



HAL
open science

Comment on Three Papers about Hardy-Weinberg Equilibrium Tests in Autopolyploids

David Gerard

► **To cite this version:**

David Gerard. Comment on Three Papers about Hardy-Weinberg Equilibrium Tests in Autopolyploids. *Frontiers in Genetics*, 2022, 13, 10.3389/fgene.2022.1027209 . hal-03754674v2

HAL Id: hal-03754674

<https://hal.science/hal-03754674v2>

Submitted on 4 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN ACCESS

EDITED BY
Matthew B. Hamilton,
Georgetown University, United States

REVIEWED BY
Sebastian E. Ramos,
Centre for Research in Agricultural
Genomics (CSIC), Spain

*CORRESPONDENCE
David Gerard,
dgerard@american.edu

SPECIALTY SECTION
This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

RECEIVED 24 August 2022
ACCEPTED 12 September 2022
PUBLISHED 04 October 2022

CITATION
Gerard D (2022), Comment on three
papers about Hardy–Weinberg
equilibrium tests in autopolyploids.
Front. Genet. 13:1027209.
doi: 10.3389/fgene.2022.1027209

COPYRIGHT
© 2022 Gerard. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Comment on three papers about Hardy–Weinberg equilibrium tests in autopolyploids

David Gerard*

Department of Mathematics and Statistics, American University, Washington, DC, United States

KEYWORDS

double reduction, equilibrium, Hardy–Weinberg, hypothesis tests, polyploids, random mating

1 Introduction

One goal in the population genetics of autopolyploids, which are organisms with more than two sets of homologous chromosomes, is to model their genotype frequencies. This modeling presents a greater challenge in autopolyploids than in diploids because certain meiotic configurations in autopolyploids can result in a phenomenon known as double reduction is the co-migration of sister chromatid segments to the same gamete (Stift et al., 2010). Double reduction affects both the segregation frequencies of genotypes from individuals to their offspring (Mather, 1936; Fisher and Mather, 1943; Bever and Felber, 1992; Huang et al., 2019) and the equilibrium genotype frequencies of a panmictic population (Geiringer, 1949; Bennett, 1968; Bever and Felber, 1992; Huang et al., 2019). Testing if a population is in equilibrium, or merely exhibits random mating, is important for the same reasons as tests for Hardy–Weinberg equilibrium in diploids, namely, for 1) determining the mating system, 2) detecting segregation distortion, 3) detecting inbreeding, and 4) detecting genotyping errors (Gerard, 2022b).

Three similar articles which attempt to test for equilibrium and random mating were recently released: one for tetraploids (Sun et al., 2021), one for hexaploids (Wang et al., 2022), and one for octoploids (Wang et al., 2021). These three articles have numerous implementation mistakes, confuse random mating and equilibrium, confuse autopolyploids and allopolyploids (organisms with homoeologous subgenomes), and provide suboptimal testing approaches. The objectives of this study are to correct the authors' mistakes (Section 2), provide examples of how random mating and equilibrium differ in autopolyploids (Section 3), provide examples of how allo- and autopolyploids differ (Section 4), and promote the better methods of Gerard (2022b) and Gerard (2022a) (Section 5).

This study requires a little notation before the issues are discussed here. Let $\mathbf{q} = (q_0, q_1, \dots, q_K)$ be the genotype frequencies at a single biallelic locus for an autopolyploid population with ploidy K ; that is, q_k is the proportion of individuals in the population with k copies of the minor allele. Let $\mathbf{x} = (x_0, x_1, \dots, x_K)$ be the genotype counts in a random sample of $n = \sum_{k=0}^K x_k$ individuals. Then \mathbf{x} is multinomially distributed with size n and probability vector \mathbf{q} . Under random mating, the genotype frequencies are (Gerard, 2022b)

TABLE 1 Segregation frequencies for an autooctoploid when there is no double reduction, either according to Table 1 from the study by Wang et al. (2021) or according to the correct calculation using the hypergeometric distribution (Eq. 4). The two approaches are different, so the general model for meiosis in the study by Wang et al. (2021) is incorrect.

