A. Constraints on individual LD values

The normalized individual LD values are $D_{ij}' = D_{ij} / D_{max}$, where $D_{max} = \min[p_{Ai}(1-p_{Bj}), (1-p_{Ai})\,p_{Bj}]$ if $D_{ij}>0$, and

$D_{max} = \min[p_{Ai}\,p_{Bj,}\,(1-p_{Ai})(1-p_{Bj})]$ when $D_{ij}<0$. The range of D_{ij}' is $(-1, 1)$.

B. Other measures of conditional LD

The conditional association between the alleles at *marker* locus B and *disease* locus A was defined by Nei and Li (1980) as

$d_* = (f_{A2B1} / p_{A2} - f_{B1A1} / p_{A1})$ for two bi-allelic loci. It was developed to account for study designs for rare diseases with known

modes of inheritance and full, or close to full, penetrance (e.g., sickle cell anemia and Duchenne muscular dystrophy) where

individuals are not randomly sampled from a single population. This measure is equivalent to Somer's D statistic, $D(C|R)$,

conditioning on the rows of the contingency table relating two categorical variables, as shown below.

$$d_* = f_{A2B1} / p_{A2} - f_{A1B1} / p_{A1} = (p_{A2}\,p_{B1} + D)/p_{A2} - (p_{A1}\,p_{B1} - D)/p_{A1}$$
$$= p_{B1} + (D/p_{A2}) - p_{B1} + (D/p_{A1}) = D/(p_{A1}p_{A2})$$

Somer's D statistic is defined as twice the difference between the number of concordant and discordant entries in the

contingency table, divided by $n^2 - \Sigma_i\,n_{i\cdot}^2$, where n is the table total and $n_{i\cdot}$ is the ith row total.

$$D(C|R) = 2(n_{11}\,n_{22} - n_{21}\,n_{12})/(n^2 - \sum_i n_{i\square}^2) = 2(f_{A1B1}f_{A2B2} - f_{A2B1}f_{A1B2})/(1^2 - \sum_i p_{Ai}^2)$$
$$= 2D/(1 - F_A) = 2D/(2p_{A1}p_{A2}) = D/(p_{A1}p_{A2})$$

C. Other measures of LD based on diversity statistics

Other measures of association and LD that are based on allelic diversity statistics (homozygosity and heterozygosity) have been

defined. However, these measures are all symmetric. Ohta (1980) suggested a measure, F', that divides the difference between

the two-locus haplotypic homozygosity (F_{AB}) and the product of the two single locus homozygosity values by the product of the

single locus heterozygosity values: $F' = (F_{AB} - F_A\,F_B) / [(1 - F_A)\,(1 - F_B)]$ (Ohta 1980). The $D*$ measure (Maruyama 1982; Hedrick

and Thomson 1986) standardizes the disequilibrium between all alleles at two loci by the product of the single locus

heterozygosity values: $D* = \Sigma_i \Sigma_j\,D_{ij}^2 / [(1 - F_A)\,(1 - F_B)]$. Note that in the bi-allelic case $D*$ is equivalent to the square of the

correlation measure (r^2) (Hedrick 1987).

D: Proofs of Special Cases (c) – (f)

D.1. Multi-Allelic Case (c): $k_A = k_B = k$, $f(A_iB_i) > 0$, $i = 1, 2, ... k$, and $f(A_iB_j) = 0$ for all $i \neq j$

Summary: $W_n = W_{A/B} = W_{B/A} = 1$. There is complete symmetry and 100% correlation of alleles at the two loci.

The A and B locus allele frequencies are equal and the notation is simplified as follows: $p_{Ai} = p_{Bi} = p_i$, $i = 1, 2, ... k$,

$f(A_iB_i) = p_i = p_i^2 + D_{ii}$, i.e., $D_{ii} = p_i(1 - p_i)$, $f(A_iB_j) = 0 = p_i\,p_j + D_{ij}$, i.e., $D_{ij} = -p_i\,p_j$, $i \neq j$.

