GENOTYPIC FREQUENCIES AT EQUILIBRIUM FOR POLYSOMIC INHERITANCE UNDER DOUBLE-REDUCTION

MATERIAL S1: THE APPENDICES

A Derivation of $\Pr(BBBB \,|\, ABBBBBBB)$

In this example, we show how to derive the transitional probability from a zygote genotype $G = ABBBBBBB$ to a gamete genotype $g = BBBB$. According to Equation (1), m_k and n_k are respectively the numbers of copies of A_k in g and G, h is the number of alleles at this locus, and v is the ploidy level. Letting $v = 8$, $h = 2$, $m_1 = 0$, $m_2 = 4$, $n_1 = 1$ and $n_2 = 7$, and expanding the sum formula in Equation (1), it follows the following expression:

$$
\begin{aligned}
T(BBBB \,|\, ABBBBBBB) = {} & \frac{1\binom{1}{0}\binom{1}{0} \times 1\binom{7}{0}\binom{7}{4}}{\binom{8}{0}\binom{8}{4}}\alpha_0 && i=0, j_1=0, j_2=0 \\
& + \frac{1\binom{1}{0}\binom{1}{0} \times 1\binom{7}{1}\binom{6}{2}}{\binom{8}{1}\binom{7}{2}}\alpha_1 && i=1, j_1=0, j_2=1 \\
& + \frac{0\binom{1}{1}\binom{0}{-2} \times 1\binom{7}{0}\binom{7}{4}}{\binom{8}{1}\binom{7}{2}}\alpha_1 && i=1, j_1=1, j_2=0 \\
& + \frac{1\binom{1}{0}\binom{1}{0} \times 1\binom{7}{2}\binom{5}{0}}{\binom{8}{2}\binom{6}{0}}\alpha_2 && i=2, j_1=0, j_2=2 \\
& + \frac{0\binom{1}{1}\binom{0}{-2} \times 1\binom{7}{1}\binom{6}{2}}{\binom{8}{2}\binom{6}{0}}\alpha_2 && i=2, j_1=1, j_2=1 \\
& + \frac{0\binom{1}{2}\binom{-1}{-4} \times 1\binom{7}{0}\binom{7}{4}}{\binom{8}{2}\binom{6}{0}}\alpha_2 && i=2, j_1=2, j_2=0 \\
= {} & \frac{1}{8}(4\alpha_0 + 5\alpha_1 + 6\alpha_2).
\end{aligned}
$$

Note that $\delta_{jk} = 0$ whenever $j_k < \max(0, m_k - n_k)$ or $j_k > \min(n_k, m_k/2)$. The coefficients of the above terms with a gray background are equal to zero.

B Derivation of α_i under octosomic inheritance

In this example, we show how to derive the values of alpha for octosomic inheritance in both CES and PES.

For the case of CES, we first calculate the values of μ_j for each possible j. Because there are $v/2$ chromosomes, the maximum value of j is $\lfloor v/4 \rfloor$. Note that in this case, we have $v = 8$, then $\lfloor v/4 \rfloor = 2$

and $j = 0, 1, 2$. Now, by Equation (4), we obtain

$$\begin{aligned}
\mu_0 &= 2^4 \binom{4}{0}\binom{4}{4} \Big/ \binom{8}{4} = \frac{8}{35}, \\
\mu_1 &= 2^2 \binom{4}{1}\binom{3}{2} \Big/ \binom{8}{4} = \frac{24}{35}, \\
\mu_2 &= 2^0 \binom{4}{2}\binom{2}{0} \Big/ \binom{8}{4} = \frac{3}{35}.
\end{aligned}$$

Second, we calculate the value of α_i under CES where $0 \leqslant i \leqslant \lfloor v/4 \rfloor$, i.e. $0 \leqslant i \leqslant 2$. By Equation (5), it follows that:

$$\begin{aligned}
\alpha_0 &= 2^0 \mu_0 \binom{0}{0} + 2^{-1}\mu_1 \binom{1}{0} + 2^{-2}\mu_2 \binom{2}{0} \\
&= 1 \cdot \frac{8}{35} \cdot 1 + \frac{1}{2} \cdot \frac{24}{35} \cdot 1 + \frac{1}{4} \cdot \frac{3}{35} \cdot 1 \\
&= \frac{83}{140},
\end{aligned}$$

$$\begin{aligned}
\alpha_1 &= 2^{-1}\mu_1 \binom{1}{1} + 2^{-2}\mu_2 \binom{2}{1} \\
&= \frac{1}{2} \cdot \frac{24}{35} \cdot 1 + \frac{1}{4} \cdot \frac{3}{35} \cdot 2 \\
&= \frac{27}{70}, \\
\alpha_2 &= 2^{-2}\mu_2 \binom{2}{2} \\
&= \frac{1}{4} \cdot \frac{3}{35} \cdot 1 \\
&= \frac{3}{140}.
\end{aligned}$$