Parent genotype	Method	Gamete genotype				
		4	3	2	1	0
8	Wang et al. (2021)	1	0	0	0	0
8	Correct	1	0	0	0	0
7	Wang et al. (2021)	9/16	3/8	1/16	0	0
7	Correct	1/2	1/2	0	0	0
6	Wang et al. (2021)	225/784	45/98	87/392	3/98	1/784
6	Correct	3/14	8/14	3/14	0	0
5	Wang et al. (2021)	25/196	75/196	285/784	45/392	9/784
5	Correct	1/14	6/14	6/14	1/14	0
4	Wang et al. (2021)	9/196	12/49	41/98	12/49	9/196
4	Correct	1/70	16/70	36/70	16/70	1/70
3	Wang et al. (2021)	9/784	45/392	285/784	75/196	25/196
3	Correct	0	1/14	6/14	6/14	1/14
2	Wang et al. (2021)	1/784	3/98	87/392	45/98	225/784
2	Correct	0	0	3/14	8/14	3/14
1	Wang et al. (2021)	0	0	1/16	3/8	9/16
1	Correct	0	0	0	1/2	1/2
0	Wang et al. (2021)	0	0	0	0	1
0	Correct	0	0	0	0	1

$$q_k = \sum_{i=\max(0, k-K/2)}^{\min(k, K/2)} p_i p_{k-i} \tag{1}$$

where $\mathbf{p} = (p_0, p_1, \dots, p_{K/2})$ are the gamete frequencies of the population; that is, p_k is the proportion of gametes in the population that have k copies of the minor allele. Suppose that a population is randomly mating, then there exists a function $f(\mathbf{q}, \alpha) = (f_0(\mathbf{q}, \alpha), \dots, f_K(\mathbf{q}, \alpha))$ that updates the genotype frequencies from the current generation \mathbf{q} to the next $f(\mathbf{q}, \alpha)$. Here, α is called the double reduction rate, which is a property of meiosis in autopolyploids (Stift et al., 2010). If the population is at equilibrium, then the genotype frequencies follow

$$\mathbf{q} = f(\mathbf{q}, \alpha). \tag{2}$$

For each ploidy, there is a \mathbf{q} that satisfies Eq. 2, which is called the “equilibrium genotype frequencies” (Huang et al., 2019). These frequencies are a function of the double reduction rate α and the allele frequency $r = \frac{1}{K} \sum_{k=0}^K k q_k$, and have been calculated for ploidies less than or equal to ten (Huang et al., 2019). If $\alpha = 0$, then these equilibrium genotype frequencies reduce to binomial proportions (Haldane, 1930),

$$q_k = \binom{K}{k} r^k (1-r)^{K-k}. \tag{3}$$

This study concerns tests for Eqs 1–3.

2 Implementation and coding errors

There are many logical and coding issues in the studies by Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022). In this section, the ones that were found are listed. However, the code from the study by Sun et al. (2021) is not available, and the code from the study by Wang et al. (2021) and Wang et al. (2022) is verbose and sparsely documented, so there might be more implementation errors that were missed. In particular, the following were found: 1) an incorrect model for meiosis for autooctoploids that results in incorrect equilibrium genotype frequencies, 2) two instances of incorrect χ^2 test statistic calculations, 3) five instances of incorrect degrees of freedom calculations, and 4) two instances of using unknown parameters in an estimation scheme.

The model for meiosis in the study by Wang et al. (2021) is incorrect. This leads to incorrect equilibrium genotype frequencies in their “recursive” test for equilibrium, and thus an incorrect test for equilibrium. It can be determined that their model is incorrect by looking at what it implies when $\alpha = 0$. In this case, the distribution of gamete dosages is known to follow a hypergeometric distribution (Table 1 from Haldane, 1930; Huang et al., 2019). If X is the parental genotype and Y is the gamete genotype, the reader can see this result by thinking of the probability of obtaining Y minor alleles out of $K/2$ chosen alleles from an individual with K total

alleles and X total minor alleles. Therefore, the correct segregation frequencies are obtained *via*

$$Pr(Y = y|X = x) = \frac{\binom{x}{y} \binom{K-x}{K/2-y}}{\binom{K}{K/2}}. \quad (4)$$

Table 1 shows that the model for meiosis from Table 1 of the study by Wang et al. (2021) does not equal the probabilities from Eq. 4 when $\alpha = 0$, indicating that their model for meiosis is incorrect. It can be empirically observed that their equilibrium frequencies also do not equal binomial proportions when $\alpha = 0$ (Supplementary Appendix S2), which they should (Haldane, 1930).

The χ^2 statistics testing hypotheses Eq. 1 and Eq. 2 are implemented incorrectly in the study by Wang et al. (2022). The χ^2 statistic in Eq. 1 of the study by Wang et al. (2022) is correct in the study, but in their code, they left out the N term. This affects both their equilibrium testing results and their random mating results. This is known because this study reproduced their 6.602 and 6.649 values from page five of their article (Supplementary Appendix S3). Thus, their tests are improperly implemented.

