Additional file 1: Computation of $E\left[ X_{im}^{d_i} X_{jm}^{d_j} \cdots X_{Km}^{d_K} \right]$ as a function of between chromosomes identity coefficients, in the case of independent markers.

**Principles**

Derivation of $E[XX'TXX']$ elements are simplified using three properties:

- $E\left( \sum_m X_{im} X_{km} \right) \left( \sum_m X_{jm} X_{im} \right) = \sum_m E\left[ X_{im} X_{km} X_{jm} X_{im} \right] + \left( \sum_m E\left[ X_{im} X_{km} \right] \left( \sum_m E\left[ X_{jm} X_{im} \right] \right) - \sum_m \left( E\left[ X_{im} X_{km} \right] E\left[ X_{jm} X_{im} \right] \right) \right.

- We will demonstrate in this supplementary material that, at least for the $K$ (1 to 4) and $d_k$ (1 to 4) values present in our derivations, any expectation $E\left[ X_{im}^{d_i} X_{jm}^{d_j} \cdots X_{Km}^{d_K} \right]$ with $\sum_l d_l$ even can be written as

$$E[p_m (1 - p_m)^{d_{i\cdots K}}] - [p_m (1 - p_m)]^{d_{i\cdots K}{d_j\cdots K}}$$

where parameters $a_{i\cdots K}^{d_{i\cdots K}}$ and $y_{i\cdots K}^{d_{i\cdots K}}$ are functions of identity states probabilities between gametes of $ij \cdots K$ individuals at marker $m$.

- When individual are repeated (e.g. $i = j$), $E\left[ \cdots X_{im}^{d_i} X_{jm}^{d_j} \cdots \right] = E\left[ \cdots X_{im}^{d_i+d_j} \cdots \right]$.

Let $\tau_z = \sum_l \left( 2p_m (1 - p_m) \right)^2$ and $a_{ij}$ the coancestry coefficient between individuals $i$ and $j$.

$$\sum_m E\left[ X_{im} X_{jm} X_{km} X_{im} \right] = \frac{1}{2} \tau a_{1111}^{1111} - \frac{1}{4} \tau_2 y_{1111}^{1111}$$

$$\sum_m E\left[ X_{im} X_{km} \right] = \frac{1}{2} \tau a_{11}^{111} - \frac{1}{4} \tau_2 y_{111}^{111} = 2a_{ij}\tau$$

$$\sum_m E\left[X_{im} X_{km} \right] E\left[ X_{jm} X_{im} \right] \left( 2 \sum_m E\left[ X_{jm} X_{im} \right] = 2a_{ij}\tau \right)$$

Thus $E[XX'TXX'] = \sum_k \sum_{jkl} \left( \left( \frac{1}{2} \tau a_{ijkl}^{1111} - \frac{1}{4} \tau_2 y_{ijkl}^{1111} + 4a_{ik}a_{jl}(\tau^2 - \tau_2) \right) \right)$. Elements of this summation are computed considering the third property given above.

**Demonstration.**

Let $S_{c_1c_2\cdots c_n}$ the identity state between $n_c$ chromosomes at a given locus. This is an extension of the identity coefficients [35-37]. We do not consider pairs of chromosomes of two individuals, but a set of $n_c$ chromosomes which may or not belong to different individuals. For instance, $c_1$ could mean the paternal allele at locus $m$ for individual $i$ (it will be noted in this case $c_1=\text{is}$). The figure SM1-1 represents the possible states, depending on the number of chromosomes. When $n_c = 2$ only two identity states are possible: locus are IBD ($S_{c_1c_2} = 1$) or not IBD ($S_{c_1c_2} = 2$). When $n_c = 3$ five identity states are possible and when there are 4 chromosomes, we find back the 15 classical identity states.

The codification for the genotypes are $X_{im} = (0, 1$ or $2) - 2p_m$. (It is equivalent to $X_{im} = g_{ims} + g_{imd}$ where $g_{ims}$ and $g_{imd}$ are the “values” of the alleles transmitted to individual $i$ by its sire and its dam, with $g_{ims}$ and $g_{imd} = (0$ or $1) - p_m$. As we only consider one locus in the following derivation the $m$ indice will be omitted: $X_i = g_{is} + g_{id}$ and $g_{is}$ and $g_{id} = (0$ or $1) - p$.

Different situations were encountered for the product $E\left[ X_i^{d_i} X_j^{d_j} \cdots X_{K}^{d_K} \right]$.