For the case of PES, we still have $v = 8$, then $\lfloor v/4 \rfloor = 2$ and the values of μ_0, μ_1 and μ_2 in Equation (6) are just those values in the case of CES. Now, according to Equation (6), we have

$$\begin{aligned}
\nu_0 &= \mu_0 \binom{0}{0} r_s^0 (1 - r_s)^0 + \mu_1 \binom{1}{0} r_s^0 (1 - r_s)^1 + \mu_2 \binom{2}{0} r_s^0 (1 - r_s)^2 \\
&= \frac{8}{35} + \frac{24}{35}(1 - r_s) + \frac{3}{35}(1 - r_s)^2 \\
&= \frac{1}{35}(35 - 30 r_s + 3 r_s^2), \\
\nu_1 &= \mu_1 \binom{1}{1} r_s^1 (1 - r_s)^0 + \mu_2 \binom{2}{1} r_s^1 (1 - r_s)^1 \\
&= \frac{24}{35} r_s + \frac{3}{35} 2 r_s (1 - r_s) \\
&= \frac{1}{35}(30 r_s - 6 r_s^2), \\
\nu_2 &= \mu_2 \binom{2}{2} r_s^2 (1 - r_s)^0 \\
&= \frac{3}{35} r_s^2.
\end{aligned}$$

Furthermore, because of Equation (5), we derive the value of α_k ($k = 0, 1, 2$) under PES as follows:

$$
\begin{aligned}
\alpha_0 &= 2^0\nu_0\binom{0}{0} + 2^{-1}\nu_1\binom{1}{0} + 2^{-2}\nu_2\binom{2}{0} \\
&= 1\cdot\frac{1}{35}(35-30r_s+3r_s^2)\cdot 1 + \frac{1}{2}\cdot\frac{1}{35}(30r_s-6r_s^2)\cdot 1 + \frac{1}{4}\cdot\frac{3}{35}r_s^2\cdot 1 \\
&= \frac{1}{140}(140-60r_s+3r_s^2), \\
\alpha_1 &= 2^{-1}\nu_1\binom{1}{1} + 2^{-2}\nu_2\binom{2}{1} \\
&= \frac{1}{2}\cdot\frac{1}{35}(30r_s-6r_s^2)\cdot 1 + \frac{1}{4}\cdot\frac{3}{35}r_s^2\cdot 2 \\
&= \frac{1}{70}(30r_s-3r_s^2), \\
\alpha_2 &= 2^{-2}\nu_2\binom{2}{2} \\
&= \frac{1}{4}\cdot\frac{3}{35}r_s^2\cdot 1 \\
&= \frac{3}{140}r_s^2.
\end{aligned}
$$

C Derivation of GFG and GFZ with nonlinear method

Here, we use the tetrasomic inheritance at a triallelic locus under equilibrium to derive the GFG by using a non-linear method as an example. Under these conditions, we have $v=4$ and $h=3$. Because $\binom{v/2+h-1}{v/2}=\binom{4}{2}=6$ and $\binom{v+h-1}{v}=\binom{6}{4}=15$, there are 6 gamete genotypes and 15 zygote genotypes, so Equations (8) and (9) determine 6 and 15 equations, respectively.

(i) Simulating meiosis. We denote AA and AB for two gamete genotypes, and P_{AA} and P_{AB} for their frequencies, and so on. Similarly, denote $AABB$ and $ABBC$ for two zygote genotypes, and P_{AABB} and P_{ABBC} for their frequencies, and so on. Now, by Equation (8), the GFG can be established as follows:

$$
\begin{cases}
P_{AA} = P_{AAAA} + \frac{2+\alpha_1}{4}(P_{AAAB}+P_{AAAC}) + \frac{1+2\alpha_1}{6}(P_{AABB}+P_{AABC}+P_{AACC}) \\
\qquad + \frac{\alpha_1}{4}(P_{ABBB}+P_{ABBC}+P_{ABCC}+P_{ACCC}), \\
P_{BB} = P_{BBBB} + \frac{2+\alpha_1}{4}(P_{ABBB}+P_{BBBC}) + \frac{1+2\alpha_1}{6}(P_{AABB}+P_{ABBC}+P_{BBCC}) \\
\qquad + \frac{\alpha_1}{4}(P_{AAAB}+P_{AABC}+P_{ABCC}+P_{BCCC}), \\
P_{CC} = P_{CCCC} + \frac{2+\alpha_1}{4}(P_{ACCC}+P_{BCCC}) + \frac{1+2\alpha_1}{6}(P_{AACC}+P_{ABCC}+P_{BBCC}) \\
\qquad + \frac{\alpha_1}{4}(P_{AAAC}+P_{AABC}+P_{ABBC}+P_{BBBC}), \\
P_{AB} = \frac{1-\alpha_1}{2}(P_{AAAB}+P_{ABBB}) + \frac{2(1-\alpha_1)}{3}P_{AABB} + \frac{1-\alpha_1}{3}(P_{AABC}+P_{ABBC}) \\
\qquad + \frac{1-\alpha_1}{6}P_{ABCC}, \\
P_{AC} = \frac{1-\alpha_1}{2}(P_{AAAC}+P_{ACCC}) + \frac{2(1-\alpha_1)}{3}P_{AACC} + \frac{1-\alpha_1}{3}(P_{AABC}+P_{ABCC}) \\
\qquad + \frac{1-\alpha_1}{6}P_{ABBC}, \\
P_{BC} = \frac{1-\alpha_1}{2}(P_{BBBC}+P_{BCCC}) + \frac{2(1-\alpha_1)}{3}P_{BBCC} + \frac{1-\alpha_1}{3}(P_{ABBC}+P_{ABCC}) \\
\qquad + \frac{1-\alpha_1}{6}P_{AABC}.
\end{cases}
\tag{A1}
$$

(ii) Simulating fertilization. By Equation (9), the GFZ can be established as follows:

$$\begin{cases} P_{AAAA} &= P_{AA}^2, \\ P_{BBBB} &= P_{BB}^2, \\ P_{CCCC} &= P_{CC}^2, \\ P_{AAAB} &= 2P_{AA}P_{AB}, \\ P_{AAAC} &= 2P_{AA}P_{AC}, \\ P_{ABBB} &= 2P_{AB}P_{BB}, \\ P_{BBBC} &= 2P_{BB}P_{BC}, \\ P_{ACCC} &= 2P_{AC}P_{CC}, \\ P_{BCCC} &= 2P_{BC}P_{CC}, \\ P_{AABB} &= 2P_{AA}P_{BB} + P_{AB}^2, \\ P_{AACC} &= 2P_{AA}P_{CC} + P_{AC}^2, \\ P_{BBCC} &= 2P_{BB}P_{CC} + P_{BC}^2, \\ P_{AABC} &= 2P_{AA}P_{BC} + 2P_{AB}P_{AC}, \\ P_{ABBC} &= 2P_{AB}P_{BC} + 2P_{AC}P_{BB}, \\ P_{ABCC} &= 2P_{AB}P_{CC} + 2P_{AC}P_{BC}. \end{cases} \tag{A2}$$

Now, substituting Equation (A2) into Equation (A1), the GFZ are eliminated, and a system of non-linear equations with 6 equations and 6 unknowns is obtained (whose expressions are more complex and omitted). On the other hand, the process that transforms the allele frequencies into GFG can be described by the linear substitution

$$\begin{cases} p_A &= P_{AA} + \frac{1}{2}(P_{AB} + P_{AC}), \\ p_B &= P_{BB} + \frac{1}{2}(P_{AB} + P_{BC}), \end{cases}$$

where p_A, p_B and p_C are the allele frequencies with $p_A + p_B + p_C = 1$. Combining the linear substitution with the system of non-linear equations mentioned above, we still obtain a system of non-linear equations with 8 equations, 6 unknowns (i.e. P_{AA}, P_{AB}, P_{AC}, P_{BB}, P_{BC} and P_{CC}) and 2 parametric variables (i.e. p_A and p_B). The solution that is with p_A and p_B as the parametric variables is unique and is shown as follows:

$$\begin{cases} P_{AA} = \frac{3\alpha_1 p_A + 2(1-\alpha_1)p_A^2}{2+\alpha_1}, \\ P_{AB} = \frac{4(1-\alpha_1)p_A p_B}{2+\alpha_1}, \\ P_{AC} = \frac{4(1-\alpha_1)(1-p_A-p_B)p_A}{2+\alpha_1}, \\ P_{BB} = \frac{3\alpha_1 p_B + 2(1-\alpha_1)p_B^2}{2+\alpha_1}, \\ P_{BC} = \frac{4(1-\alpha_1)(1-p_A-p_B)p_B}{2+\alpha_1}, \\ P_{CC} = \frac{(1-p_A-p_B)(2+\alpha_1-2p_A+2\alpha_1 p_A-2p_B+2\alpha_1 p_B)}{2+\alpha_1}. \end{cases}$$

