# Supplementary materials

## Appendix A. Wright’s standardized variance estimator

We will use the symbol to denote an allele copy in the first two appendices.

For the situation of a single subpopulation (say ), the BP is . We use the binary variable to denote the state of an allele copy in individual , whose value is if , or if , where is in . Then and .

The kinship coefficient is the probability of randomly sampling two IBD allele copies with replacement (say and ) from individual with replacement relative to . Then is equal to one at the probability if and are IBD. If and are not IBD, they are independent, and so is equal to one at the probability . Hence

.

Similarly, by the definition of the kinship coefficient , we have

,

where and are randomly sampled from . Therefore

Also,

.

Since the operator is linear, we obtain

,

.

Therefore, by and , it follows

For the situation of multiple subpopulations, the BP is the total population. We will use the binary variable to denote the state of any allele copy , where , and satisfy the nested relationship, i.e., and . Then and .

By the definition of , we have

,

,

.

By the same derivation as above, the following equations hold:

(A3)

,

(A4)

Also,

,

(A5)

Note that , it follows

,

.

Hence

The same way can be used to derive the following expressions:

## Appendix B. Nei’s heterozygosity estimator

We use the same binary variables as Appendix A. For simplicity, we omit the identifier in the subscripts of , , and .

For the situation of a single subpopulation , the BP is . Because the probability that two distinct allele copies in are IBD is , then

Noticing that , we have , then

Also, by Equation (A2), then

.

Now, by the definitions of and , we obtain

So and by . Hence

For the situation of multiple subpopulations, the BP is . Then

Because of , we have by Equation (A5). Also, we also have by Equation (A4). Now, by the same derivation as the first half, we obtain

,

,

,

where is a representative in . Therefore

## Appendix C. Li’s correlation estimator

Unlike the first two appendices, we here denote for the ordinal number of an allele instead of the allele itself. For example, is the allele copy within individual at a target locus.

In a single subpopulation (say), the BP is . Also, the estimate of Li & Horvitz’s (1953) estimator can be expressed as

where and are two random variables, whose possible values are the assignments of all possible allele copies in individual at a target locus, and each pair of their values is the assignments of allele copies and in which is randomly sampled from subpopulation , and and are randomly sampled from 1 to without replacement. Let be the assignment of (the value assigned can be arbitrary). Then the above expression can be rewritten as

Let and be two allele copies randomly sampled from individual at this target locus with replacement, and let and . Then if and are IBD, otherwise . Especially, . So, , i.e.,

.

Hence .

Similarly, , and so

As the derivation in Appendix B, we have . Because the operator is linear, by and , we obtain

.

Also,

Then, by Equation (2), it follows

.

So, Moreover, by . Hence

## Appendix D. Weir & Cockerham’s estimator

We perform our derivation based on Huang *et al*.’s (2019a) generalized AMOVA framework, and choose the IAM distance as an example to define various sums of squares (SS). In this framework, the derivation of the expected value of the within each hierarchy is clearer. The symbol ( or ) is simply called the *sum* *of* *squared genetic distances* between alleles within individuals, subpopulations or the total population.

For a single subpopulation (say ), the BP is and there are only the two squared sums and , which is calculated by

where and are the ordinal numbers of two allele copies, is the genetic distance between two distinct allele copies and within individual at a target locus, whose definition in the infinity allele model (IAM) is as follows: if and are IBD, otherwise .

For independent samples, the probability that and are IBD is , and the expectation of is ; the probability that and are IBD is 0. So, by the definition of IAM distance, the following hold:

,

.

Since and , we have

and .

Then ,

=

Now, replacing by the , by the , by , and by , we obtain two relational expressions as follows:

(A6)

Noticing that the estimates and are unbiased for independent samples.

For non-independent samples, the probability that and are IBD is also , and thus . However, the probability that and are IBD is instead of zero, and thus

.

Denote for , then . Following the above method, we can derive that

and .

On the other hand, by Equation (A6), the following hold:

Then ,

So and . Because , we obtain

For the situation of multiple subpopulations, the BP for each unit or each pair of units is the total population, and the three squared sums are defined as follows:

For this situation, the probability that and are IBD is still , whose expectation relative to the BP is ; the probability that and are IBD is , and if the samples are independent; and the probability that and are IBD is zero if the samples are independent, or is if they are non-independent.

