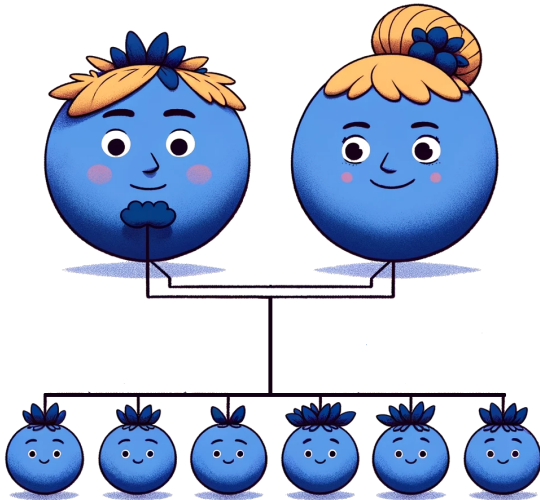# Tests for Segregation Distortion in F1 Populations of Tetraploids

David Gerard, Mira Thakkar, Luis Felipe Ventorim Ferrão

Setting: F1 Population of Tetraploids

Setting: Single Biallelic Marker

- $0 =$ AAAA
- $1 =$ AAAa
- $2 =$ AAaa
- $3 =$ Aaaa
- $4 =$ aaaa
- Parent Genotypes: $g_1, g_2 \in \{0, 1, 2, 3, 4\}$.

Tests for Segregation Distortion

▶ Quality Control: Verify if biallelic marker follows rules of Mendelian segregation.

▶ Run a χ^2 test against known distribution of offspring genotypes.

▶ E.g., when $g_1 = 1$ and $g_2 = 2$, then the distribution of offspring genotypes is

$$\mathbf{q} = (q_0, q_1, q_2, q_3, q_4) = (1, 5, 5, 1, 0)/12 \tag{1}$$

▶ q_k is the frequency of genotype k.

Issues

▶ Double Reduction: Co-migration of sister chromatid segments to the same gamete.

▶ (Partial) Preferential Pairing: Homologs preferentially pair during meiosis.

▶ Genotype Uncertainty

Solution for Genotype Uncertainty

▶ Use genotype likelihoods

▶ Sum over genotype uncertainty (à la Li (2011)).

$$\Pr(\text{Data}|\mathbf{q}) = \sum_{\text{Genotypes}} \Pr(\text{Data}|\text{Genotypes})\Pr(\text{Genotypes}|\mathbf{q})$$

▶ Use this likelihood to run a likelihood ratio test (LRT).

Model for Meiosis

▶ Distribution of offspring genotypes boils down to distribution of gamete genotypes for each parent.

▶ Let p_{ik} be the probability a gamete will have genotype $k \in \{0, 1, 2\}$ given parent $i \in \{1, 2\}$. Then offspring dosages are

$$
\begin{aligned}
q_0 &= p_{10}p_{20}, \\
q_1 &= p_{10}p_{21} + p_{11}p_{20}, \\
q_2 &= p_{10}p_{22} + p_{11}p_{21} + p_{12}p_{20}, \\
q_3 &= p_{11}p_{22} + p_{12}p_{21}, \text{ and} \\
q_4 &= p_{12}p_{22}.
\end{aligned}
\tag{2}
$$

▶ Accounting for double reduction and partial preferential pairing amounts to models on the p_{ik}'s.

Muller (1914)

	$g = 1$	$g = 2$	$g = 3$
$y = 0$	1/2	1/6	0
$y = 1$	1/2	2/3	1/2
$y = 2$	0	1/6	1/2

▶ Basic polysomic inheritance (Hypergeometric)

▶ $g \in \{0, 1, 2, 3, 4\}$: Parent genotype

▶ $y \in \{0, 1, 2\}$: Gamete genotype

Fisher and Mather (1943)

	$g = 1$	$g = 2$	$g = 3$
$y = 0$	$\frac{1}{2} + \frac{1}{4}\alpha$	$\frac{1}{2}\alpha + \frac{1}{6}(1 - \alpha)$	$\frac{1}{4}\alpha$
$y = 1$	$\frac{1}{2} - \frac{1}{2}\alpha$	$\frac{2}{3}(1 - \alpha)$	$\frac{1}{2} - \frac{1}{2}\alpha$
$y = 2$	$\frac{1}{4}\alpha$	$\frac{1}{2}\alpha + \frac{1}{6}(1 - \alpha)$	$\frac{1}{2} + \frac{1}{4}\alpha$

▶ Accounts for double reduction, but not preferential pairing.