$W_{A/B}^2 = \{\Sigma_i \Sigma_j\,[f_{ij}^2 / p_{Bj}] - F_A\} / (1 - F_A) = [\{\Sigma_i f_{ii}^2/p_i\} - F_A]\} / (1 - F_A)$

G. Thomson and R. M. Single

$$= [\{\Sigma_i \, p_i^2/p_i\} - F_A] / (1 - F_A) \;=\; [\{\Sigma_i \, p_i\} - F_A] / (1 - F_A) = (1 - F_A) / (1 - F_A) \;=\; 1.$$

Similarly, $W_{B/A}^2 = 1$.

$$W_n^2 \quad = \{\Sigma_i \Sigma_j \, D_{ij}^2 / (p_{Ai} \, p_{Bj})\} / (k-1) \;=\; [\{\Sigma_i \, D_{ii}^2 / (\, p_i^2)\} + \{\Sigma_i \Sigma_{j\neq i} \, D_{ij}^2 / (p_i \, p_j)\}] \;/(k-1)$$

$$= [\{\Sigma_i \, p_i^2 \, (1 - p_i)^2 / (p_i^2)\} + \{\Sigma_i \Sigma_{j\neq i} \, (p_i \, p_j)^2 / (p_i \, p_j)\}] / (k-1) = [\Sigma_i \, (1 - p_i)^2 + \{\Sigma_i \Sigma_{j\neq i} \, (p_i \, p_j)\}] / (k-1)$$

$$= [\Sigma_i \, (1 - p_i)^2 + \, p_i \, (1 - p_i)] / (k-1) = [\Sigma_i \, (1 - p_i)] / (k-1) = (k-1) / (k-1) = 1.$$

D.2. Multi-Allelic Case (d): $k_A = k_B = k$, $f(AiBi) > 0$, $i = 1, 2, \ldots k$, and $f(A1B2) \neq 0$

In this case $W_{A/B} < 1$, $W_{B/A} < 1$, and $W_n < 1$ (Proof given below for $k = 3$, with the same result holding for any value of k).

<u>Summary</u>: $W_{A/B}^2 = \{1 - 2\,\delta\,[p_{A2} / p_{B2}] - F_A\} / (1 - F_A)$, $W_{B/A}^2 = \{1 - 2\,\delta\,[p_{B1} / p_{A1}] - F_B\} / (1 - F_B)$,

$$W_n^2 = 1 - \{\delta\,(p_{A1} + p_{A2}) / [2\,p_{A1}\,p_{B2}]\}$$

Allele frequencies are (A locus): p_{A1}, p_{A2}, and p_{A3}; (B locus): $p_{B1} = p_{A1} - \delta$, $p_{B2} = p_{A2} + \delta$, and $p_{B3} = p_{A3}$, with $\delta > 0$ and $\delta < p_{A1}$.

Haplotype	Frequency*	Frequency*	LD
A1B1	$p_{A1} - \delta$	$p_{A1} \, (p_{A1} - \delta) + D_{11}$	$D_{11} = (p_{A1} - \delta) \, (1 - p_{A1})$
A1B2	δ	$p_{A1} \, (p_{A2} + \delta) + D_{12}$	$D_{12} = -\, p_{A1} \, (p_{A2} + \delta) + \delta$
A1B3	0	$p_{A1} \, p_{A3} + D_{13}$	$D_{13} = -\, p_{A1} \, p_{A3}$
A2B1	0	$p_{A2} \, (p_{A1} - \delta) + D_{21}$	$D_{21} = -\, p_{A2} \, (p_{A1} - \delta)$
A2B2	p_{A2}	$p_{A2} \, (p_{A2} + \delta) + D_{22}$	$D_{22} = p_{A2} \, (1 - p_{A2} - \delta)$
A2B3	0	$p_{A2} \, p_{A3} + D_{23}$	$D_{23} = -\, p_{A2} \, p_{A3}$
A3B1	0	$p_{A3} \, (p_{A1} - \delta) + D_{31}$	$D_{31} = -\, p_{A3} \, (p_{A1} - \delta)$
A3B2	0	$p_{A3} \, (p_{A2} + \delta) + D_{32}$	$D_{32} = -\, p_{A3} \, (p_{A2} + \delta)$
A3B3	p_{A3}	$p_{A3} \, p_{A3} + D_{33}$	$D_{33} = p_{A3} \, (1 - p_{A3})$