The random mating test is implemented incorrectly in the study by Wang et al. (2022), even after the N term is included; that is, the authors calculate it differently than what they state in their article. Particularly, in their code, they first estimate the gamete frequencies *via* maximum likelihood, and then put the resulting genotype frequencies through the recursive formula to come up with equilibrium values. However, using this recursive formula just results in the same genotype frequencies as the equilibrium “recursive” test. So, the 6.602 value and the 6.649 value aforementioned are different merely because the authors ran the recursive relationship for a different number of iterations.

A total of five instances of incorrect degrees of freedom calculations were counted for the chi-squared tests in the studies by Sun et al. (2021), Wang et al. (2022), and Wang et al. (2021). These calculations are described in Supplementary Appendix S6. Thus, most of the test statistics from these three articles are compared to the incorrect null distribution, resulting in incorrect p -values.

The estimates of α are implemented incorrectly in the study by Wang et al. (2022). The authors do not modularize their code into functions, and this led to some logical errors. They have a variable in their simulations called `alpha`, that is, the true double reduction rate. Their code returns `alpha1`, that is, the estimated double reduction rate. However, their EM algorithm uses `alpha`, not the current version of `alpha1`, to update the parental gamete frequencies. Thus, they use the true value of α in their code that estimates α . This clearly results in unwarranted advantages. Their simulations were rerun after that bug was fixed, obtaining very biased estimates of α (Supplementary Appendix S5). This indicates that either their

EM algorithm is wrong or their code is incorrect. It is hard to judge if their EM algorithm is wrong since the EM algorithm used to estimate α is neither in the article nor in the Supplementary Materials. It is also noted that when the authors’ “estimate_alpha.R” was run, which should produce the simulation results in their Table 3, it was not able to actually reproduce their Table 3.

From what can be understood through their code (in files “table2_power.R” and “LR.R”), Wang et al. (2021) implemented their tests by using the true genotype frequencies when constructing their test statistics. Needless to say, researchers would not have access to the true genotype frequencies in reality. In their “table2_power.R,” they set some genotype frequencies q_1 and then obtain the underlying true genotype frequencies *via* a perturbation of $q = f(q_1, \alpha)$, where $\alpha = 0$. They obtain two equivalently valued variables called `prob` and `prob1`. They use a perturbation of `prob` to generate the data, and `prob1` to construct the test statistic, but both `prob` and `prob1` are equal to q . An annotated version of “table2_power.R” is provided in the Supplementary Material so that it is easier for the reader to see the issue here. Though the reader is warned, their code is rather verbose and spans 49 8.5 × 11” pages. Because their test statistic is impossible to be calculated in real analyses (because it uses the true genotype frequencies), the simulation results of Wang et al. (2021) are invalid. It is also noted that when “table2_power.R” was run using the authors’ original code, it was not able to actually reproduce the power results in their Table 2.

3 Distinction between random mating and equilibrium

Sun et al. (2021), Wang et al. (2022), and Wang et al. (2021) suggest that Eqs 1 and 2 are the same hypothesis, or at least approximately so. In their articles, they have a “recursive” test and a “gamete-based” test that they claim both tests for “asymptotic Hardy–Weinberg equilibrium.” Their “recursive” test does indeed evaluate Eq. 2 (assuming α is known). However, the “gamete-based” test actually evaluates Eq. 1.

Since the authors say that Eq. 1 is about the same as Eq. 2 for any choice of α , this is worth some exploration. As an extreme counterexample (Supplementary Appendix S1), let $p = (0, 0, 1, 0)$, then hypothesis 1) states that

$$q_1 = (0, 0, 0, 0, 1, 0, 0). \quad (5)$$

But q_1 is not at equilibrium, and one can use q_1 as the starting point for many rounds of random mating to reach equilibrium (Eq. 2). When one does, one obtains

$$q_2 = (0.001, 0.016, 0.082, 0.219, 0.329, 0.263, 0.088), \quad (6)$$

when $\alpha = 0$, the lower bound of the double reduction rate. One also obtains

$$q_3 = (0.005, 0.032, 0.098, 0.204, 0.277, 0.251, 0.133), \quad (7)$$

when $\alpha = 0.3$, the upper bound of the double reduction rate (Huang et al., 2019). Clearly, q_1 , q_2 , and q_3 are very different. But Sun et al. (2021), Wang et al. (2022), and Wang et al. (2021) suggest that they should be about the same.

