

**Appendix 3 to:**

***Measures of Genetic Diversity***

**Analysis of molecular variance: Example 2**

As described in Appendix 2, this model (AMOVA) measures gene diversity among populations, this time with specific reference to populations within an area of a region in a continent (situation 4, slide 26). It has a new hierarchical level (region), with its respective parametric values and estimators of the mean squares.

We have: i = individuals, j = alleles, k = populations, l = regions

$$Y_{lki(j)} = Y + r_l + a_{l(k)} + b_{lk(i)} + w_{lki(j)}$$

Where,

- $Y_{lki(j)}$  = a value between 0 and 1 for the  $j^{\text{th}}$  allele of the  $i^{\text{th}}$  individual of the  $k^{\text{th}}$  population in the  $l^{\text{th}}$  region
- $r_l$  = the effect of the  $l^{\text{th}}$  region, with variance  $\sigma^2r$
- $a_{l(k)}$  = the effect of the  $k^{\text{th}}$  population within the  $l^{\text{th}}$  region, with variance  $\sigma^2a$
- $b_{lk(i)}$  = the effect of the  $i^{\text{th}}$  individual within the  $k^{\text{th}}$  population in the  $l^{\text{th}}$  region, with variance  $\sigma^2b$
- $w_{lki(j)}$  = effect of the  $j^{\text{th}}$  locus within individual  $i$  of the  $k^{\text{th}}$  population of the  $l^{\text{th}}$  region, with variance  $\sigma^2c$
- $n$  = the product of  $i, j, k$  and  $l$ , that is the total number of observations

Source of variation	df	MS	EMS
Between regions	$l - 1$	$MS_r$	$\sigma^2w + 2\sigma^2b + 2n\sigma^2a + 2nl\sigma^2r$
Between pops. within reg.	$l(k - 1)$	$MS_a$	$\sigma^2w + 2\sigma^2b + 2n\sigma^2a$
Between indiv./pop./reg.	$lk(i - 1)$	$MS_b$	$\sigma^2w + 2\sigma^2b$
Within individuals	$lki(j - 1)$	$MS_w$	$\sigma^2w$
Total	$lkij - 1$		
Total variance (%)	$\% \sigma^2r = (\sigma^2r/\sigma^2) * 100$		$\% \sigma^2a = (\sigma^2a/\sigma^2) * 100$
$\sigma^2 = \sigma^2r + \sigma^2w + \sigma^2b + \sigma^2a$	$\% \sigma^2b = (\sigma^2b/\sigma^2) * 100$		$\% \sigma^2w = (\sigma^2w/\sigma^2) * 100$

$\sigma^2r$  is the parametric value of the variance between regions and is estimated by  $(CMA - CMB)/2nl$ .

In the variance estimates, the sign '%' is added because we can express the variance accounted for each source (region, population within a region, individuals within a population) as a function of the total variance and, as such, we can determine which of the variation components is the most important. For example, if the value of the variation due to regions was high and those of the other sources was low, we could conclude that populations within regions have homogeneous allelic frequencies, but populations from different regions differ notably in their allelic frequencies.