

Validation of the `pedtrans` Chromosome Transmission Simulator: Kinship Coefficient Tests

Bruce Rannala

January 6, 2026

Abstract

We describe a validation test for the `pedtrans` pedigree chromosome transmission simulator. The test verifies that simulated identity-by-descent (IBD) sharing between relatives matches theoretical expectations for kinship coefficients. We compare observed kinship coefficients from 1000 simulation replicates against expected values for siblings (0.25), parent-child pairs (0.25), and aunt/uncle-niece/nephew pairs (0.125). All tests confirm that `pedtrans` correctly simulates Mendelian chromosome transmission with recombination.

1 Introduction

The `pedtrans` program simulates chromosome transmission through a pedigree, tracking which genomic segments in each individual trace back to which founder chromosomes. This document describes validation tests that verify the simulator produces correct patterns of identity-by-descent (IBD) sharing between relatives.

1.1 Kinship Coefficient

The kinship coefficient $\phi(A, B)$ between two individuals A and B is defined as the probability that a randomly drawn allele from A is identical by descent (IBD) to a randomly drawn allele from B :

$$\phi(A, B) = \Pr(\text{random allele from } A \text{ is IBD to random allele from } B) \quad (1)$$

For diploid individuals, this is equivalent to:

$$\phi(A, B) = \mathbb{E} \left[\frac{\text{number of IBD allele pairs}}{4} \right] \quad (2)$$

where there are four possible pairings between the two alleles of A and the two alleles of B at any locus.

1.2 Expected Kinship Coefficients

For a pedigree without inbreeding, the theoretical kinship coefficients are:

| Relationship | Kinship Coefficient |
|-------------------------|---------------------|
| Parent–child | 0.25 |
| Full siblings | 0.25 |
| Half siblings | 0.125 |
| Grandparent–grandchild | 0.125 |
| Aunt/Uncle–Niece/Nephew | 0.125 |
| First cousins | 0.0625 |

2 Calculation Method

2.1 Simulator Output

The `pedtrans` simulator outputs, for each individual, a list of genomic segments with their founder chromosome origins. Each segment is defined by:

- Start position $s \in [0, 1)$
- End position $e \in (s, 1]$
- Founder origin: (founder ID, homolog), where homolog $\in \{0, 1\}$ indicates paternal or maternal chromosome of the founder

For example, an individual might have:

Paternal:

```
0.000000 0.456789 Father:pat
0.456789 1.000000 Father:mat
```

Maternal:

```
0.000000 1.000000 Mother:pat
```

2.2 IBD Determination

Two alleles are identical by descent if and only if they originate from the *same founder chromosome*—that is, both the founder identity and the homolog (paternal vs. maternal) must match.

2.3 Algorithm

Given the segment data for two individuals A and B , we calculate the kinship coefficient as follows:

Algorithm 1 Kinship Coefficient Calculation

```
1: Collect all segment breakpoints from both individuals
2: Sort breakpoints to partition [0, 1] into intervals
3:  $\phi \leftarrow 0$ 
4: for each interval  $[p_i, p_{i+1})$  do
5:    $\ell \leftarrow p_{i+1} - p_i$                                  $\triangleright$  Interval length
6:   Determine founder origins at midpoint for both individuals:
7:      $A_{\text{pat}}, A_{\text{mat}}$  for individual  $A$ 
8:      $B_{\text{pat}}, B_{\text{mat}}$  for individual  $B$ 
9:   Count IBD pairs among the four pairings:
10:     $n_{\text{IBD}} \leftarrow \mathbf{1}[A_{\text{pat}} = B_{\text{pat}}] + \mathbf{1}[A_{\text{pat}} = B_{\text{mat}}]$ 
11:     $+ \mathbf{1}[A_{\text{mat}} = B_{\text{pat}}] + \mathbf{1}[A_{\text{mat}} = B_{\text{mat}}]$ 
12:     $\phi \leftarrow \phi + \frac{n_{\text{IBD}}}{4} \cdot \ell$ 
13: end for
14: return  $\phi$ 
```

Here, $\mathbf{1}[\cdot]$ is the indicator function, and two origins are equal if they have both the same founder ID and the same homolog.