* These two columns are different, but equivalent, formats for the same haplotype frequency.

$$W_{A/B}^2 = \{\Sigma_i \Sigma_j \, [f_{ij}^2 / p_{Bj}] - F_A\} / (1 - F_A)$$

$$= \{[(p_{B1}^2) / (p_{B1}) + (\delta^2) / (p_{B2}) + (p_{A2}^2) / (p_{B2}) + (p_{A3}^2) / (p_{A3})] - F_A\} / (1 - F_A)$$

$$= \{[p_{B1} + (\delta^2) / (p_{B2}) + (p_{A2}^2) / (p_{B2}) + p_{A3}] - F_A\} / (1 - F_A)$$

$$= \{[(p_{A1} - \delta) + (\delta^2) / (p_{A2} + \delta) + (p_{A2})^2 / (p_{A2} + \delta) + p_{A3} + p_{A2} - p_{A2}] - F_A\} / (1 - F_A)$$

$$= \{[\, p_{A1} + p_{A2} + p_{A3} - [(\delta \, (p_{A2} + \delta) - (\delta^2) - (p_{A2})^2 + p_{A2} \, (p_{A2} + \delta)] / (p_{A2} + \delta)] - F_A\} / (1 - F_A)$$

$$= \{1 - 2\,\delta\,[p_{A2} / p_{B2}] - F_A\} / (1 - F_A) < 1 \text{ always since } \delta > 0$$

$$W_{B/A}^2 = \{\Sigma_i \Sigma_j \, [f_{ij}^2 / p_{Ai}] - F_B\} / (1 - F_B)$$

$$= \{[(p_{A1} - \delta)]^2) / (p_{A1}) + (\delta^2) / (p_{A1}) + p_{A2} + p_{A3}] - F_B\} / (1 - F_B)$$

$$= \{[\, p_{A1} + p_{A2} + p_{A3} - [(p_{A1})^2 - (\delta^2) - (p_{A1} - \delta)^2] / (p_{A1})] - F_B\} / (1 - F_B)$$

$$= \{1 - 2\,\delta\,[(p_{A1} - \delta) / p_{A1}] - F_B\} / (1 - F_B) = \{1 - 2\,\delta\,[p_{B1} / p_{A1}] - F_B\} / (1 - F_B) < 1 \text{ always since } \delta > 0.$$

$$W_n^2 = \{\Sigma_i \Sigma_j \, D_{ij}^2 / (p_{Ai} \, p_{Bj})\} / (k-1)$$

$$= \{[(p_{A1} - \delta) \, (1 - p_{A1})^2 / p_{A1}] + p_{A1} \, (p_{A2} + \delta) - 2\delta + \delta^2 / [\, p_{A1} \, (p_{A2} + \delta)] + p_{A1} \, p_{A3} + p_{A2} \, (p_{A1} - \delta) + [p_{A2} / (p_{A2} + \delta)] - 2 \, p_{A2}$$

$$+ p_{A2} (p_{A2} + \delta) + p_{A2} p_{A3} + p_{A3} (p_{A1} - \delta) + p_{A3} (p_{A2} + \delta) + (1 - p_{A3})^2 \} / 2$$

$$= \{(1 - p_{A1})^2 - [\delta (1 - p_{A1})^2 / p_{A1}] + p_{A1} p_{A2} + \delta p_{A1} - 2\delta + \delta^2 /[p_{A1} (p_{A2} + \delta)] + p_{A1} p_{A3} + p_{A2} p_{A1} - \delta p_{A2} + [p_{A2} / (p_{A2} + \delta)]$$

$$+ [1 - (p_{A2} + \delta) / (p_{A2} + \delta)] - 2 p_{A2} + p_{A2}^2 + \delta p_{A2} + p_{A2} p_{A3} + p_{A3} p_{A1} - \delta p_{A3} + p_{A3} p_{A2} + \delta p_{A3} + (1 - p_{A3})^2 \} / 2$$

