1

2 **Measuring drift by mean deviation: unequal breeding sex ratio revisited**

3

4 Ben Qin^{1,2,*}

5

6 ¹ Guangxi Key Laboratory of Medicinal Resources Protection and Genetic
7 Improvement, Guangxi Botanical Garden of Medicinal Plants, Nanning 530023,
8 China

9 ² Guangxi Engineering Research Center of TCM Resource Intelligent Creation,
10 Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, China

11

12

13 Measuring drift by mean deviation

14 Keywords: genetic drift, sex ratio, effective population size, mean deviation,

15 coalescent theory

16

17 * Corresponding Author:

18 Ben Qin

19 No. 189 Changgang Road,

20 Guangxi Botanical Garden of Medicinal Plants,

21 Nanning 530023, China

22 Tel: +8615307815487

23 Email: qinben@gxyyzwy.com

24

25 **Abstract**

26 When it comes to estimating the magnitude of genetic drift, there is hardly any
 27 indexes other than the effective population size. Starting from the binomial sampling
 28 distribution, this research proposed using mean deviation of allele frequency change
 29 as a direct measurement of drift, then tested it in a classical example concerning
 30 unequal breeding sex ratio. This study found that: (1) mean deviation offers a new
 31 dimension in measuring the magnitude of drift; (2) the measurement displays a “half-
 32 half” pattern; (3) allele frequency determines the efficacy of hitchhiking effect of rare
 33 alleles, and in what way that “half-half” pattern should be divided.

34

35 **Introduction**

36 As one of the fundamental evolutionary forces, genetic drift has been explored using
 37 models like random walk and coalescent process. The former looks forward in time,
 38 measuring drift by following the fluctuation of allele frequency—bigger step size
 39 translates into more significant drift; the latter looks backward in time^[1], measuring
 40 drift by effective population size (N_e)—smaller N_e corresponds to stronger drift^[2, 5].
 41 Another approach, the Wright-Fisher model, treats genetic drift as a binomial
 42 sampling process, and becomes the most common model to study drift^[4]. Started from
 43 binomial sampling, this research incorporated both allele frequency change (i.e., step
 44 size in random walk model) and N_e (key component in coalescent theory), and
 45 proposed a direct and exhaustive method to measure the magnitude of drift.

46 In terms of allele frequency, drift is by definition the change in allele frequency due to
 47 random sampling, and binomial distribution provides the probability of this change,
 48 so the combination of these two factors gives the mean deviation (MD), a direct

49 measurement of drift (see below).

50 In terms of N_e , breeding males and females (N_m and N_f , respectively) in an idealized
51 population are the same, but unequal sex ratio reduces N_e and intensifies genetic drift^[6]
52 by:

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

53 For instance, a population where 1 male breeding with many females would
54 experience genetic drift as strong as an idealized population of 4. And a 2-male
55 population would withstand drift as an idealized population of 8 does, and 3-male as
56 another one of 12 does, etc. Classical coalescent theory treats alleles as identity by
57 descent, so their frequencies are left out from this equation. This research introduced
58 allele frequency into the calculation of the magnitude of drift, for populations of 1, 2,
59 and 3 males breeding with many females, then compared the results with the idealized
60 populations.

61

62 **Methods**

63 **1. Magnitude of drift.**

64 In a neutral model of one locus and two alleles (A_1 , A_2), binomial distribution gives
65 the probability of randomly sampling k A_1 into the next generation (containing $2N$
66 alleles), based on A_1 frequency (p) in the current generation:

$$P_k = \binom{2N}{k} p^k (1 - p)^{2N-k}$$

67 Considering the absolute change in allele frequency is $|k/2N - p|$, then by definition,

the average magnitude of drift can be written as the weighted mean deviation (MD) of allele frequency change:

$$MD = \sum_{k=0}^{2N} \binom{2N}{k} p^k (1-p)^{2N-k} \left| \frac{k}{2N} - p \right|$$

2. MD in extremely female-biased population where only 1 male breeds.

In this case, $N_e \approx 4$, so there are 7 values of interest for p ($p=0/8$ and $p=8/8$ are not meaningful to discuss with respect to drift).