2.4 Example: Sibling Comparison

Consider two siblings who inherited:

- Sibling 1: paternal = Father:pat, maternal = Mother:mat
- Sibling 2: paternal = Father:pat, maternal = Mother:pat

The four allele pairings at any position are:

| Pair | Sib1 Allele | Sib2 Allele | IBD? |
|------|-------------|-------------|--------------------------------------|
| 1 | Father:pat | Father:pat | Yes |
| 2 | Father:pat | Mother:pat | No (different founder) |
| 3 | Mother:mat | Father:pat | No (different founder) |
| 4 | Mother:mat | Mother:pat | No (same founder, different homolog) |

In this realization, $n_{\text{IBD}} = 1$, giving kinship = $1/4 = 0.25$.

3 Test Design

3.1 Test Pedigrees

We constructed two test pedigrees:

Simple Pedigree. Two founders (Father, Mother) and two sibling offspring (Sib1, Sib2).

Extended Pedigree. A four-generation pedigree with 8 founders enabling tests of multiple relationship types:

- Generation 0: 8 founders (4 unrelated couples)
- Generation 1: 4 individuals (children of founder couples)
- Generation 2: 4 individuals (two pairs of siblings)
- Generation 3: 4 individuals (offspring)

3.2 Simulation Parameters

- Recombination rate: 1.0 (expected crossovers per meiosis)
- Number of replicates: 1000
- Random seeds: Sequential (1 through 1000) for reproducibility

3.3 Relationships Tested

1. **Siblings:** Individuals sharing both parents (expected $\phi = 0.25$)
2. **Parent–child:** Direct parent-offspring pairs (expected $\phi = 0.25$)
3. **Aunt/Uncle–Niece/Nephew:** Second-degree relatives (expected $\phi = 0.125$)

4 Results

4.1 Simple Pedigree: Sibling Test

| Relationship | Mean $\hat{\phi}$ | Std Dev | Expected |
|----------------------|-------------------|---------|----------|
| Siblings (Sib1–Sib2) | 0.2472 | 0.1080 | 0.25 |

4.2 Extended Pedigree: Multiple Relationships

| Relationship | Mean $\hat{\phi}$ | Std Dev | Expected | Status |
|-------------------------|-------------------|---------|----------|--------|
| Siblings (G2) | 0.2535 | 0.1078 | 0.25 | OK |
| Siblings (G3) | 0.2510 | 0.1085 | 0.25 | OK |
| Parent–Child | 0.2500 | 0.0000 | 0.25 | OK |
| Aunt/Uncle–Niece/Nephew | 0.1288 | 0.0720 | 0.125 | OK |

4.3 Observations

1. **Sibling kinship:** The observed mean of ≈ 0.25 matches the theoretical expectation. The standard deviation of ≈ 0.11 reflects the variance in IBD sharing due to random segregation and recombination.
2. **Parent–child kinship:** The observed value is exactly 0.25 with zero variance. This is expected because a child always inherits exactly one allele from each parent, making the IBD count deterministic (always 1 out of 4 pairings involves the transmitted allele matching itself).

3. **Aunt/Uncle–Niece/Nephew:** The observed mean of ≈ 0.129 closely matches the expected 0.125. The higher variance compared to siblings reflects the additional meiosis separating these relatives.

5 Statistical Validation

For $n = 1000$ replicates, the standard error of the mean is:

$$\text{SE} = \frac{\sigma}{\sqrt{n}} \quad (3)$$

For siblings with $\sigma \approx 0.108$:

$$\text{SE} = \frac{0.108}{\sqrt{1000}} \approx 0.0034 \quad (4)$$

The observed mean of 0.2472 differs from the expected 0.25 by approximately 0.8 standard errors, well within the range expected from sampling variation.

6 Conclusion

The `pedtrans` chromosome transmission simulator produces kinship coefficients that match theoretical expectations for all tested relationship types:

- Siblings: $\hat{\phi} = 0.25$ (expected 0.25)
- Parent–child: $\hat{\phi} = 0.25$ (expected 0.25)
- Aunt/Uncle–Niece/Nephew: $\hat{\phi} = 0.125$ (expected 0.125)

These results validate that the simulator correctly implements:

1. Mendelian segregation (random selection of parental chromosomes)
2. Recombination (crossover events during meiosis)
3. Proper tracking of founder chromosome identity through the pedigree

The test script is available at `testing/test_pedtrans_sharing.py`.