- $E[X_i X_j] i \neq j$

- $E[X_i X_j^2] i \neq j$
In all case, we will decompose the $X$ genotypes in their $g$ values

$$E[X_i^2 X_j^2] \neq j$$

$$E[X_i X_j X_k^2] \neq j \neq k$$

$$E[X_i X_j X_k X_l] \neq j \neq k \neq l$$

This formula turns to be the weighted sum of elements such as $E[g_{is}^{d_i} g_{js}^{d_j-1} g_{jd} \cdots | S_{is,js,jd,\ldots}]$ which are very simple to derive from the allele frequency $p$.

Computations are simplified by the fact that those expectations are null when one of the elements of the product, here say $g_{jd}$, is uniq and the state $S_{is,js,jd,\ldots}$ such that the $jd$ locus is not IBD with any other locus.

1) $E[X_i X_j] \neq j$

$$E[X_i X_j] = E[g_{is} g_{js}] + E[g_{is} g_{jd}] + E[g_{id} g_{js}] + E[g_{id} g_{jd}]$$

$$E[g_{is} g_{js}] = p(S_{is,js} = 1)\{(1-p)(0-p)^2 + p(1-p)^2\} + p(S_{is,js} = 0)\{(1-p)^2(0-p)^2 + 2p(1-p)(0-p) + p^2(1-p)^2\}$$

$$E[g_{is} g_{jd}] = p(S_{is,js} = 1)\{p(1-p)\}$$

$$E[X_i X_j] = p(1-p)[p(S_{is,js} = 1) + p(S_{is,jd} = 1) + p(S_{id,js} = 1) + p(S_{id,jd} = 1)]$$

It must be noted that the event $S_{is,js} = 1$ etc. corresponds to the classical identity states 1, 2, 4, 9, 10, that is $p(S_{is,js} = 1) = \delta_1 + \delta_2 + \delta_4 + \delta_5 + \delta_9 + \delta_{10};$ that $S_{is,jd} = 1$ corresponds to the 1, 3, 4, 12, 13 etc. The results being that

$$E[X_i X_j] = p(1-p)[4\delta_1 + 2[\delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_9 + \delta_{12}] + \delta_{10} + \delta_{11} + \delta_{13} + \delta_{14}] = 4p(1-p) a_{ij},$$

as expected

$$E[X_i X_j] = 4p(1-p) a_{ij}$$

2) $E[X_i X_j^3] \neq j$

$$E[X_i X_j^3] = E[(g_{is} + g_{id})(g_{js} + g_{jd})^2] = E[g_{is} g_{js}^2] + 3E[g_{is} g_{js} g_{jd}] + 3E[g_{is} g_{js} g_{jd}] + 3E[g_{id} g_{js} g_{jd}] + 3E[g_{id} g_{js} g_{jd}] + E[g_{id} g_{jd} g_{jd}]$$

$$E[g_{is} g_{js}^2] = p(S_{is,js} = 1)\{(1-p)(0-p)^4 + p(1-p)^4\} = p(S_{is,js} = 1)(p(1-p)(1-3p(1-p))$$

$$E[g_{is} g_{js} g_{jd}] = p(S_{is,js,jd} = 1)\{(1-p)(0-p)^4 + p(1-p)^4\} = p(S_{is,js,jd} = 4)\{(1-p)^2(0-p)^4 + 2p(1-p)(0-p)^2(1-p)^2 + p^2(1-p)^4\} = p(S_{is,js,jd} = 1)(p(1-p)(1-3p)) + p(S_{is,js,jd} = 4)[p(1-p)\]$$

$$E[X_i X_j^3] = p(1-p)[1-3p(1-p)]\{p(S_{is,js} = 1) + p(S_{is,jd} = 1) + p(S_{id,js} = 1) + p(S_{id,jd} = 1) + 6p(S_{is,js,jd} = 1) + 6p(S_{is,js,jd} = 1) + [p(1-p)]^2 3p(S_{is,js,jd} = 4) + 3p(S_{is,js,jd} = 2) + 3p(S_{id,js,jd} = 4) + 3p(S_{id,js,jd} = 2)\]$$
\[ E[X_i X_j^2] = p(1-p)[(1-3p(1-p))[16\delta_1 + 2(\delta_2 + \delta_3) + 8(\delta_4 + \delta_5) + 2(\delta_6 + \delta_12) + \delta_10 + \delta_{11} + \delta_13 + \delta_14) + [p(1-p)]^2[6(\delta_2 + \delta_3 + \delta_9 + \delta_{12}) + 3(\delta_{10} + \delta_{11} + \delta_{13} + \delta_{14})] \]