Therefore, the generalized form for GFG at equilibrium can be directly written as follows:

$$\Pr(g \mid v = 4) = \begin{cases} \frac{3\alpha_1 p_A + 2(1-\alpha_1)p_A^2}{2+\alpha_1} & \text{if } g = AA, \\ \frac{4(1-\alpha_1)p_A p_B}{2+\alpha_1} & \text{if } g = AB. \end{cases} \tag{A3}$$

Using Equation (9), we can derive the generalized form for GFZ at equilibrium as follows:

$$\Pr(G\,|\,v=4)=\begin{cases}\frac{[2p_A+\alpha_1(3-2p_A)]^2p_A^2}{(2+\alpha_1)^2} & \text{if } G=AAAA,\\ \frac{8(\alpha_1-1)[2(\alpha_1-1)p_A-3\alpha_1]p_A^2p_B}{(2+\alpha_1)^2} & \text{if } G=AAAB,\\ \frac{6p_Ap_B\lambda_1}{(2+\alpha_1)^2} & \text{if } G=AABB,\\ \frac{24(\alpha_1-1)[2(\alpha_1-1)p_A-\alpha_1]p_Ap_Bp_C}{(2+\alpha_1)^2} & \text{if } G=AABC,\\ \frac{96(\alpha_1-1)^2p_Ap_Bp_Cp_D}{(2+\alpha_1)^2} & \text{if } G=ABCD.\end{cases} \tag{A4}$$

Where $\lambda_1 = 4p_Ap_B + 2\alpha_1(p_A + p_B - 4p_Ap_B) + \alpha_1^2(3 + 4p_Ap_B - 2p_A - 2p_B)$.

D GFG and GFZ in hexasomic inheritance

The generalized form of GFG for hexasomic inheritance at equilibrium derived from the linear method is given by

$$\Pr(g\,|\,v=6)=\begin{cases}\frac{p_A[20\alpha_1^2-45(\alpha_1-3)\alpha_1p_A+27(\alpha_1-3)(\alpha_1-1)p_A^2]}{(\alpha_1+9)(2\alpha_1+9)} & \text{if } g=AAA,\\ \frac{9(\alpha_1-3)p_A[9(\alpha_1-1)p_A-5\alpha_1]p_B}{(\alpha_1+9)(2\alpha_1+9)} & \text{if } g=AAB,\\ \frac{162(\alpha_1-3)(\alpha_1-1)p_Ap_Bp_C}{(\alpha_1+9)(2\alpha_1+9)} & \text{if } g=ABC.\end{cases}$$

Where AAA, AAB and ABC are the genotypic patterns of g. With Equation (9), the generalized form of GFZ for hexasomic inheritance at equilibrium is given by

$$\Pr(G\,|\,v=6)=\begin{cases}\lambda_9^2p_A^2\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AAAAAA,\\ 18\lambda_9\lambda_2\lambda_7p_A^2p_B\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AAAAAB,\\ 9\lambda_2[2\lambda_9\lambda_8+9\lambda_2(\alpha_1\lambda_6+9p_A)^2p_B]p_A^2p_B\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AAAABB,\\ 810\lambda_2(23\alpha_1^2-13\alpha_1^3+36\alpha_1\lambda_2\lambda_3p_A+27\lambda_2\lambda_3^2p_A^2)p_A^2p_Bp_C\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AAAABC,\\ 2(\lambda_9\lambda_{10}+81\lambda_2^2\lambda_7\lambda_8p_Ap_B)p_Ap_B\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AAABBB,\\ 180\lambda_2[10\alpha_1^3+243\lambda_2\lambda_3^2p_A^2p_B+54\alpha_1\lambda_2\lambda_3p_A(p_A+3p_B) & \text{if } G=AAABBC,\\ \quad+9\alpha_1^2(5\lambda_2p_A+2\lambda_3p_B)]p_Ap_Bp_C\lambda_4^{-2}\lambda_5^{-2} & \\ 3240\lambda_2\lambda_3(2\alpha_1^2+18\alpha_1\lambda_2p_A+27\lambda_2\lambda_3p_A^2)p_Ap_Bp_Cp_D\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AAABCD,\\ 810\lambda_2^2(81p_Ap_Bp_C+18\alpha_1\lambda_{11}+\alpha_1^2\lambda_{12})p_Ap_Bp_C\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AABBCC,\\ 1620\lambda_2^2[5\alpha_1^2+81\lambda_3^2p_Ap_B+18\alpha_1\lambda_3(p_A+p_B)]p_Ap_Bp_Cp_D\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AABBCD,\\ 29160\lambda_2^2\lambda_3(2\alpha_1+9\lambda_3p_A)p_Ap_Bp_Cp_Dp_E\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=AABCDE,\\ 524880\lambda_2^2\lambda_3^2p_Ap_Bp_Cp_Dp_Ep_F\lambda_4^{-2}\lambda_5^{-2} & \text{if } G=ABCDEF.\end{cases}$$