$$= \{(1 - p_{A1})^2 + 2 p_{A1} p_{A2} + 2 p_{A1} p_{A3} + 2 p_{A2} p_{A3} + (1 - p_{A2})^2 + (1 - p_{A3})^2 - [\delta (1 - p_{A1})^2] / [p_{A1}] + \delta p_{A1} - 2\delta + \delta^2 / [p_{A1} (p_{A2} + \delta)]$$

$$- [\delta / (p_{A2} + \delta)]\} / 2$$

$$= \{3 - 2 (p_{A1} + p_{A2} + p_{A3}) + (p_{A1} + p_{A2} + p_{A3})^2 - [\delta / p_{A1}] + 2\delta - \delta p_{A1} + \delta p_{A1} - 2\delta + \delta^2 / [p_{A1} (p_{A2} + \delta)] - [\delta / (p_{A2} + \delta)]\} / 2$$

$$= 1 - \{\delta (p_{A1} + p_{A2}) / [2 p_{A1} p_{B2}]\} < 1 \text{ always since } \delta > 0.$$

D.3. Multi-Allelic Case (e): $k_A < k_B$, with each B_j allele occurring with only one A_i allele

In this case $W_n = W_{A/B} = 1$, and $W_{B/A} < 1$ (Proof given below for $k_A = 3$ and $k_B = 4$, and the equivalent result holds if $k_A > k_B$.)

Summary: $W_{A/B}^2 = 1$, $W_{B/A}^2 = \{1 - [2 p_{B3} p_{B4} / (p_{B3} + p_{B4})] - F_B\} / (1 - F_B)$, $W_n^2 = 1$

Haplotype	Frequency*	Frequency*	LD
A1B1	p_{B1}	$p_{B1}^2 + D_{11}$	$D_{11} = p_{B1} (1 - p_{B1})$
A1B2	0	$p_{B1} p_{B2} + D_{12}$	$D_{12} = - p_{B1} p_{B2}$
A1B3	0	$p_{B1} p_{B3} + D_{13}$	$D_{13} = - p_{B1} p_{B3}$
A1B4	0	$p_{B1} p_{B4} + D_{14}$	$D_{14} = - p_{B1} p_{B4}$
A2B1	0	$p_{B2} p_{B1} + D_{21}$	$D_{21} = - p_{B2} p_{B1}$
A2B2	p_{B2}	$p_{B2}^2 + D_{22}$	$D_{22} = p_{B2} (1 - p_{B2})$
A2B3	0	$p_{B2} p_{B3} + D_{23}$	$D_{23} = - p_{B2} p_{B3}$
A2B4	0	$p_{B2} p_{B4} + D_{24}$	$D_{24} = - p_{B2} p_{B4}$
A3B1	0	$p_{B1} (p_{B3} + p_{B4}) + D_{31}$	$D_{31} = - p_{B1} (p_{B3} + p_{B4})$
A3B2	0	$p_{B2} (p_{B3} + p_{B4}) + D_{32}$	$D_{32} = - p_{B2} (p_{B3} + p_{B4})$
A3B3	p_{B3}	$p_{B3} (p_{B3} + p_{B4}) + D_{33}$	$D_{33} = p_{B3} (1 - p_{B3} - p_{B4})$
A3B4	p_{B4}	$p_{B4} (p_{B3} + p_{B4}) + D_{34}$	$D_{34} = p_{B4} (1 - p_{B3} - p_{B4})$

Allele frequencies at the A locus are: p_{A1}, p_{A2}, and $p_{A3} = (p_{B3} + p_{B4})$, and at the B locus they are: $p_{B1} (= p_{A1})$, $p_{B2} (= p_{A2})$, p_{B3} and p_{B4}.