As a less contrived example, S1 populations (a single generation of selfing) are technically random mating populations, but hardly any researcher would claim that an S1 population is at equilibrium (see Gerard (2022b) for details).

The only real data example used in the study by Wang et al. (2022) consists of four markers from an F1 population. This is insufficient to explore their methods, as F1 populations exhibit neither random mating (Eq. 1) nor equilibrium (Eq. 2). Furthermore, they did not apply their test for random mating on these data, but rather a test for binomial frequencies (Eq. 3), which is a standard approach, though an incorrect one for F1 populations.

4 Distinction between allo- and autopolyploids

Wang et al. (2021) stated on page four that “The case of no double reduction in the autopolyploid model reduces to allopolyploids if no preferential pairing is assumed.” Sun et al. (2021) stated on page three that “When $\alpha = 0$, the pattern of allelic inheritance reduces from autotetraploids to allotetraploids.” Since allopolyploids exhibit disomic inheritance within each subgenome (Stift et al., 2010), this is true only if all subgenomes of an allopolyploid have the exact same allele frequency. This is likely not the case in true allopolyploids. In an extreme example, suppose that there is an allooctoploid population with an allele frequency of 0 in two of its subgenomes, and an allele frequency of one in the other two subgenomes; then the overall allele frequency is 0.5, and the allooctoploid equilibrium genotype frequencies are

$$q_{allo} = (0, 0, 0, 0, 1, 0, 0, 0, 0), \quad (8)$$

because every individual will have two minor alleles each from two subgenomes, and two major alleles each from two subgenomes, and therefore, all individuals will have genotype 4. Compare this to the genotype frequencies of an autooctoploid with allele frequency 0.5 at equilibrium when there is no double reduction

$$q_{auto} = (0.004, 0.031, 0.109, 0.219, 0.273, 0.219, 0.109, 0.031, 0.004). \quad (9)$$

Clearly, q_{allo} and q_{auto} are very different. This is not a contrived example, as it might be the case that some subgenomes have fixed an allele before the polyploidization event (“fixed heterozygosity,” Cornille et al., 2016).

The tests created in the studies by Sun et al. (2021), Wang et al. (2022), and Wang et al. (2021) are only applicable to autopolyploids, but the only real data example in the studies by Sun et al. (2021) and Wang et al. (2021) are allopolyploids. So the authors did not adequately evaluate their method on a reasonable dataset.

5 Hypothesis testing strategies

The test for equilibrium (Eq. 2) in the studies by Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022) assumes that the double reduction rate is known. But it would not be clear to the reader that this is the case from a reading of the articles. The double reduction rate is never known in practice.

The “recursive” approach in the studies by Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022) for equilibrium testing is unnecessary. The equilibrium frequencies of tetraploids, hexaploids, and octoploids in the presence of double reduction are well documented in the excellent article of Huang et al. (2019). For example, for hexaploids, the equilibrium gamete frequencies are

$$p_0 = \left(1 - \frac{9(3-\alpha)(6-\alpha)}{(9+\alpha)(9+2\alpha)}r + \frac{27(1-\alpha)(3-\alpha)}{(9+\alpha)(9+2\alpha)}r^2 \right) (1-r), \quad (10)$$

$$p_1 = \left(\frac{9(3-\alpha)(9-4\alpha)}{(9+\alpha)(9+2\alpha)} - \frac{81(1-\alpha)(3-\alpha)}{(9+\alpha)(9+2\alpha)}r \right) r(1-r), \quad (11)$$

$$p_2 = \left(\frac{45\alpha(3-\alpha)}{(9+\alpha)(9+2\alpha)} + \frac{81(1-\alpha)(3-\alpha)}{(9+\alpha)(9+2\alpha)}r \right) r(1-r), \text{ and} \quad (12)$$

$$p_3 = \left(\frac{20\alpha^2}{(9+\alpha)(9+2\alpha)} + \frac{45\alpha(3-\alpha)}{(9+\alpha)(9+2\alpha)}r + \frac{27(1-\alpha)(3-\alpha)}{(9+\alpha)(9+2\alpha)}r^2 \right) r. \quad (13)$$

The equilibrium genotype frequencies are discrete linear convolutions of these proportions. Eqs 10–13 look complicated, but they are not complicated for a computer. It is easy to implement a likelihood approach to test for equilibrium using these gamete frequencies, and such an approach, advantageously, does not depend on knowing the double reduction rate, which is a huge benefit over the iterative approach of Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022). Indeed, this likelihood approach is what was used in the study by Gerard (2022b).