Where $\lambda_2 = 3 - \alpha_1$, $\lambda_3 = 1 - \alpha_1$, $\lambda_4 = 9 + \alpha_1$, $\lambda_5 = 9 + 2\alpha_1$, $\lambda_6 = 5 - 9p_A$, $\lambda_7 = 5\alpha_1 + 9\lambda_3p_A$, $\lambda_8 = 5\alpha_1 + 9\lambda_3p_B$, $\lambda_9 = 20\alpha_1^2 + 45\alpha_1\lambda_2p_A + 27\lambda_2\lambda_3p_A^2$, $\lambda_{10} = 20\alpha_1^2 + 45\alpha_1\lambda_2p_B + 27\lambda_2\lambda_3p_B^2$, $\lambda_{11} = p_Ap_B + p_Ap_C + p_Bp_C - 9p_Ap_Bp_C$, $\lambda_{12} = 5p_A + 5p_B + 5p_C - 18p_Ap_B - 18p_Ap_C - 18p_Bp_C + 81p_Ap_Bp_C$.

E Derivation of co-ancestry coefficients in different relationships

The expression of co-ancestry coefficient θ in mating individuals can be obtained by taken the weighted average of co-ancestry coefficient between different relationships, where the weight is the frequencies of those relationships in mating individuals. We first derive the co-ancestry coefficients in the following four relationships.

In selfing, the individual self-fertilizes. For a pair of alleles sampled from an individual with replacement, the probability is $1/v$ if the same allele is sampled twice and these are indeed IBD; otherwise the probability is F. Hence $\theta_{ID} = \frac{1}{v} + \frac{v-1}{v}F$.

In backcrossing, the offspring is fertilized by, or fertilizes, its parent (says f). Let m be the other parent. Denote g_f and g_m for the two gametes which are respectively produced by f and m to form an offspring. The allele pairs between the offspring and f can be classified into two categories: (i) between g_f and f. In this case, for each allele pair, the probability that the two alleles are IBD is equal to θ_{ID}; (ii) between g_m and f. In this case, for each allele pair, the probability that the two alleles are IBD alleles is equal to θ, i.e. the co-ancestry coefficient between mating individuals. Hence $\theta_{PO} = (\theta_{ID} + \theta)/2$.

In matings between full-siblings (says a and b), it is assumed that the parents are f and m, and let g_{af} and gam be the gametes forming a, where g_{af} is produced by f and g_{am} is produced by m. Similarly, g_{bf} and g_{bm} denote the gametes forming b. For each pair of alleles, the probability that the two alleles are IBD between g_{af}-g_{bf} is θ_{ID}, the same as that between g_{am}-g_{bm}; and the probability that they are IBD between g_{af}-g_{bm} is θ, the same as that between g_{am}-g_{bf}. Hence $\theta_{FS} = (\theta_{ID} + \theta)/2$.

In matings between nonrelatives, $\theta_{UN} = 0$.

Second, we derive the F in population with a selfing ratio s as an example. Assuming a proportion s of individuals is produced by selfing and the remaining proportion $1 - s$ of individuals is produced from matings between nonrelatives, then $\theta = s\theta_{ID} + (1 - s)\theta_{UN}$. Because $\theta_{UN} = 0$, this expression can be simplified into $\theta = s\theta_{ID}$. By substituting this expression into Equation (11), we obtain inbreeding coefficient F at equilibrium: $F = \frac{8\lambda+sv}{8\lambda+v(s+v-sv)}$.