$$W_{A/B}^2 = \{\Sigma_i \Sigma_j [f_{ij}^2 / p_{Bj}] - F_A\} / (1 - F_A) = \{p_{B1} + p_{B2} + p_{B3} + p_{B4} - F_A\} / (1 - F_A) = 1$$

$$W_{B/A}^2 = \{\Sigma_i \Sigma_j [f_{ij}^2 / p_{Ai}] - F_B\} / (1 - F_B)$$

$$= \{p_{B1} + p_{B2} + [p_{B3}^2 / (p_{B3} + p_{B4})] + [p_{B4}^2 / (p_{B3} + p_{B4})] - F_B\} / (1 - F_B)$$

$$= \{p_{B1} + p_{B2} + [(p_{B3}^2 + p_{B4}^2) / (p_{B3} + p_{B4})] - F_B\} / (1 - F_B)$$

$$= \{p_{B1} + p_{B2} + [(p_{B3} + p_{B4})^2 / (p_{B3} + p_{B4})] - [2 p_{B3} p_{B4} / (p_{B3} + p_{B4})] - F_B\} / (1 - F_B)$$

$$= \{1 - [2 p_{B3} p_{B4} / (p_{B3} + p_{B4})] - F_B\} / (1 - F_B) < 1 \text{ always}$$

$$W_n^2 = \{\Sigma_i \Sigma_j D_{ij}^2 / (p_{Ai} p_{Bj})\} / (k_A - 1)$$

$$= \{(1 - p_{B1})^2 + p_{B1} p_{B2} + p_{B1} p_{B3} + p_{B1} p_{B4} + p_{B2} p_{B1} + (1 - p_{B2})^2 + p_{B2} p_{B3} + p_{B2} p_{B4} + p_{B1} (p_{B3} + p_{B4}) + p_{B2} (p_{B3} + p_{B4}) + 1 - 2 p_{A3} + p_{A3}^2 \} / 2$$

$$= \{3 - 2 (p_{B1} + p_{B2} + p_{A3}) + (p_{B1} + p_{B2} + p_{A3})^2 \} / 2 = 1$$

D.4. One Locus Bi-Allelic and the Other Multi-Allelic, Case (f): $k_A = 2$, $k_B > 2$.

G. Thomson and R. M. Single

In this case $W_n = W_{A/B} \neq W_{B/A}$ (Proof given below for $k_A = 2$ and $k_B = 3$, with the obvious extension for any value of k_B)

Summary: $W_{A/B}^2 = [\{D_{11}^2 / (p_{B1})\} + \{D_{12}^2 / (p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{B3})\}] / (p_{A1} p_{A2})$, $W_{B/A}^2 = [\{D_{11}^2 + D_{12}^2 + (D_{11} + D_{12})^2\}] / \{(p_{A1} p_{A2}) (1 - F_B)\}$,

$W_n^2 = W_{A/B}^2$

Haplotype	Frequency	LD*
A1B1	$p_{A1} p_{B1} + D_{11}$	D_{11}
A1B2	$p_{A1} p_{B2} + D_{12}$	D_{12}
A1B3	$p_{A1} p_{B3} + D_{13}$	$- D_{11} - D_{12}$
A2B1	$p_{A2} p_{B1} + D_{21}$	$- D_{11}$
A2B2	$p_{A2} p_{B2} + D_{22}$	$- D_{12}$
A2B3	$p_{A2} p_{B3} + D_{23}$	$D_{11} + D_{12}$

* See Table 1 footnote for constraints on LD parameters, such that two LD variables, D_{11} and D_{12}, along with three allele

frequencies (p_{A1}, p_{B1}, and p_{B2}) are sufficient to describe the five independent haplotype frequencies in this case.

$W_{A/B}^2 = \{\Sigma_i \Sigma_j D_{ij}^2 / (p_{Bj})\} / (1 - F_A)$

$= [\{D_{11}^2 / (p_{B1})\} + \{D_{12}^2 / (p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{B3})\} + \{D_{11}^2 / (p_{B1})\} + \{D_{12}^2 / (p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{B3})\}] / (2 p_{A1} p_{A2})$

$= [\{D_{11}^2 / (p_{B1})\} + \{D_{12}^2 / (p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{B3})\}] / (p_{A1} p_{A2})$