Under no selection and constant population size, when forming the next generation, N gametes of each sex would be sampled. Each of the two alleles from that single male is expected to drift to $N/2$ copies, while each allele from females has only about 1/2 chance to reach the next generation.

For that male individual, the gametes are generated through a process of sampling with replacement. Denote x as the number of A_1 allele(s) in each of the 3 possible genotypes (hereafter categories of A_1 number in breeding males, see below), and m as the numbers of A_1 alleles entering the next generation from that male, then the probability of sampling m A_1 alleles from that male (P_m) can be written as:

$$P_m = \begin{cases} P_{100} = 1, & x = 2 \\ \binom{N}{m} \left(\frac{1}{2}\right)^N, & x = 1 \\ P_0 = 1, & x = 0 \end{cases}$$

The number of A_1 in breeding females is subject to that in breeding males because the total has to sum up to $2Np$, so the sampling process (N from $2N-2$ alleles) is without replacement. Thus, the probability of sampling f A_1 alleles from female (P_f) follows hypergeometric distribution:

$$P_f = \frac{\binom{2Np-x}{f} \binom{(2N-2)-(2Np-x)}{N-f}}{\binom{2N-2}{N}}$$

86 The change in p is:

$$\Delta p = \frac{f+m}{2N} - p$$

87 Then generally, the MD of each genotype can be calculated through:

$$MD = \sum_m \sum_f P_m \cdot P_f \cdot |\Delta p|$$

88 The following example concerns a population of 99 females and 1 male, and the
89 population size remains constant ($N=100$).

90 **2.1 $x=2$ (A_1A_1)**

91 In this category, all the 100 sperms contain A_1 , so $m=100$ and $P_m = P_{100}=1$, thus,

$$MD_{\delta(A_1A_1)} = \sum_f \frac{\binom{200p-2}{f} \binom{198-(200p-2)}{100-f}}{\binom{198}{100}} \cdot \left| \frac{f+100}{200} - p \right|$$

92 **2.2 $x=1$ (A_1A_2)**

93 In this category, the male gamete pool is sampled 100 times and the probability of
94 getting A_1 is $\frac{1}{2}$ for each sample, thus,

$$MD_{\delta(A_1A_2)} = \sum_m \sum_f \binom{100}{m} \left(\frac{1}{2}\right)^{100} \cdot \frac{\binom{200p-1}{f} \binom{198-(200p-1)}{100-f}}{\binom{198}{100}} \cdot \left| \frac{f+m}{200} - p \right|$$

95 **2.3 $x=0$ (A_2A_2)**

96 In this category, none of the 100 sperms contains A_1 , so $m=0$ and $P_m = P_0=1$, thus,

$$MD_{\delta(A_2A_2)} = \sum_f \frac{\binom{200p}{f} \binom{198-200p}{100-f}}{\binom{198}{100}} \cdot \left| \frac{f}{200} - p \right|$$

2.4 Weighted MD

Because the parental population are of fixed size and allele frequencies, determining which two gametes get into that male is also a sampling process without replacement.

Thus, the genotype frequency can be calculated through $\frac{\binom{2Np}{x} \binom{2N-2Np}{2-x}}{\binom{2N}{2}}$, and the MD weighted by genotype frequencies is:

$$MD = \frac{\binom{2Np}{2}}{\binom{2N}{2}} \cdot MD_{\delta(A_1A_1)} + \frac{4N^2p(1-p)}{\binom{2N}{2}} \cdot MD_{\delta(A_1A_2)} + \frac{\binom{2N-2Np}{2}}{\binom{2N}{2}} \cdot MD_{\delta(A_2A_2)}$$

3. More than 1 male breeds.

As long as the number of females is far more than males, it does not affect the essence of the result (data not shown). To avoid interruption during the binomial sampling due to non-integers, this research set the population size to 1000 for the 2-male case, and to 600 for the 3-male case.

