajectory.
Read-error model	HMM with symmetric average error	locus A; reported $s = 0.063559$; error 0.037697	no	Stage 2	Uses the visible average error as symmetric error even though the C/T damage process is directional.
Compound error	Wrong orientation + no-error HMM	locus B; reported $s = -0.019205$; error 0.120461	no	Stages 1–2	Combines reversed allele polarization with no read-error correction.
Compound error	Wrong orientation + symmetric HMM	locus B; reported $s = -0.027877$; error 0.129133	no	Stages 1–2	Combines reversed allele polarization with a symmetric-damage emission model.
Damage calibration	HMM with $e_{\mathrm{da}} = 0.25$	locus A; reported $s = 0.076408$; error 0.024848	no	Stage 2	Under-corrects directional damage relative to the visible average-error contract.
Damage calibration	HMM with $e_{\mathrm{da}} = 0.35$	locus A; reported $s = 0.131365$; error 0.030109	no	Stage 2	Over-corrects the directional damage rate and overshoots the likelihood peak.
Damage calibration	HMM with $e_{\mathrm{da}} = 0.40$	locus A; reported $s = 0.170298$; error 0.069042	no	Stage 2	Strongly over-corrects damage and creates an inflated selection estimate.

Table 2: Unified decision-point and ablation walkthrough for the Wright–Fisher selection problem.

Reference and target-equivalent sensitivity. The reference analysis and one sensitivity analysis preserve the released selected locus and remain inside the s tolerance. They are separated from the failure-mode table so that the landscape page carries the main diagnostic comparison without an awkward trailing fragment.

Analysis	Quantitative output	Interpretation
Correct asymmetric HMM	locus A; $s_A = 0.101255$, $s_B = 0.024774$, reported 0.101255; error 0.000001	Reference polarized, asymmetric-emission Wright–Fisher HMM.
Robustness check, $e_{\mathrm{ad}} = 0.02$	locus A; reported $s = 0.096888$; error 0.004368	Directional damage and locus choice are preserved inside the public tolerance.

11 References

[1] J. P. Bollback, T. L. York, and R. Nielsen. *Estimation of 2Nes From Temporal Allele Frequency Data*. Genetics 179(1):497–502, 2008. DOI: `https://doi.org/10.1534/genetics.107.085019`.

[2] A.-S. Malaspinas, O. Malaspinas, S. N. Evans, and M. Slatkin. *Estimating Allele Age and Selection Coefficient from Time-Serial Data*. Genetics 192(2):599–607, 2012. DOI: `https://doi.org/10.1534/genetics.112.140939`.

[3] L. R. Rabiner. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE 77(2):257–286, 1989. DOI: `https://doi.org/10.1109/5.18626`.

[4] A. W. Briggs et al. *Patterns of damage in genomic DNA sequences from a Neandertal*. Proceedings of the National Academy of Sciences 104(37):14616–14621, 2007. DOI: `https://doi.org/10.1073/pnas.0704665104`. PubMed: `https://pubmed.ncbi.nlm.nih.gov/17715061/`.

[5] H. Jonsson et al. *mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters*. Bioinformatics 29(13):1682–1684, 2013. DOI: `https://doi.org/10.1093/bioinformatics/btt193`. URL: `https://academic.oup.com/bioinformatics/article/29/13/1682/184965`.

[6] P. D. Keightley and B. C. Jackson. *Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site*. Genetics 209(3):897–906, 2018. DOI: `https://doi.org/10.1534/genetics.118.301120`.

[7] P. Virtanen et al. *SciPy 1.0: fundamental algorithms for scientific computing in Python*. Nature Methods 17:261–272, 2020. DOI: `https://doi.org/10.1038/s41592-019-0686-2`.

[8] SciPy documentation. *scipy.special.logsumexp*. URL: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.special.logsumexp.html`.

[9] SciPy documentation. *minimize_scalar(method='bounded')*. URL: `https://docs.scipy.org/doc/scipy/reference/optimize.minimize_scalar-bounded.html`.