3) \[ E[X_i^2 X_j^2] \ i \neq j \]

\[ E[X_i^2 X_j^2] = E \left[ (g_{iS} + g_{id})^2 (g_{js} + g_{jd})^2 \right] = E[g_{iS}^2 g_{js}^2] + E[g_{iS}^2 g_{jd}^2] + E[g_{id}^2 g_{js}^2] + E[g_{id}^2 g_{jd}^2] + 2E[g_{iS}^2 g_{js} g_{jd}] + 2E[g_{iS} g_{id} g_{js} g_{jd}] + 2E[g_{id} g_{js} g_{jd}] + 4E[g_{iS} g_{id} g_{js} g_{jd}] \]

\[ E[g_{iS}^2 g_{js}^2] = p(S_{iS,js} = 1)((1-p)(0-p)^4 + p(1-p)^4) + p(S_{iS,js} = 2)((1-p)^2(0-p)^2 + 2p(1-p)(1-p)^2(0-p)^2 + p(S_{iS,js} = 1)(p(1-p)[p^3 + (1-p)^3]) + p(S_{iS,js} = 2)[p(1-p)]^2 \]

\[ E[g_{iS}^2 g_{js} g_{jd}] = p(S_{iS,js,jd} = 1)((1-p)(0-p)^4 + p(1-p)^4) + p(S_{iS,js,jd} = 3)((1-p)^2(0-p)^2 + 2p(1-p)(1-p)^2(0-p)^2 + p^2(1-p)^4) \]

The state \( S_{iS,js,jd} = 3 \) corresponds to \( g_{js} \) IBD to \( g_{jd} \) and \( g_{is} \) not IBD to the others. In the other states we have terms like \( E[g_{jd}] = 0 \) in the conditional expectation \[ E[g_{iS}^2 g_{js} g_{jd} | S_{iS,js,jd}] \].

\[ E[g_{iS}^2 g_{jd} g_{js} g_{jd}] = p(S_{iS,js,jd} = 1)(p(1-p)[p^3 + (1-p)^3]) + p(S_{iS,js,jd} = 3)[p(1-p)]^2 \]

\[ E[g_{iS} g_{id} g_{js} g_{jd}] = p(S_{iS,js,jd} = 1)((1-p)(0-p)^4 + p(1-p)^4) + p(S_{iS,js,jd} = 6) + p(S_{iS,js,jd} = 9) + p(S_{iS,js,jd} = 12)[(1-p)^2(0-p)^2 + 2p(1-p)(1-p)^2 + p^2(1-p)^4] \]

Finally

\[ E[X_i^2 X_j^2] = (p(1-p)[p^3 + (1-p)^3)] [p(S_{iS,js} = 1) + p(S_{iS,js} = 1) + p(S_{id,js} = 1) + 2p(S_{id,js} = 1) + 2p(S_{id,js} = 1) + 4p(S_{id,js} = 1) + [p(1-p)]^2 [p(S_{iS,js} = 2) + p(S_{iS,js} = 2) + p(S_{iS,js} = 2) + 2p(S_{iS,js} = 3) + 2p(S_{id,js} = 3) + 2p(S_{id,js} = 3) + 2p(S_{id,js} = 3) + 4p(S_{id,js} = 3)] \]
\[ S_{ls,js} = 1 \quad x \quad x \quad x \quad x \quad x \quad x \quad 1 \]
\[ S_{ls,jd} = 1 \quad x \quad x \quad x \quad x \quad x \quad x \quad 1 \]
\[ S_{id,js} = 1 \quad x \quad x \quad x \quad x \quad x \quad x \quad 1 \]
\[ S_{id,jd} = 1 \quad x \quad x \quad x \quad x \quad x \quad x \quad 1 \]
\[ S_{is,js,jd} = 1 \quad x \quad x \quad x \quad x \quad x \quad x \quad 1 \]
\[ S_{id,js,jd} = 3 \quad x \quad x \quad x \quad x \quad x \quad x \quad 2 \]
\[ S_{is,ids} = 2 \quad x \quad x \quad x \quad x \quad x \quad x \quad 2 \]
\[ S_{is,js,jd} = 2 \quad x \quad x \quad x \quad x \quad x \quad x \quad 2 \]
\[ S_{is,js,jd} = 6 \quad x \quad x \quad x \quad x \quad x \quad x \quad 4 \]
\[ S_{is,js,jd} = 9 \quad x \quad x \quad x \quad x \quad x \quad x \quad 4 \]
\[ S_{is,js,jd} = 12 \quad x \quad x \quad x \quad x \quad x \quad x \quad 4 \]