3.1 2 males breed with 998 females

In this case $N_e \approx 8$, meaning that there are 15 meaningful values for p . In these 2 males there are 4 slots for alleles to fill, so x (number of A_1 in both males) ranges from 0 to 4, corresponding to 5 categories, then we have:

$$P_m = \begin{cases} P_{1000} = 1, & x = 4 \\ \binom{1000}{m} \left(\frac{x}{4}\right)^m \left(1 - \frac{x}{4}\right)^{1000-m}, & x = 1, 2, 3 \\ P_0 = 1, & x = 0 \end{cases}$$

$$P_f = \frac{\binom{2000p-x}{f} \binom{(2000-4)-(2000p-x)}{1000-f}}{\binom{2000-4}{1000}}$$

111 The weight for each category is:

$$\frac{\binom{2000p}{x} \binom{2000-2000p}{4-x}}{\binom{2000}{4}}$$

112 Other calculations are in the same procedure as the 1-male case.

113 3.2 3 males breed with 597 females

114 In this case $N_e \approx 12$, meaning that there are 23 meaningful values for p . In these 3
115 males there are 6 slots for alleles to fill, so x (number of A_1 in all 3 males) ranges
116 from 0 to 6, corresponding to 7 categories, then we have:

$$P_m = \begin{cases} P_{600} = 1, & x = 6 \\ \binom{600}{m} \left(\frac{x}{6}\right)^m \left(1 - \frac{x}{6}\right)^{600-m}, & x = 1, 2, 3, 4, 5 \\ P_0 = 1, & x = 0 \end{cases}$$

$$P_f = \frac{\binom{1200p-x}{f} \binom{(1200-6)-(1200p-x)}{600-f}}{\binom{1200-6}{600}}$$

117 The weight for each category is:

$$\frac{\binom{1200p}{x} \binom{1200-1200p}{6-x}}{\binom{1200}{6}}$$

118 Other calculations are in the same procedure as the 1-male case.

119 All MD s were calculated with R^[3].

120

121 Results and discussion

122 1. 1-male case.

123 Table 1 shows that, with MD as an index of drift, the two populations of the same N_e

can still experience slightly different degrees of genetic drift, and when $p=1/2$ the two populations are the closest.

Besides, subdividing the *MD* of the female-biased population allows us to zoom in for more detail. Traditionally, this classical case has been used to illustrate how a male's rare allele can drift a big step by luck, but the low probability for having this allele could largely cancel out this effect. Take $p=1/8$ for example, when the male is A_1A_1 , every offspring would carry A_1 and p would have a dramatic increase, translating to a very high *MD* (>0.4). However, since A_1 is a fairly rare allele ($p=1/8$), the chance to have an A_1A_1 male is even lower, so the weighted *MD* is about 0.0065, which does not contribute much to the total *MD* (Fig. 1A).

A common concern about highly biased breeding sex ratio is that deleterious alleles could soar in number, but this is the case only if those alleles can find their place in the breeding male, which is not very likely. Even if the male is selected to breed for traits on certain loci, hitchhiking alleles on other independent neutral loci are sampled randomly based on their frequencies. In other words, biased sex ratio alone is not enough to efficiently boost up a rare allele in number.

2. 2-male and 3-male cases.

Consistent with the 1-male case, two types of population have similar but not the same *MDs*, which are closest to each other at $p=1/2$ (Table 2, 3).

Combining the 1-male case, an interesting pattern emerges: the total *MD* splits in half, each of which is composed of certain categories of A_1 in breeding males, and the composition varies with A_1 frequency. Take the 3-male case as an example: from $p=1/24$ - $3/24$, half of the total *MD* comes from $0A_1$ category, and when $p=4/24$, the weighted *MD* of $0A_1$ category plus $1/2$ of that from $1A_1$ category makes up exactly $1/2$

of the total *MD*. This pattern carries on for all the subsequent frequencies: from $p=5/24-7/24$, the sum of $0A_1$ and $1A_1$ equals $\frac{1}{2}$ of the total *MD*, and when p reaches $8/24$, an extra $\frac{1}{2}$ of $2A_1$'s *MD* is needed to maintain this “half-half” pattern, and same cycle is repeated afterwards (Table 3, Fig. 1C).

