Additional file 1 - G matrix computation based on similarities

Similarity (S) can be measured as the number of shared alleles between individuals \( j \) and \( k \) for each genotype at locus \( i \), such as [1]

\[
S_{jk, i} = \frac{I_{j1k1} + I_{j1k2} + I_{j2k1} + I_{j2k2}}{4}
\]

For a single locus, \( S_{jk} = \)

<table>
<thead>
<tr>
<th>( \text{ind}_j / \text{ind}_k )</th>
<th>AA</th>
<th>AB</th>
<th>BB</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>1</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>AB</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>BB</td>
<td>0</td>
<td>0.5</td>
<td>1</td>
</tr>
</tbody>
</table>

We shall now show that this is similar to computing relationships using the following equations as outlined by Yang et al [2], with allele frequency \( p_i \) fixed at 0.5 for all variants.

\[
G_{jk} = \frac{1}{N} \sum_l G_{ijk} = \begin{cases} 
\frac{1}{N} \sum_i \frac{(x_{ij} - 2p_i)(x_{ij} - 2p_i)}{2p_i(1 - p_i)}, & j \neq k \\
1 + \frac{1}{N} \sum_i x_{ij}^2 - (1 + 2p_i)x_{ij} + 2p_i^2}{2p_i(1 - p_i)}, & j = k
\end{cases}
\]

Let’s consider a unique locus,

\[
(1) \quad \frac{(x_{ij} - 2p_i)(x_{ij} - 2p_i)}{2p_i(1 - p_i)} = \frac{(x_{ij} - 1)(x_{ij} - 1)}{0.5}
\]

\[
(2) \quad 1 + \frac{x_{ij}^2 - (1 + 2p_i)x_{ij} + 2p_i^2}{2p_i(1 - p_i)} = 1 + \frac{x_{ij}^2 - 2x_{ij} + 0.5}{0.5}
\]

\[
= \frac{x_{ij}^2 - 2x_{ij} + 1}{0.5} = \frac{(x_{ij} - 1)(x_{ij} - 1)}{0.5}
\]

If \( x_{ij} = x_{ik} \) then \( 1 = 2 \) and only one equation is needed to calculate both diagonal and off-diagonal elements.

In this case, \( G_{jk} = \)
\[
\begin{array}{cccc}
\text{ind}_j / \text{ind}_k & \text{AA [2]} & \text{AB [1]} & \text{BB [0]} \\
\hline
\text{AA [2]} & 2 & 0 & -2 \\
\text{AB [1]} & 0 & 0 & 0 \\
\text{BB [0]} & -2 & 0 & 2 \\
\end{array}
\]

S and G are linked by the following transformation: \(4 \cdot (S_{jk} - \overline{S_{jk}}) = G_{jk}\), with \(\overline{S_{jk}} = 0.5\).

In the case of multiple loci,

\[
G_{jk} = \frac{1}{N} \sum_i G_{ijk} = \frac{1}{N} \sum_i (x_{ij} - 1)(x_{ik} - 1) \cdot \frac{0.5}{0.5}
\]

Which is equivalent, in matrix notation, to \(G = \frac{(M-1)(M-1)'}{(N/2)}\), where M is the genotype matrix containing values of 0, 1 and 2.