\[
E[X_i^2 X_j^2] = [p(1 - p)[1 - 3p(1 - p)](16\delta_1 + 4(\delta_2 + \delta_3 + \delta_4 + \delta_5) + 2(\delta_9 + \delta_{12}) + \delta_{10} + \delta_{11} + \delta_{13} + \delta_{14}) + p(1 - p)^2(4(\delta_2 + \delta_3 + \delta_4 + \delta_5 + \delta_{15}) + 16\delta_6 + 8(\delta_7 + \delta_8) + 6(\delta_9 + \delta_{12}) + 3(\delta_{10} + \delta_{11} + \delta_{13} + \delta_{14}))]
\]

4) \[E[X_i X_j X_k^2] \quad i \neq j \neq k\]

\[
E[X_i X_j X_k^2] = E[(g_{ls} + g_{ld})(g_{js} + g_{jd})(g_{ks} + g_{kd})^2] = E[g_{ls} g_{js} g_{ks}^2] + E[g_{ls} g_{jd} g_{kd}^2] + E[g_{id} g_{js} g_{ks}^2] + E[g_{id} g_{jd} g_{kd}^2] + 2E[g_{ls} g_{js} g_{ks} g_{kd}] + 2E[g_{id} g_{jd} g_{ks} g_{kd}] + 2E[g_{ls} g_{jd} g_{ks} g_{kd}] + 2E[g_{id} g_{js} g_{ks} g_{kd}]
\]

\[
E[g_{ls} g_{js} g_{ks}^2] = p(S_{is,js,ks} = 1)(1 - p)(0 - p)^2 + p(1 - p)^2 + p(S_{is,js,ks} = 2)(1 - p)^2(0 - p)^2 + 2p(1 - p)(1 - p)^2 + p(1 - p)^2 = p(S_{is,js,ks} = 1)p(1 - p)[1 - 3p(1 - p)] + p(S_{is,js,ks} = 2)p(1 - p)^2
\]

\[
E[g_{ls} g_{jd} g_{kd}^2] = p(S_{is,js,ks,kd} = 1)p(1 - p)[1 - 3p(1 - p)] + p(S_{is,js,ks,kd} = 6) + p(S_{is,js,ks,kd} = 9) + p(S_{is,js,ks,kd} = 12)p(1 - p)^2
\]

5) \[E[X_i X_j X_k X_l] \quad i \neq j \neq k \neq l\]

\[
E[X_i X_j X_k X_l] = E[(g_{ls} + g_{ld})(g_{js} + g_{jd})(g_{ks} + g_{kd})(g_{ls} + g_{ld})] = \\
\sum_{a_i \in \{s,d\}} \sum_{a_j \in \{s,d\}} \sum_{a_k \in \{s,d\}} \sum_{a_l \in \{s,d\}} E[g_{ia} g_{ja} g_{ka} g_{la}]
\]
\[ E[g_{is}g_{js}g_{ks}g_{ls}] = p(S_{is,js,ks,ls} = 1)p(1 - p)[1 - 3p(1 - p)] + \{p(S_{is,js,ks,ls} = 6) + p(S_{is,js,ks,ls} = 9) + p(S_{is,js,ks,ls} = 12)\}[p(1 - p)]^2 \]

\[ E[X_iX_jX_kX_l] = p(1 - p)[1 - 3p(1 - p)] \sum_{a_i \in \{s,d\}} \sum_{a_j \in \{s,d\}} \sum_{a_k \in \{s,d\}} \sum_{a_l \in \{s,d\}} p(S_{ia,ja,ka,la} = 1) + [p(1 - p)]^2 \sum_{a_i \in \{s,d\}} \sum_{a_j \in \{s,d\}} \sum_{a_k \in \{s,d\}} \sum_{a_l \in \{s,d\}} p(S_{ia,ja,ka,la} = 6) + p(S_{ia,ja,ka,la} = 9) + p(S_{ia,ja,ka,la} = 12) \]
Figure SM1-1

Possible IBD states, depending on the number of chromosomes

2 locus

S1  S2
1
2

3 locus

S1  S2  S3  S4  S5
1
2
3

4 locus

S1  S2  S3  S4  S5  S6  S7  S8  S9  S10  S11  S12  S13  S14  S15
1
2
3
4