▶ $g \in \{0, 1, 2, 3, 4\}$: Parent genotype

▶ $y \in \{0, 1, 2\}$: Gamete genotype

▶ α: Double reduction rate

Our Work

	$g = 1$	$g = 2$	$g = 3$
$y = 0$	$\frac{1}{2} + \frac{1}{4}\alpha$	$\frac{1}{2}\alpha + \frac{1}{4}(1-\alpha)(1-\xi)$	$\frac{1}{4}\alpha$
$y = 1$	$\frac{1}{2} - \frac{1}{2}\alpha$	$\frac{1}{2}(1-\alpha)(1+\xi)$	$\frac{1}{2} - \frac{1}{2}\alpha$
$y = 2$	$\frac{1}{4}\alpha$	$\frac{1}{2}\alpha + \frac{1}{4}(1-\alpha)(1-\xi)$	$\frac{1}{2} + \frac{1}{4}\alpha$

▶ We modified Stift et al. (2008) to biallelic loci.

  ▶ Original model assumes chromosomes are distinguishable.

▶ $g \in \{0, 1, 2, 3, 4\}$: Parent genotype

▶ $y \in \{0, 1, 2\}$: Gamete genotype

▶ α: Double reduction rate

▶ ξ: Preferential pairing parameter

 ▶ $1/3$ means tetrasomic, 0 or 1 means disomic.

Takeaways

▶ Only double reduction (not preferential pairing) is important when parent genotypes are 1 and 3.

▶ Any gamete frequency between $(0.25, 0.5, 0.25)$ and $(0, 1, 0)$ is possible given a parent genotype of 2.

  ▶ If you assume either fully disomic or fully polysomic, only possible gamete frequencies are:

    $(0.25, 0.5, 0.25)$, $(0.167, 0.667, 0.167)$, and $(0, 1, 0)$.

▶ Double reduction and preferential pairing are not separately identifiable.

  ▶ $\alpha = 0$ and $\xi = 1/9$ is the same as $\alpha = 1/6$ and $\xi = 1/3$.

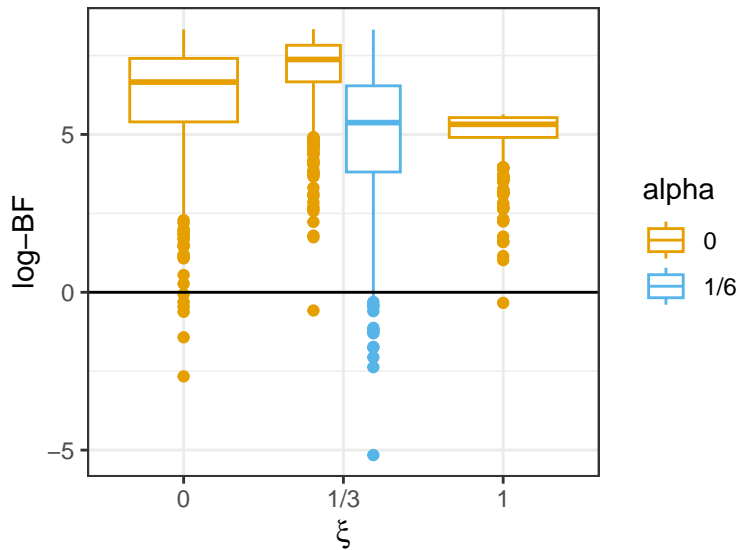 ▶ Only important when parent genotype is 2.

Our Tests

- Likelihood ratio tests
  - Asymptotically conservative (parameters can be on boundary).
- Bayesian Tests:
  - $\mathrm{BF} = \frac{\mathrm{Prob}(\mathrm{data}|H_0)}{\mathrm{Prob}(\mathrm{data}|H_1)}$
  - log-BF > 0 means supports null.
  - log-BF < 0 means support alternative.