Genotype uncertainty is a major issue in polyploids (Gerard et al., 2018; Gerard and Ferrão, 2019; Gerard, 2021a,b), and so methods should be adjusted to account for this uncertainty. The standard approach to do so is using genotype likelihoods (Li et al., 2011), and this is what was used in the studies by Gerard (2022b) and Gerard (2022a). However, Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022) approach this by aggregating heterozygous genotypes into a single count, which leaves them

with only enough degrees of freedom to test for binomial frequencies Eq. 3. They thus provide no way to evaluate hypotheses Eq. 1 and Eq. 2 in the presence of genotype uncertainty.

6 Discussion

Here, some implementation mistakes and some misconceptions about the genotype frequencies of autopolyploid organisms presented in the studies by Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022) have been discussed. Examples of how random mating and equilibrium differ in autopolyploids, and also how allo- and autopolyploid equilibrium genotype frequencies differ, have been provided. Finally, it was suggested that users consider the approaches of Gerard (2022b) and Gerard (2022a), which do not assume that the double reduction rate is known, and can account for genotype uncertainty through the use of genotype likelihoods.

Sun et al. (2021), Wang et al. (2021), and Wang et al. (2022) could have averted many of their issues if they would have adhered to standard practices in the validation of numeric analysis and coding. In the future, the authors are encouraged to 1) apply unit testing (Wickham, 2011), 2) set up continuous integration (Hilton et al., 2016), 3) implement code review (Vable et al., 2021), 4) modularize their code into functions, ideally, in a package (Wickham, 2015), 5) use a workflow management software to aid in reproducibility and decrease the chance for coding errors (Blischak et al., 2019), 6) provide instructions (ideally automation) on specifically how to reproduce their methods (Heil et al., 2021), and 7) post their code on a repository that is committed to permanency and produces DOI's, such as Zenodo or Figshare, as this extends the lifetime of a work's reproducibility. It is also recommended to encourage greater validation checks, such as demonstrating that the authors' test statistic produce p -values that are uniform under the null. This alone could have detected the test statistic and degrees of freedom issues discussed in Section 2 (Supplementary Appendix S4).

Data availability statement

Additional materials related to this work is available on Zenodo: <https://doi.org/10.5281/zenodo.7019205>.

- The file “hwesupp.Rmd” is an R Markdown file that contains Appendices S1–S6, and is sufficient to reproduce all of the results of this paper. It has been knitted into “hwesupp.pdf”.
- The file “sims.csv” contains the simulation output from Appendix S5 of “hwesupp.Rmd”.

- The file “table 2_power.Rmd” contains one iteration of “table 2_power.R” from Wang et al., 2021, annotated to demonstrate the mistakes here. It has been knitted into “table 2_power.pdf”.

Much of the code from Wang et al., 2021 and Wang et al., 2022 was packaged by me in the `hexocto` package on Zenodo <https://doi.org/10.5281/zenodo.7019230>.

A fork of the original code from Wang et al., 2021 and Wang et al., 2022 may be found at <https://github.com/dcgerard/hexaploid> and <https://github.com/dcgerard/OctoploidDeer>.

Author contributions

DG conceived of the study, implemented the study, and wrote the manuscript.

Funding

This material is based upon the work supported by the National Science Foundation under grant no. 2132247.

Acknowledgments

Most analyses were performed using the R statistical language (R Core Team, 2022).

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.1027209/full#supplementary-material>