Now, by the definitions of , and , for the independent samples, we have

,

,

.

Then , and

,

.

So, as the above derivation, the following three relational expressions are established:

(A7)

Next, for the dependent samples, if we let

and ,

then ,

,

.

By the same derivation as above, one has

,

,

.

On the other hand, by Equation (A7), one obtains

,

.

Comparing the corresponding values of at each hierarchy, one gets

, and .

Noticing that , it follows

,

,

.

## Appendix E. Genotypic frequencies under inbreeding

Huang et al. (2019b) derived the genotypic frequency at equilibrium state for polysomic inheritance under double-reduction and Huang *et al*. (2019a) use a Dirichlet distribution to approximate the genotypic frequencies under the conditions of inbreeding. The derivation are shown in Appendix B of Huang *et al*. (2019a). Here, we only present the resulting equation

where denotes the genotype, denotes the vector consisting of the frequencies of alleles in a unit, denotes the inbreeding coefficient in a unit, is the number of copies of the allele in , and (Pritchard *et al.* 2000).

## Appendix F. Conditional distribution of genotypes

In this appendix, we focus on a *band*, which is defined as a virtual hierarchy inferior to the background population, denoted by ­. The symbols and have the same meanings as those in Appendix E, and we will also use to denote the vector consisting of the frequencies of alleles in band ­. In order to facilitate the derivation, we use to represent the background population. Meanwhile, the kinship coefficient relative to is denoted by regardless of the value of .

To generate the inbred offspring, the expected kinship coefficient between the father and the mother must be equal to the inbreeding coefficient , or equivalently equal to . Coincidentally, distinct alleles within a parent or between parents are all IBD at a probability of , therefore and can be considered as drawn from a band without inbreeding ( and ).

Since , the true allele frequencies in each individual ( and ) are equal to and the alleles in - can be considered as independent relative to . Also, by , one has . Therefore, it can be inferred that are drawn from the Dirichlet distribution , where and (Pritchard *et al.* 2000). Now, by using the results given in Appendix E, the expression of a joint probability mass function of a father’s genotype and a mother’s genotype conditional on and can be derived as follows:

where is the probability density function of and the integral domain can be expressed as

.

Therefore, the conditional distribution function of a mother’s genotype given a father’s genotype conditional on and can be expressed as

Similarly, the conditional distribution function of father’s genotypes can be expressed as

## Appendix G. Calculation of the kinship coefficient from pedigree

Karigl (1981) developed an iterative algorithm to calculate the kinship coefficients and from pedigree. Huang *et al*. (2015) generalized this algorithm into the situation of polysomic inheritance, and the generalized calculating formulas are as follows:

where represents the kinship coefficient at the background of a subpopulation, and are distinct individuals, is the father of , and is the mother of .

For natural individuals without pedigree information (i.e., the father and the mother are unknown), the expected kinship coefficient is within individuals, between mates, or zero between other dyads in the parental generation.

Let denote the kinship coefficient between the full-sibs -, and let denote the kinship coefficient between the half-sibs - (same father but different mothers). If and are without any pedigree information, the expectation can be expressed as

;