The exact pattern is detected in 1-male and 2-male cases (Table 1-2, Fig. 1A-B), and it carries on beyond (e.g., 4-male, data not shown). A chart (Fig. 2) illustrates this pattern for one to refer to: it is the abstract representation of a table, and one can locate the boundary of “half of the drift” for a certain A_1 frequency and number of breeding males.

Previous studies on coalescent theory and N_e have discussed that, random chance experienced by a diploid organism comes from two sources: the first is within individual, acting when an allele gets into the successful gamete during Mendelian Segregation; and the second is between individuals, acting during survival and reproduction. Genetic drift can result from both, and under some peculiar circumstances, these two sources of randomness turn out to be equal^[4].

This research found that the magnitude of drift is divisible in another aspect (category of A_1 number) for populations of extremely biased sex ratio. The division helps in understanding the property of drift: as illustrated in Fig. 2, how to break up the *MD* is a function of the number of breeding males and allele frequency on that locus; when variation of allele frequency is very small (p close to 0 or 1), those two halves of the *MD* are distributed extremely unevenly. This finding could be of especial interest when studying the impact of neutral drift on rare alleles.

Conclusions

When measuring drift in a population with unequal breeding sex ratio, an idealized population and a female-biased population consistently have similar *MDs*; for the latter, *MD* can always be divided into two equal parts, each of which regularly contains certain categories of A_1 number in breeding males; a rare neutral allele can hardly multiply through hitchhiking, because it requires a great deal of luck to end up in the breeding males. In the future, it would be interesting to test this method in another classical case: fluctuating population size over generations. Besides, incorporating selection, migration and directionality in frequency change are also worth of pursuing.

Acknowledgement

I thank Sean Rice for insightful discussion, Meijun Dong for helps in some graphs, and Maosheng Zhang for suggestion in simplifying some equations.

References

- [1] Kingman, J.F.C. (2001). Origins of the Coalescent: 1974-1982. *Genetics*, 156(4), 1461-1463.
- [2] Nomura, T. (2003). Effective size of populations with heritable variation in fitness. *Heredity*. 89, 413–416. 10.1038/sj.hdy.6800169.
- [3] R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [4] Rice, S. (2004). *Evolutionary theory: mathematical and conceptual*

- 195 *foundations*. Sunderland, MA: Sinauer Associates, 73-117.
- 196 [5] Wang, J., Santiago, E., & Caballero, A. (2016). Prediction and estimation of
- 197 effective population size. *Heredity*, 117(4), 193–206.
- 198 <https://doi.org/10.1038/hdy.2016.43>
- 199 [6] Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97-159.

200 **Figure Legends**

201 Fig. 1 Weighted *MDs* at various frequencies of A_1 for 1-male (A), 2-male (B), and 3-
 202 male (C) cases. For B and C, the second half is omitted due to symmetry. *MDs* are
 203 subdivided according to categories of A_1 number in the breeding male(s), and these
 204 categories always form a “half-half” pattern (the pie charts).

205 Fig. 2 Generalization of the “half-half” pattern. The nesting zigzag shape is to show
 206 that half or all of the *MD* from the following allele frequency level are added
 207 periodically.

Fig. 1

Fig. 2

212 **Tables**

213 Table 1. *MDs* of the 1-male case.

<i>MD</i>	<i>A</i> ₁ frequency (<i>p</i>)						
	1/8	2/8	3/8	4/8	5/8	6/8	7/8
idealized (<i>N_e</i> =4)	0.08590223	0.1167984	0.132015	0.1367188	0.132015	0.1167984	0.08590223
♀biased (<i>N_e</i> ≈4)	0.09466843	0.1389715	0.1447774	0.1353594	0.1447774	0.13897142	0.094668416
♂(<i>A</i> ₁ <i>A</i> ₁)	0.4330808	0.3712121	0.3093434	0.2474747	0.1856061	0.1237374	0.06186869
	0.006528856	0.022851	0.04313707	0.06155778	0.07228377	0.06948569	0.04733421
♂(<i>A</i> ₁ <i>A</i> ₂)	0.1856061	0.1237375	0.06231416	0.02436533	0.06231416	0.1237375	0.1856061
	0.04080536	0.0466347	0.02935654	0.01224388	0.02935654	0.04663473	0.04080535
♂(<i>A</i> ₂ <i>A</i> ₂)	0.06186869	0.1237374	0.1856061	0.2474747	0.3093434	0.3712121	0.4330808
	0.04733421	0.06948576	0.07228378	0.06155778	0.04313709	0.022851	0.006528856