References

- Bennett, J. H. (1968). Mixed self- and cross-fertilization in a tetrasomic species. *Biometrics* 24, 485–500. doi:10.2307/2528313
- Bever, J. D., and Felber, F. (1992). The theoretical population genetics of autopolyploidy. *Oxf. Surv. Evol. Biol.* 8, 185–217.
- Blischak, J. D., Carbonetto, P., and Stephens, M. (2019). Creating and sharing reproducible research code the workflow way. *F1000Res*, 8, 1749. doi:10.12688/f1000research.20843.1
- Cornille, A., Salcedo, A., Kryvokhyzha, D., Glémin, S., Holm, K., Wright, S. I., et al. (2016). Genomic signature of successful colonization of Eurasia by the allopolyploid shepherd's purse (*Capsella bursa-pastoris*). *Mol. Ecol.* 25, 616–629. doi:10.1111/mec.13491
- Fisher, R. A., and Mather, K. (1943). The inheritance of style length in *Lythrum salicaria*. *Ann. Eugen.* 12, 1–23. doi:10.1111/j.1469-1809.1943.tb02307.x
- Geiringer, H. (1949). Chromatid segregation of tetraploids and hexaploids. *Genetics* 34, 665–684. doi:10.1093/genetics/34.6.665
- Gerard, D. (2022a). Bayesian tests for random mating in autopolyploids. *bioRxiv*. doi:10.1101/2022.08.11.503635
- Gerard, D. (2022b). Double reduction estimation and equilibrium tests in natural autopolyploid populations. *Biometrics*. (in press). doi:10.1111/biom.13722
- Gerard, D., Ferrão, L. F. V., Garcia, A. A. F., and Stephens, M. (2018). Genotyping polyploids from messy sequencing data. *Genetics* 210, 789–807. doi:10.1534/genetics.118.301468
- Gerard, D., and Ferrão, L. F. V. (2019). Priors for genotyping polyploids. *Bioinformatics* 36, 1795–1800. doi:10.1093/bioinformatics/btz852
- Gerard, D. (2021a). Pairwise linkage disequilibrium estimation for polyploids. *Mol. Ecol. Resour.* 21, 1230–1242. doi:10.1111/1755-0998.13349
- Gerard, D. (2021b). Scalable bias-corrected linkage disequilibrium estimation under genotype uncertainty. *Heredity* 127, 357–362. doi:10.1038/s41437-021-00462-5
- Haldane, J. (1930). Theoretical genetics of autopolyploids. *Journ. Gen.* 22, 359–372. doi:10.1007/BF02984197
- Heil, B. J., Hoffman, M. M., Markowitz, F., Lee, S.-I., Greene, C. S., and Hicks, S. C. (2021). Reproducibility standards for machine learning in the life sciences. *Nat. Methods* 18, 1132–1135. doi:10.1038/s41592-021-01256-7
- Hilton, M., Tunnell, T., Huang, K., Marinov, D., and Dig, D. (2016). Usage, costs, and benefits of continuous integration in open-source projects. *2016 31st IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE) (IEEE)*, 426–437. doi:10.1145/2970276.2970358
- Huang, K., Wang, T., Dunn, D. W., Zhang, P., Cao, X., Liu, R., et al. (2019). Genotypic frequencies at equilibrium in polysomic inheritance under double-reduction. *G3 Genes |Genomes |Genetics* 9, 1693–1706. doi:10.1534/g3.119.400132
- Li, Y., Sidore, C., Kang, H. M., Boehnke, M., and Abecasis, G. R. (2011). Low-coverage sequencing: Implications for design of complex trait association studies. *Genome Res.* 21, 940–951. doi:10.1101/gr.117259.110
- Mather, K. (1936). Segregation and linkage in autotetraploids. *Journ. Genet.* 32, 287–314. doi:10.1007/BF02982683
- R Core Team (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Stift, M., Reeve, R., and Van Tienderen, P. H. (2010). Inheritance in tetraploid yeast revisited: Segregation patterns and statistical power under different inheritance models. *J. Evol. Biol.* 23, 1570–1578. doi:10.1111/j.1420-9101.2010.02012.x
- Sun, L., Gan, J., Jiang, L., and Wu, R. (2021). Recursive test of Hardy-Weinberg equilibrium in tetraploids. *Trends Genet.* 37, 504–513. doi:10.1016/j.tig.2020.11.006
- Vable, A. M., Diehl, S. F., and Glymour, M. M. (2021). Code review as a simple trick to enhance reproducibility, accelerate learning, and improve the quality of your team's research. *Am. J. Epidemiol.* 190, 2172–2177. doi:10.1093/aje/kwab092
- Wang, J., Feng, L., Mu, S., Dong, A., Gan, J., Wen, Z., et al. (2022). Asymptotic tests for Hardy-Weinberg equilibrium in hexaploids. *Hortic. Res.* 9. doi:10.1093/hr/uha014
- Wang, J., Lv, X., Feng, L., Dong, A., Liang, D., and Wu, R. (2021). A tracing model for the evolutionary equilibrium of octoploids. *Front. Genet.* 12. doi:10.3389/fgene.2021.794907
- Wickham, H. (2015). R packages: Organize, test, document, and share your code (*O'reilly media*).
- Wickham, H. (2011). testthat: Get started with testing. *R J.* 3, 5–10. doi:10.32614/RJ-2011-002