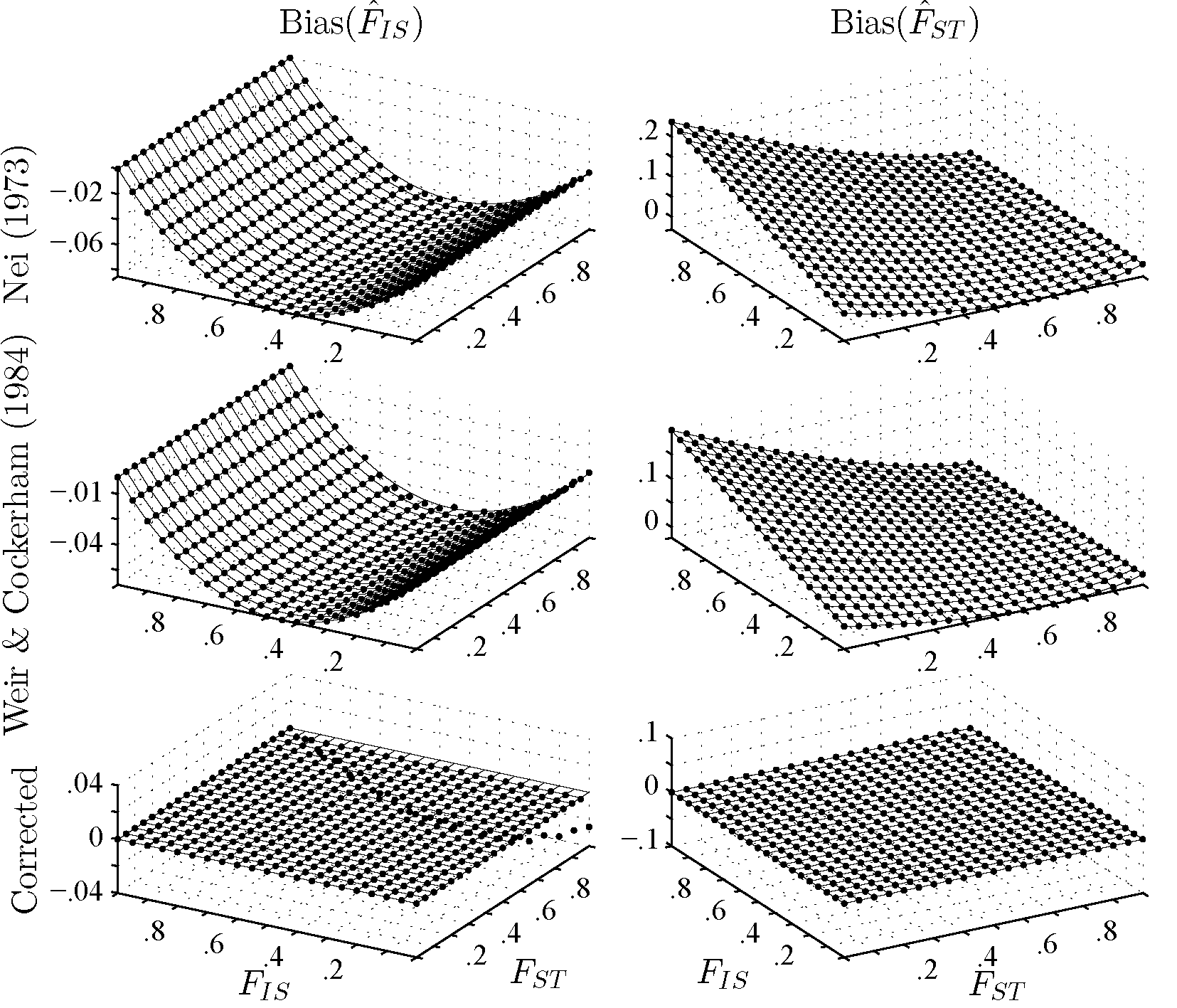
Besides, the kinship between the two mothers and of half-sibs is identified by the following probability:

The value of this probability is calculated by using Equations (A8-A10). We derived that is equal to , therefore the expectation can be expressed as

.

Therefore and . (A11)

## Supplemental Figures



**Figure S1.** The bias of and as a binary function of and of Nei’s (1973), Weir & Cockerham’s (1984) and our corrected estimators for autotetraploids. The mesh plots denote the theoretical biases and the black dots denote the simulated biases.

# References

Huang K, Guo ST, Shattuck MR*, et al.* (2015) A maximum-likelihood estimation of pairwise relatedness for autopolyploids. *Heredity* **114**, 133-142.

Huang K, Li Y, Dunn DW, Zhang P, Li B (2019a) A generalized framework of AMOVA with any number of hierarchies and any level of ploidies. *bioRxiv*, doi: 10.1101/608117.

Huang K, Wang TC, Dunn DW*, et al.* (2019b) Genotypic frequencies at equilibrium for polysomic inheritance under double-reduction. *G3: Genes|Genomes|Genetics* **9**, 1693-1706.

Karigl G (1981) A recursive algorithm for the calculation of identity coefficients. *Annals of Human Genetics* **45**, 299-305.

Li CC, Horvitz DG (1953) Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics* **5**, 107.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3323.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Weir BS, Cockerham CC (1984) Estimating *F*‐statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.