

# Homework #6

*Ezgi Karaesmen*

*November 3, 2016*

## Chapter 13

### Problem 13.1

Clark's algorithm was applied to determine haplotypes of 6 individuals consisting of 3 diallelic SNPs.

Individual	Haplotype	SNP1	SNP2	SNP3
1	1	G	C	A
	2	G	C	A
2	1	G	C	A
	2	G	A	A
3	1	G	C	A
	2	G	A	G
4	1	G	A	G
	2	G	A	G
5	1	G	C	A
	2	A	C	G
6	1	G	C	G
	2	A	A	G

1. Step 1: Identify all unambiguous haplotypes. Consider these haplotypes "resolved".

Resolved haplotypes:  $h_1 = h(GCA)$  ;  $h_2 = h(GAA)$  ;  $h_3 = h(GAG)$

2. Step 2: Determine whether any of the resolved haplotypes could be one of the haplotypes in a still unresolved individual. If not, STOP. Otherwise continue.

The given haplotypes for individual 6 are  $h(GCG)$  and  $h(AAG)$ . However, haplotypes for the same individual can also be determined as  $h(ACG)$  and  $h(GAG)$  without losing information. Theoretically, the algorithm is designed to search for possible "resolved haplotypes" in each individual and should be able to handle such cases, individual 6's haplotypes were changed in the table given below to ease the demonstration.

Individual	Haplotype	SNP1	SNP2	SNP3	Status	Assigned.haplotype
1	1	G	C	A	resolved	$h_1$
	2	G	C	A	resolved	$h_1$
2	1	G	C	A	resolved	$h_1$
	2	G	A	A	resolved	$h_2$
3	1	G	C	A	resolved	$h_1$
	2	G	A	G	resolved	$h_3$
4	1	G	A	G	resolved	$h_3$
	2	G	A	G	resolved	$h_3$
5	1	G	C	A	resolved	$h_1$
	2	A	C	G	not resolved	
6	1	A	C	G	not resolved	
	2	G	A	G	resolved	$h_3$

- Step 3: Each time a resolved haplotype is identified as one of the possible alleles in an unresolved individual, identify the complement and consider this newly identified homologue resolved. Go back to Step 2.

Newly resolved haplotype:  $h_4 = h(ACG)$

- Return to Step 2

Individual	Haplotype	SNP1	SNP2	SNP3	Status	Assigned.haplotype
1	1	G	C	A	resolved	$h_1$
	2	G	C	A	resolved	$h_1$
2	1	G	C	A	resolved	$h_1$
	2	G	A	A	resolved	$h_2$
3	1	G	C	A	resolved	$h_1$
	2	G	A	G	resolved	$h_3$
4	1	G	A	G	resolved	$h_3$
	2	G	A	G	resolved	$h_3$
5	1	G	C	A	resolved	$h_1$
	2	A	C	G	resolved	$h_4$
6	1	A	C	G	resolved	$h_4$
	2	G	A	G	resolved	$h_3$

STOP.

Upon applying the Clark's algorithm 4 haplotypes were identified:

$h_1 = h(GCA)$  ;  $h_2 = h(GAA)$  ;  $h_3 = h(GAG)$  ;  $h_4 = h(ACG)$

## Extra Problem

Considering a haplotype block made up of 4 SNPs and every SNP is diallelic,  $2^4 = 16$  haplotypes are possible.

There are only the following 8 haplotypes occurring in a population. Is there any redundancy, i.e., can we ignore one or more SNPs without losing any information?

In order to answer this question, the given haplotypes were sorted and color coded. It can easily be seen that SNP2 and SNP3 are in perfect LD and for every A allele in SNP2, T allele is observed in SNP4 and consequently for every C allele in SNP2, A allele is observed in SNP4. Therefore, we can conclude that SNP4 can be ignored without losing information.

Haplotype	SNP1	SNP2	SNP3	SNP4
3	C	A	C	T
1	C	A	G	T
7	C	C	C	A
4	C	C	G	A
8	T	A	C	T
6	T	A	G	T
5	T	C	C	A
2	T	C	G	A