$W_{B/A}^2 = \{\Sigma_i \Sigma_j D_{ij}^2 / (p_{Ai})\} / (1 - F_B)$

$= [\{D_{11}^2 / (p_{A1})\} + \{D_{12}^2 / (p_{A1})\} + \{D_{11} + D_{12})^2 / (p_{A1})\} + \{D_{11}^2 / (p_{A2})\} + \{D_{12}^2 / (p_{A2})\} + \{D_{11} + D_{12})^2 / (p_{A2})\}] / (1 - F_B)$

$= [\{D_{11}^2 + D_{12}^2 + (D_{11} + D_{12})^2\}] / \{(p_{A1} p_{A2}) (1 - F_B)\}$

$W_n^2 = \{\Sigma_i \Sigma_j D_{ij}^2/(p_{Ai} p_{Bj})\} / (2-1)$

$= [\{D_{11}^2/(p_{A1} p_{B1})\} + \{D_{12}^2/(p_{A1} p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{A1} p_{B3})\} + \{D_{11}^2/(p_{A2} p_{B1})\} + \{D_{12}^2/(p_{A2} p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{A2} p_{B3})\}]$

$= [\{D_{11}^2 / (p_{B1})\} + \{D_{12}^2 / (p_{B2})\} + \{D_{11} + D_{12})^2 / (p_{B3})\}] / (p_{A1} p_{A2}) = W_{A/B}^2$
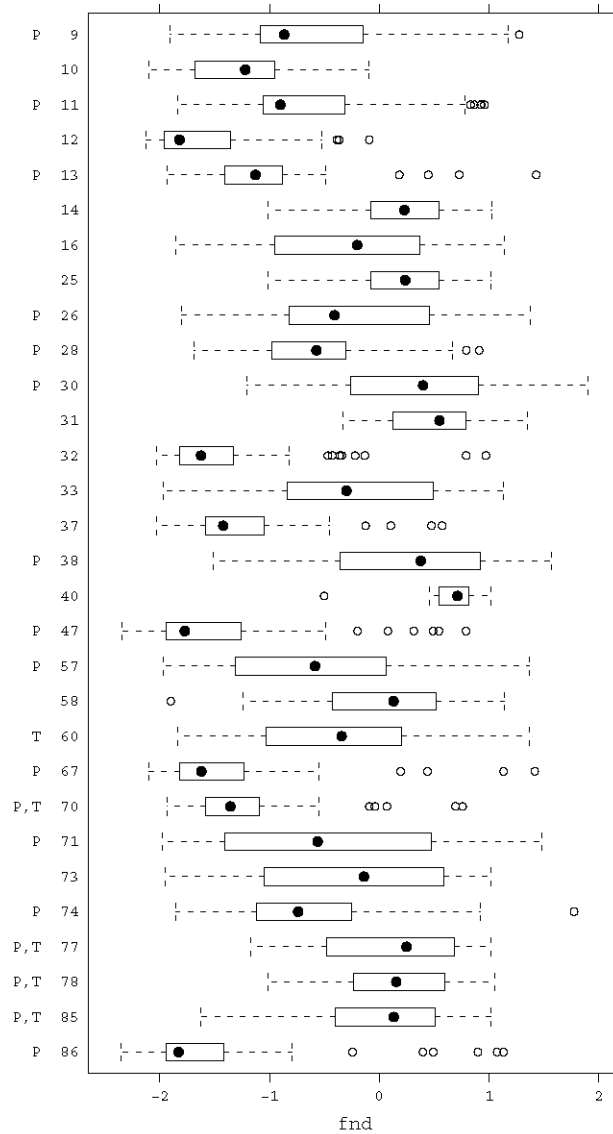
Figure S1 Boxplots of F_{nd} values for *DRB1* Amino Acids 9 to 86 from 57 Populations (Lancaster 2006)*

*AA position is indicated on the vertical axis with P indicating a peptide interacting site and T indicating a T-cell interacting site. The boxes indicate the middle 50% of the Fnd values across the 57 populations with a line drawn at the median value. The spread of the central half of the data is called the interquartile range. Extreme observations that are more than 1.5 times the interquartile range away from the central box are identified with circles.

G. Thomson and R. M. Single