214 The *MDs* of female-biased population are subdivided according to the male's genotypes, and weighted values are in bold.

215 Table 2. *MDs* of the 2-male case.

<i>MD</i>	<i>A</i> ₁ frequency (<i>p</i>)							
	1/16	2/16	3/16	4/16	5/16	6/16	7/16	8/16
idealized (<i>N_e</i> =8)	0.04450927	0.06325023	0.07563979	0.08444965	0.09071627	0.09494085	0.09738807	0.09819031
♀biased (<i>N_e</i> ≈8)	0.04817337	0.0730945	0.08149348	0.08168797	0.0949357	0.102715	0.1019123	0.09636719
4 <i>A</i> ₁	6.821037e-06	0.0001043707	0.0004946072	0.001448772	0.003250111	0.006136599	0.01024366	0.01554689
3 <i>A</i> ₁	0.0003075057	0.002112767	0.005983152	0.01166014	0.01828841	0.02464498	0.02936829	0.03118735

216

217

218

MDs subdivided according to categories of A_1 number in breeding males; only shows weighted values and the first half of the A_1 frequency spectrum (the other half is symmetrical).

Table 3. *MDs* of the 3-male case.

<i>MD</i>	A_1 frequency (p)											
	1/24	2/24	3/24	4/24	5/24	6/24	7/24	8/24	9/24	10/24	11/24	12/24
idealized ($N_e=12$)	0.03000662	0.04317759	0.052367	0.05940697	0.06499195	0.06947143	0.07304815	0.07584952	0.07795895	0.07943146	0.08030203	0.08059013
♀biased ($N_e \approx 12$)	0.032096942	0.04913652	0.055738827	0.058387799	0.067052887	0.073641601	0.075413233	0.075709863	0.080016574	0.083166919	0.082440734	0.080659188
6 A_1	1.849796e-09	1.327306e-07	1.519557e-06	8.341719e-06	3.069468e-05	8.771458e-05	0.0002104102	0.0004436486	0.0008467416	0.001492078	0.00246125	0.003838119
5 A_1	2.343074e-07	7.544684e-06	5.34465e-05	0.0002053346	0.0005637799	0.001248818	0.00237855	0.00404338	0.006279208	0.009042898	0.01219335	0.01548149
4 A_1	1.155117e-05	0.0001679435	0.0007341357	0.001962334	0.00398733	0.006772968	0.01010143	0.01359706	0.01677634	0.01911588	0.02012995	0.01944927
3 A_1	0.0002758736	0.001810524	0.004831236	0.008836589	0.01298487	0.01638283	0.01828587	0.01822074	0.0160418	0.01193272	0.006512489	0.00312143
2 A_1	0.003146291	0.009120055	0.01424072	0.01668576	0.01593751	0.01232856	0.006784793	0.003100204	0.006815464	0.01252548	0.01689294	0.01944927
1 A_1	0.01261452	0.01346206	0.008018969	0.00299108	0.008117812	0.01477756	0.01941946	0.02183749	0.02222073	0.02096359	0.01855249	0.01548149
0 A_1	0.01604847	0.02456826	0.0278588	0.02769836	0.02543089	0.02204315	0.01823272	0.01446734	0.01103629	0.008094273	0.005698265	0.003838119

219

220

MDs subdivided according to categories of A_1 number in breeding males; only shows weighted values and the first half of the A_1 frequency spectrum (the other half is symmetrical).

A **B** **C**