Simulation Example

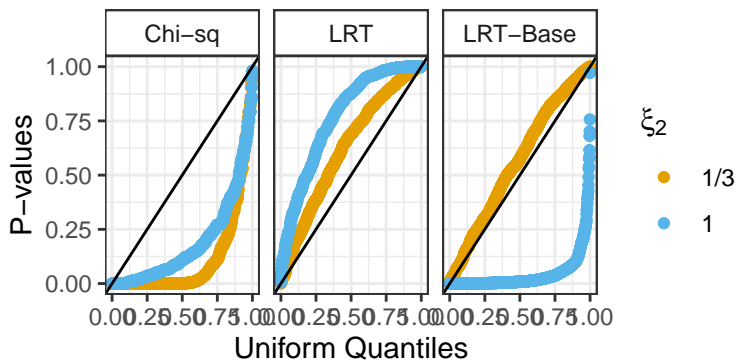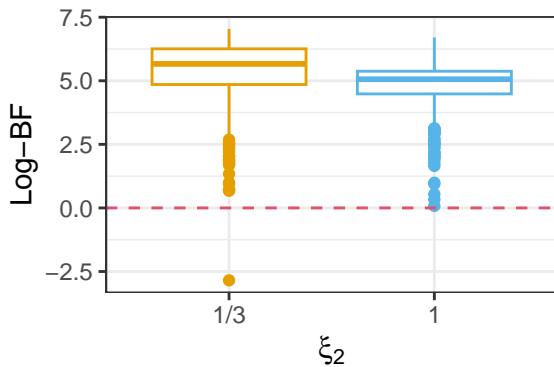▶ $n = 100$, $g_1 = 1$, $g_2 = 2$, genotypes known, 1000 replicates.
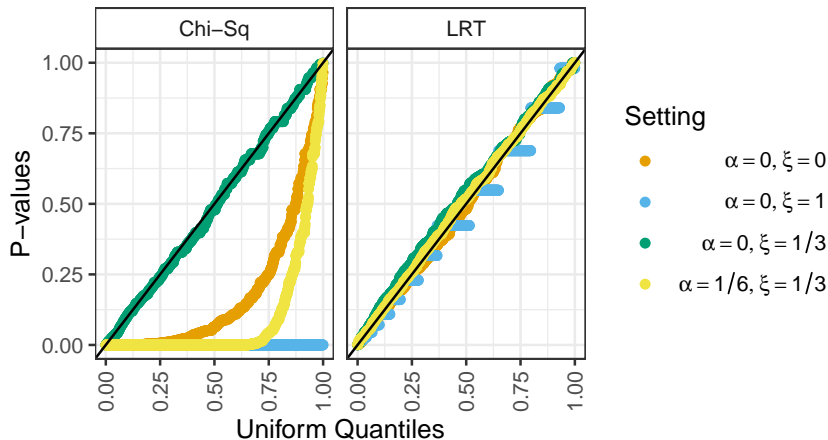
Bayes Factor

Genotype Likelihoods

▶ $n = 100$, $g_1 = 1$, $g_2 = 2$, $\alpha = 0$, $\xi_1 = 1/3$, $\xi_2 = 1/3$ or 1, read-depth of 10, genotype likelihoods generated via updog, 1000 replicates.
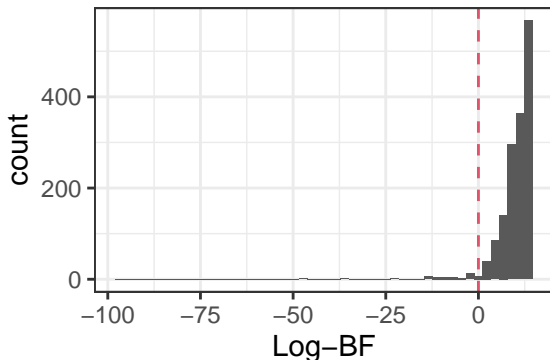
Bayes Factors

Real Data Example

- $n = 240$ blueberries from an F1 population of Cappai et al. (2020), 1547 SNPs on Chromosome 1.

- Using {qvalue}, estimated proportion of null genes is 1 using LRT, versus 0.075 using the chi-squared test with posterior modes.

- Bayes factors indicate mostly null

Thanks!

Soon to appear in an R packaged called menbayes.

Appendix

References

▶ Cappai, F., Amadeu, R. R., Benevenuto, J., Cullen, R., Garcia, A., Grossman, A., Ferrão, L.F.V., & Munoz, P. (2020). High-resolution linkage map and QTL analyses of fruit firmness in autotetraploid blueberry. Frontiers in plant science, 11, 562171.

▶ Fisher, R. A., & Mather, K. (1943). The inheritance of style length in *Lythrum salicaria*. *Annals of Eugenics*, 12(1), 1-23.

▶ Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993.

▶ Muller, H. J. (1914). A new mode of segregation in Gregory's tetraploid primulas. *The American Naturalist*, 48(572), 508-512.

▶ Stift, M., Berenos, C., Kuperus, P., & van Tienderen, P. H. (2008). Segregation models for disomic, tetrasomic and intermediate inheritance in tetraploids: a general procedure applied to Rorippa (yellow cress) microsatellite data. *Genetics*, 179(4), 2113-2123.

Prior Specification

▶ τ: Proportion of quadrivalent formation.

▶ β: Probability of double reduction given quadrivalent formation.

▶ γ: Probability that AA/aa pairing given bivalent formation and parent genotype is 2.

▶ Relationship:

$$\alpha = \beta\tau \tag{3}$$

$$\xi = \eta\frac{1}{3} + (1-\eta)\gamma, \text{ where } \eta = \frac{(1-\beta)\tau}{(1-\beta)\tau + (1-\tau)}. \tag{4}$$

▶ Prior:

$$\tau \sim \text{Unif}(0,1), \tag{5}$$

$$\beta \sim \text{Unif}(0,1/6), \text{ and} \tag{6}$$

$$\gamma_1, \gamma_2 \sim \text{Beta}(5/9, 10/9). \tag{7}$$

A little bit of partial preferential pairing